Universidad de Concepción
Dirección de Postgrado
Facultad de Humanidades y Arte - Programa de Magíster en Lingüística Aplicada

The effects of recast and metalinguistic oral corrective feedback on the perception and production of intonation patterns (fall and rise) of English wh-questions in L2 learners

Tesis para optar al grado académico de Magíster en Lingüística Aplicada

JUAN PABLO CRUCES PÉREZ
CONCEPCIÓN–CHILE
2023

Profesor Guía: Dr. Yasna Ivonne Pereira Reyes
Profesor Co-guía: Dr. Mauricio Véliz-Campos
Departamento de Idiomas Extranjeros, Facultad de Humanidades y Arte
Universidad de Concepción

Proyecto de investigación financiado por ANID – BECA MAGÍSTER NACIONAL 22210834.

Proyecto de investigación desarrollado y patrocinado por proyecto FONDECYT Regular 2022 1220209.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABSTRACT

The current study aimed to establish the effects of perceptual training and production training using recast and metalinguistic oral corrective feedback (OCF) for the intonation patterns (fall and rise) of English wh-questions in L2 learners. Additionally, a non-word repetition task (NWRT) to measure phonological working memory (PWM), a listening span task (LST) to measure central executive (CE) ability, and a listening proficiency test were used to explore their associations with the perception and production of the intonation patterns of English wh-questions, and the two types of OCF (recast and metalinguistic) provided. 40 participants were randomly allocated to two experimental groups; all of them received the same perceptual training. As for the production training, the groups were given either recast or metalinguistic OCF individually. Participants were given pre- and post-tests to measure their perception and production of the intonation patterns for English wh-questions. Results showed that both types of OCF had a similar impact on improving the production accuracy of the patterns for wh-questions after training. As for the perception ability, participants showed a near ceiling effect before training which did not leave much room for improvement. Measures of PWM, CE ability and proficiency level did not show any connections with perception or production abilities of these L2 learners nor the two types of OCF. Finally, production training with OCF may seem to be a useful strategy to improve L2 speech production beyond the level of segments.

1. INTRODUCTION AND PRESENTATION OF THE PROBLEM

The teaching of pronunciation has been increasingly integrated into the L2 classrooms since the paradigm shift from a form-focused to a meaning-focused approach in teaching English as a Second Language (ESL). This transition resulted in a change of the learning objectives, where intelligibility became the goal. In order to attain an intelligible pronunciation, it is imperative to identify which pronunciation features interplay in the perception and production of intelligible speech. As reported by some studies, suprasegmental errors (e.g., misplaced nuclear stress and intonation) cause the most serious communication breakdowns, and therefore they have a greater impact on intelligibility (Levis, 2018; Scheuer & Horgues, 2021). In this regard, a comparison of the intonation patterns of English and Chilean Spanish wh-questions is relevant for this study. English uses a falling tone (default) to request new information (Wells, 2009), whereas Chilean Spanish uses a rising tone for that same purpose (Rebolledo, 2021; Valenzuela, 2013). Conversely, a rising tone in English wh-questions is used to show interest, sympathy, or deference. As a result of these differences, negative transfer is likely to occur, which may lead speakers to erroneously perceive or produce them (Ortega-Llebaria & Colantoni, 2014; Quesada & Romero, 2018).

In view of this problem, some researchers advocate that prosodic aspects should be given emphasis and priority in pronunciation instruction (Gordon et al., 2013;

Gordon & Darcy, 2016). In addition to the integration of suprasegmental instruction in L2 classrooms, corrective feedback (CF) has been theoretically established as a reliable strategy to prevent and reduce linguistic errors. Given that L1 negative transfer is thought to be notably persistent in prosody (Geçkin, 2020; Lu & Kim, 2016), being intonation particularly challenging for L2 learners (Brandl et al., 2020; Cardinali & Barbeito; 2018; Naranjo, 2020; Puga et al., 2017), the provision of oral corrective feedback (OCF) becomes even more important.

Studies comparing recast and metalinguistic OCF to correct segmental or suprasegmental features are still scarce as the majority of them have only studied the effects of one type of OCF. For instance, A study that looked into the effect of recast on the acquisition of the English sound /ɹ/ indicates that this is effective (Saito, 2013). Another study found that the combination of recast with explicit instruction of pronunciation makes recast more evident and thus more effective (Lyster et al., 2013). Furthermore, Gooch et al. (2016) found that recast only had a facilitative role in the perception of English sounds. Suprasegmental and OCF research has yielded mixed results. On the one hand, recast to correct lexical stress was not found to produce statistically significant results (Parlak & Ziegler, 2016). On the other hand, a comparison between recast and metalinguistic OCF on Mandarin tones showed recast to be more effective (Bryfonski & Ma, 2019).

In light of the limited number of previous empirical studies, and mixed results, the present study's main aim is to determine and compare the impact of recast and metalinguistic OCF on the learning of suprasegmentals in English. Additionally, individual differences have been found to predict second language (L2) learning success (Dörnyei & Ryan, 2015) and the learners' ability to notice CF; for example, L2 proficiency (Allen & Mills, 2016; Li, 2014), working memory (Yilmaz, 2013), motivation (Azizi & Nemati, 2018), and anxiety (Luquin & Roothooft, 2019; Teimouri et al., 2019). Among these factors, L2 proficiency and working memory capacity (WMC) are still largely understudied. Therefore, these two variables were included in this study.

In summary, the present study's main aim was to determine which OCF type (recast and metalinguistic) was more effective to produce gains in the perception and production of English intonation (fall and rise) for wh-questions in L2 learners after training. Furthermore, secondary aims included (a) to determine the role of individual differences (working memory capacity such as PWM and central executive ability, and L2 proficiency) in mediating the effects of these two types of OCF and (b) to establish any associations among the perception and production of the intonation patterns, the OCF provided, and individual differences.

2. LITERATURE REVIEW

2.1. Suprasegmentals: definition and its status

Suprasegmentals usually refer to speech features such as stress, intonation and rhythm which extend beyond the limits of a single segment (Lee & Thomson, 2022). In contrast, segmental features can be described as the individual sounds into which human speech can be divided (Lee & Thomson, 2022). For many years, the study and description of human speech sounds were particularly concerned with segments, and consequently suprasegmentals were neglected (Cabrelli, 2023; Prieto, 2015; Solon & Long, 2018). However, with the use of methods and approaches that favour communication (e.g., Communicative Language Teaching, Content and Language Integrated Learning, Task-Based Language Teaching) where meaning assumes a central role, intelligibility and comprehensibility are identified as a more realistic and rightful goal of L2 pronunciation attainability and, by implication, L2 teaching (Derwing & Munro, 2015; Foote & McDonough, 2017).

Intelligibility refers to the listeners' actual understanding of L2 speech, whereas comprehensibility is defined as the listeners' perceptions of understanding (Derwing & Munro, 2009). Park et al. (2017) state that if L2-English speakers fail to produce intelligible and comprehensible speech, they will very likely encounter problems communicating with either English native speakers or other non-native speakers of English. Among the features that may cause serious intelligibility and comprehensibility difficulties, suprasegmentals have been identified (Bozorgian & Shamsi, 2020; Scheuer & Horgues, 2021; Yurtbasi, 2017). In addition to this, and considering the multiple features of prosody, it is unsurprising that suprasegmental properties are particularly difficult to acquire (Hattori, 2023; Tan, 2016). Studies suggest that this complication has its grounds on our L1 suprasegmental patterns being deeply engrained in the brain, given the fact that babies acquire these features during the first month of life (Yang, 2016). On this, Geçkin (2020) points out that L1 negative transfer is very common and persistent in L2 prosody acquisition.

As intonation is particularly relevant for the current study, this will be now examined.

2.1.1. Intonation: definitions, functions, realisations, dimensions, and relevant research

Intonation is usually defined as a complex unit of melody (pitch variation), "stress, temporal components (tempo, length, pausing), rhythm and voice timbre colour" (Xasanov & Mardieva, 2023, p. 488). In Wells' opinion (2009), the study of intonation concerns understanding how melody (pitch variation) conveys linguistic and pragmatic meaning as well as how stressed and unstressed syllables interplay to make intonation patterns emerge. Other definitions that highlight the role of pitch include Peng's (2016) in which intonation is described as pitch changes (the rising and falling of the voice) people produce when they speak.

Along similar lines, Hamawand and Hasan AL-JAF (2023) refer to intonation as pitch fluctuations used by a speaker to express meanings. These different definitions point to two main characteristics of intonation. Firstly, intonation is realised in terms of pitch movements (i.e., the rising and falling of the voice) and, secondly, these pitch movements are used to convey different meanings (e.g., linguistic, or pragmatic).

Pitch variations are usually referred to as intonation patterns. According to Tench (2015), the most common English intonation patterns are described as fall, rise, fall-rise, and rise-fall. Depending on the language, these intonation patterns can fulfil a lexical, grammatical, or pragmatic function. Mandarin Chinese, for instance, uses pitch to distinguish lexical meaning, that is, a change in the intonation pattern of a syllable can produce different words; for example, mā (mother), má (hemp), mǎ (horse), mà (scold), and ma (interrogative particle). Grammatical function refers to the use of pitch to distinguish between different types of sentences (e.g., statements, questions, commands, requests, etc.). For example, in the following English examples: “the dog ↘bit you” and “the dog ↗bit you?”, the falling tone ↘ makes the sentence a statement, whereas the rising tone ↗ makes it a question. In this way, a variation in the intonation pattern in English can change the grammatical function of the utterance. English native speakers also use pitch pragmatically. For example, yes-no questions use a rising tone as default pattern; however, depending on the attitude or emotion of the speaker, a yes-no question can be asked using a falling tone, which may convey insistence, seriousness or even threat (Wells, 2009).

English intonation can also be described in terms of tonality, tonicity, and tone (Halliday, 1967; Wells, 2009). Thus, every time a speaker produces speech, they first divide the sentence(s) into chunks known as intonation phrases - IPs (Wells, 2009), process known as tonality. Additionally, the speaker highlights a word as important (tonicity). This highlighted word or part of word (syllable) is usually contained in the last lexical item within the IP and is known as nucleus (Wells, 2009). Finally, the speaker must decide which intonation pattern (tone) the nucleus will carry to convey linguistic and communicative purposes.

In the current study, intonation will be understood as the falling and rising variations that the voice makes when we produce speech to convey the following pragmatic functions: to request information; to show respect, sympathy, or interest; confirm information; and to show surprise, amazement, or shock. Therefore, the analysis of intonation in this study only considers the nucleus (found in the last lexical item of each IP), and tone (fall and rise).

According to Mennen (2015), languages may differ in their intonations along four dimensions (modified from Ladd, 1996): (a) the systemic dimension, (b) the semantic dimension, (c) the frequency dimension, and (d) the realisational dimension. The systemic dimension deals with the inventory of structural elements (e.g., pitch accents and boundary tones). The semantic dimension is concerned with what linguistic functions these elements convey (e.g., interrogativity, encouragement, etc.). The frequency dimension refers to how often the structural elements are used and how they combine. Finally, the realisational dimension relates to the way in which the elements of intonation are phonetically implemented. Therefore, for a complete acquisition of the intonation of an L2, the learner is expected to master all four dimensions. Consequently, the acquisition of L2 intonation is notably challenging for learners.

Empirical studies that have investigated the acquisition of L2 intonation include Brandl et al.'s study (2020) which examined the perception of intonation in questions and statements in L2-Spanish learners. Data was collected from 189 L2-Spanish university students with proficiency levels ranging from beginner to advanced, and 20 L1-Spanish speakers who served as a control group. Both groups participated in matching intonation perception tasks where they had to decide whether the presented aural stimuli matched the sentence shown on a screen. The measures used were learners' accuracy and reaction time. Results showed that the L1 group was significantly more accurate and faster than the L2 group in perceiving the intonation patterns. Individual results suggest that intermediate L2-Spanish learners are not able to process the intonation cues to differentiate the declarative sentences from the questions. With similar results, Cardinali and Barbeito (2018) explored whether the instruction of English intonation contributed to the development of intonation (i.e., tonality, tonicity, and tone) in L2-English university students with L1-Spanish. The instruction focused on segmental and suprasegmental features using the systemic functional theory of language. Activities used during the sessions included awareness-raising exercises, oral practice and reading out loud tasks. With regard to intonation, raters observed that the participants' overall performance was affected by errors in tonality and tone, which led raters to perceive the intonation patterns as less intelligible.

Puga et al. (2017) studied the perception of English intonation patterns of German-speaking learners of advanced English. Participants had to look at and listen to sentences, and then choose the tonal pattern that seemed most appropriate to them. It is important to note that participants had access to a printed version of the story to use to extract contextual information (for sarcasm, in particular). Also, they were allowed to replay the sentences as many times as needed. The results showed that the participants perceived statements (fall pattern) and yes-no questions (rise pattern) close to that of the British English control group, but their perception performance was significantly less accurate than the control group in the open (rise) and closed (fall) tag questions, and the expressions of sarcasm (rise-fall/fall). Despite German having the rising tone in its inventory, the participants failed to identify the rising tone in tag questions. An explanation for this may be the lack of tag questions in German, which may have resulted in the participants heavily relying on the syntactic features of the written sentences to determine the intonation pattern. A similar explanation was given to explain the low accuracy for the sentences involving sarcasm. As sarcasm is a purely pragmatic phenomenon, participants may have failed to use the syntactic features of the sentences to choose correctly.

Naranjo (2020) conducted a study to determine the perception ability of the default intonation pattern in English wh-questions in L1-Spanish speakers learning English. English proficiency levels and perception test response time were measured to establish relations with the participants' ability to perceive the default pattern. The results suggested that proficiency does not play a significant role in enhancing the perception ability of the wh-question default pattern in the learners since there was no significant differences between the two groups of participants (more proficient vs. less proficient). Data gathered from the perception test response time showed that learners took a longer time to answer incorrectly despite their proficiency level. Overall, these findings point to the challenge that the identification of the default pattern of the wh-question poses to L2 learners.

A description of English and Spanish intonations is presented in the following section to have a clear understanding on how they differ.

2.1.2. English and Spanish intonation: characterisation and differences

English and Spanish have been classified as intonational languages as they use intonation to convey syntactic, pragmatic, and attitudinal meaning (Henriksen, 2009; Wells, 2009). To achieve this, the subsystems of tonality, tonicity, and tone (Wells, 2009) are put into effect. In the following paragraphs, a description and ensuing comparison of the intonation systems of English and Chilean Spanish using these subsystems will be presented.

In respect of English tonality, the decision where to break the utterance into intonation phrases (IPs) has been suggested as a matter of common sense (Wells, 2009). The intonation break usually corresponds to a syntactic boundary; in other words, breaks are placed between successive (a) sentences, (b) clauses, (c) phrases, (d) words, or even (e) within words. The decision where to break a stream of speech into IPs is determined exclusively by the speaker's intentions (Wells, 2009).

Spanish tonality has been described in terms of intonation groups (Hidalgo & Quilis, 2012). An intonation group is defined as a unit of speech which has its own melodic unit (tone) and is divided by pauses (Hidalgo & Quilis, 2012). Therefore, an utterance can be divided into intonation groups which, in Wells' terms, would constitute different IPs. The decision where an intonation group starts and ends in Spanish follows a similar logic as English (Gutiérrez, 2012). Thus, syntactic boundaries correspond with the intonation breaks. Differences in tonality between English and Spanish are found in a small number of cases (Gutiérrez, 2012): (a) restrictive relative clauses, (b) final vocative, (c) transferred vs non-transferred negation, and (d) nominal clauses introduced by zero conjunction in English.

In terms of tonicity, English and Chilean Spanish speakers must decide where the nucleus goes within the IP. The nucleus refers to the syllable containing the nuclear tone (a fall, a rise, a fall-rise, or a rise-fall). In general, function words (e.g., articles, possessive determiners, prepositions, and conjunctions) do not receive tone unit pitch prominence in either language. However, in Spanish personal pronouns and auxiliary verbs are realised with rhythmic stress and a full vowel which makes them susceptible of receiving onset pitch prominence when they are the first rhythmically stressed word in the IP. Conversely, in English these word classes do not have rhythmic stress and are realised with a weak vowel, therefore being unsusceptible of bearing an onset (Gutiérrez, 2012).

Wells (2009) states that English always places prominence on the last or near last content word of the IP, whereas Spanish places prominence on the word further to the right irrespective of the word type (content or function word). Examples of the difference in the placement of prominence (underlined syllable) in (a) English and (b) Spanish are as follows: (a) "John asked me to talk to him" (Chela-Flores, 2003, p.259), (b) "John me pidió que hablara con él" (Chela-Flores, 2003, p.259). This difference may lead Spanish speakers to negatively transfer these prominence placement patterns into English in situations where it is unnecessary. Furthermore, Chela-Flores (2003) mentions that Spanish uses two syntactic procedures to highlight words in an utterance: changing the order of the elements or adding other words to the utterance. Examples of the latter are: "Juan me prestó su bicicleta, a mí me prestó Juan su bicicleta, or fue Juan el que me prestó su bicicleta" (Chela-Flores, 2003, p.264). In English, word order change is also acceptable; however, using phonological prominence seems to be more common.

In English, only the lexically stressed syllable of a word can take the nucleus, which means we need to know which syllable in each word bears the stress to produce intonation patterns correctly. Therefore, to accent the word *always,* we place the nucleus on the first syllable. For *between*, as its stress is on the last syllable, we accent the last syllable. In addition to this, some syllables in English may be subject to vowel reduction when the schwa (/ə/) sound is present, making it more challenging to know where the stress falls; for example, the word cotton can be produced with its two syllables made evident as in /ˈkɒt.ən/ or with /ə/ being elided resulting in /n/ becoming a syllabic consonant as in /ˈkɒt n̩/. Many Spanish speakers only recognise the syllables that follow the phonotactics of their L1, which may result in confusion when exposed to syllabic consonants.

Finally, English tones have been categorised as falling or non-falling (Wells, 2009). According to Wells (2009), falling tones are those where the pitch of the voice starts relatively high and then moves downwards (i.e., fall, and rise-fall tones). In non-falling tones, on the contrary, the pitch of the voice starts relatively low and then moves upwards (i.e., rise, and fall-rise tones). Roach (2009) uses the concepts of simple and complex to refer to fall and rise, and fall-rise and rise-fall respectively. For Spanish, three tones have been identified: rise, fall, and sustained (Rogers et al., 2021; Urrutia, 2007).

Default tones in both languages help determine the grammatical category of the utterances. The falling tone (as default) in English is used for statements, wh-questions, exclamations, commands, interjections and closed lists. In Peninsular

Spanish, similarly, the falling tone (as default) is used for statements, wh-questions, exclamations, commands, and closed lists (Hidalgo & Quilis, 2012). However, in Latin American Spanish varieties, the default wh-question intonation pattern varies from country to country (Sosa, 2003). In the Chilean Spanish variety, the default intonation pattern for wh-questions is a rising tone (Rebolledo, 2021; Valenzuela, 2013). Comparatively, the default rising tone in English is used for polar questions (e.g., yes-no questions), declarative questions, dependent clauses, topic clauses, and open lists (Wells, 2009). As a result of these differences, transfer of Spanish patterns in the perception and production of English wh-questions is likely to occur.

The default complex rising tone (fall-rise) in English is used to show that the clause is dependent, or that the information in the clause constitutes a comment (i.e., what we say about the topic). In Spanish, a default fall-rise tone does not commonly occur (Navarro Tomás, 1966). As for the complex falling tone (rise-fall), this is not used grammatically neither in English nor Spanish. Spanish also has a sustained tone which is not common in English. A sustained (or level) tone is used to show that an utterance is unfinished or has been interrupted (Navarro Tomás, 2004). Conversely, in English, unfinished or interrupted utterances are conveyed using a rising tone.

Pragmatically, tones have varied uses in English. In general, English uses a falling tone on yes-no questions to convey insistence, and a rising tone on declarative sentences to convey encouragement, reservation, tentativeness, polite corrections, partial statements, or warning. Wh-questions can also be asked with a rising tone with the pragmatic function of encouragement, and deference. A rise-fall tone in English is pragmatically used in statements, exclamations, and yes-no questions to show that the speaker is impressed. It may also convey challenge or disapproval when used in wh-questions (Wells, 2009); however, this tone is not usually considered important for L2 learners to acquire (Roach, 2009). For this reason, the rise-fall tone was excluded from this study. Some studies that have looked into the pragmatic intonation configurations of Peninsular Spanish wh-questions discovered that a rising tone is used to tone down or reiterate the question, and that a rise-fall is used to convey surprise or amazement (Henriksen, 2009; Sosa, 2003). Unfortunately, there are no studies that have examined the pragmatic variations of wh-questions of the Chilean Spanish language to draw a comparison between English and this Spanish variety.

In conclusion, English and Chilean Spanish intonation patterns vary in some of their grammatical and pragmatic realisations and uses, which may lead to negative transfer from the L1 to the L2. Methods typically used to approach this problem are corrective feedback, and perception or production training with or without feedback.

2.2. Corrective feedback (CF): definitions, types, and relevant research

Corrective feedback (CF) is usually defined as any reaction to a learner's erroneous utterance (Li, 2018). The characteristics of these reaction moves vary considerably in their realisation in the classroom, which led Lyster and Ranta (1997) to identify six CF types: recasts, explicit correction, elicitation, repetition, clarification request, and metalinguistic clue. Furthermore, CF types can be classified in terms of how explicit or implicit they are, or whether they provide (input-providing) or withhold (output-prompting) the correct form (Lyster et al., 2013). From an explicit/implicit perspective, implicit CF is provided inconspicuously to the learner expecting them alone to notice the error (e.g., recast, repetition, clarification request); whereas explicit CF overtly indicates that an error has been committed, and then provides the learner with the correct form (e.g., explicit correction, elicitation, metalinguistic clue). The input-providing and output-prompting distinction places emphasis on whether the correction provides the correct form (providing) or encourages learners to self-correct (prompting). By this logic, recast is a type of input-providing CF, whereas metalinguistic feedback falls within the output-prompting one. In the current study, only recast and metalinguistic CF were used, thus the importance to define and discuss any relevant research into their effectiveness on the perception and production of English pronunciation.

With reference to metalinguistic feedback, Lyster and Ranta (1997) define it as explicit comments, information or questions related to grammaticality of learner's utterance. In this study, metalinguistic feedback will describe a comment meant to explicitly correct a learner's error.

Defining recasts, Nassaji (2013) states that these are reformulations of all or part of a learner's incorrect utterance within the context of a communicative activity. Despite the fact that recast is typically classed as implicit CF, there are abundant examples of what are known as didactic recasts which are characterised for being remarkably explicit (Hanh & Tho, 2018; Sheen, 2013). The degree of explicitness that recasts can adopt is regulated by emphatic stress or pitch movements (R. Ellis et al., 2006). Respecting this last point, Sheen (2007) indicates that the degree of explicitness rather than the CF strategy would be a determining factor in the efficacy of the feedback. Thus, didactic recasts would help diminish the ambiguity and be more effective (Llinares & Lyster, 2014). Along similar lines, Mackey et al. (2007) concluded that as pronunciation corrections are more difficult to be interpreted as such, it would be best to be as direct and explicit as possible. In the present study, recast will be understood as a complete reformulation of a learner's utterance. The reformulation will only contain the corrected intonation pattern.

Some studies have investigated the effects of explicit and implicit oral corrective feedback (OCF) to correct pronunciation errors. For example, Saito (2013) conducted an experiment which looked into the pedagogical potential of recasts together with form-focused instruction (FFI) on the perception and production of the English consonant sound /ɹ/ by Japanese speakers. Three groups were formed (FFI + recast, FFI-only, and control). Results revealed that both the FFI + recast and FFI-only groups attained perception and production gains; however, only the FFI + recast group's gains showed to be generalisable in both trained and untrained word contexts. This finding suggests that recasts play an important role in L2 pronunciation development. Also, this study reveals that the provision of explicit instruction combined with pronunciation-focused recasts may help make the recast more evident (Lyster et al., 2013).

Saito (2015) conducted a similar study in which recast was provided on word-initial /ɹ/ to Japanese learners of English. 54 learners were divided into an experimental group and a control group. All participants received FFI, but recast was only given to the experimental group. Results showed that both groups reduced their F2 values, but only the experimental group reduced their F3 values. According to these results, recasts together with FFI are a useful tool for getting learners to shift their focus from their default interlanguage strategy (the F2 variation) to new phonetic cues (the F3 variation). Overall, both studies conducted by Saito (2013, 2015) seem to point towards a positive effect of explicit phonetic instruction (which includes CF) in L2 pronunciation development (Saito & Plonsky, 2019).

Furthermore, Gooch et al. (2016) carried out a study comparing the effects of prompts (type of explicit CF) and recasts on the acquisition of the English /ɹ/ by Korean native speakers. FFI was also provided, however, this was given before the provision of OCF. The researchers reported that recasts only had a facilitative role in helping learners comprehend the sounds in controlled tasks whereas prompts were effective for both controlled and free production. Darabad (2014) also compared the effects of recast and prompts on -s and -es endings of English words. Participants were 72 Azari-Turkish and Persian bilinguals learning English and were assigned to three groups: recast experimental group, prompt experimental group, and control group. The findings indicated that both OCF conditions had a positive effect on the learners' pronunciation accuracy; however, recast produced greater positive effects than those of prompts.

Abedi et al. (2015) compared the impact of recast and explicit OCF on general pronunciation accuracy. 40 Iranian learners of English formed two experimental groups which either received recast or explicit OCF. After treatment, the post-tests indicated that recast had a significant effect on the participants' scores compared to explicit OCF.

Finding opposing results, Jalal and Alahmed (2022) examined the impact of two types of OCF (recast vs explicit) on general pronunciation in low-proficiency Iraqi learners of English. To achieve this, two experimental groups (one received recast and the other explicit OCF) and a control group (no feedback) were formed. The statistical analyses found that both explicit OCF and recast were beneficial for enhancing learners' English pronunciation. Nonetheless, explicit OCF showed to be more beneficial. These results may indicate that explicit types of OCF are more beneficial when correcting pronunciation errors in learners with a low-proficiency level.

Zohrabi and Behboudnia (2017) carried out a study comparing the effects of implicit OCF (recast or clarification requests) and explicit OCF (direct error correction, elicitation or metalinguistic cues) on segmental word-level pronunciation. 60 Irani EFL learners formed three groups: implicit group, explicit group and control group. Results showed that both implicit and explicit OCF were beneficial in significantly reducing pronunciation errors compared to the control group. Also, both experimental groups showed significant immediate and delayed effects in reducing the treated errors.

Furthermore, Luquin and Roothooft (2019) compared the effects of recast and metalinguistic OCF on the pronunciation of English -ed endings. 64 L1 Spanish speakers were allotted to three conditions: recast experimental group, metalinguistic experimental group, and control group. Results revealed that both experimental groups made significant improvements from pre- to post-test, but no significant differences were found between groups. Furthermore, significant differences were only found between the recast group and the control group, which again may suggest an advantage of recast over explicit types of OCF to treat pronunciation errors.

Some research studies have also informed into the impact of OCF on suprasegmentals. For instance, Parlak and Ziegler (2016) investigated the impact of recasts on the development of Englis lexical stress by Arabic native speakers. Participants were assigned to four different OCF conditions: FTF (face-to-face) recast, FTF control, SCMC (computer-mediated communication) recast, and SCMC control. Results showed that there were no statistically significant differences across groups. It is noteworthy that the recast used in this study was conversational (i.e., unmarked), which may have influenced the effectiveness of the treatment.

Karimi and Esfandiari (2016) also conducted a study investigating the role of OCF on stress patterns. In their study, they provided Iranian EFL learners with either recast or explicit CF. In this way, two experimental groups were constituted as well as a control group. The results reported that both recast and explicit CF produced positive gains in the participants. However, the effect of recast showed to be stronger than that of the explicit CF. These findings may indicate that recast may be more beneficial than explicit types of CF to treat suprasegmental errors.

Another study was conducted by Bryfonski and Ma (2019), who examined the effects of metalinguistic and recast OCF on the perception and production of Mandarin tones in beginner L2 learners. The findings showed that the learners in the recast group had greater improvement in tone production compared to the metalinguistic group. In terms of tone identification, both groups significantly improved from pre- to post-test; however, no statistically significant differences were found between them. The authors suggest that individual differences (e.g., anxiety, attention and attentional control, working memory, and developmental level) should be taken into account for further studies to explain their influence in the processing of OCF.

A study which also dealt with the perception of Mandarin tones was conducted by Saito and Wu (2014). In their study, 41 Cantonese learners of Mandarin were assigned into three groups: FFI + recast, FFI-only, and control. Statistical analyses of the post-tests revealed that both FFI-only and FFI + recast groups attained significant improvement in all lexical and tonal contexts. The FFI + recast only showed marginally significant gains under the trained lexical conditions. These findings may suggest that FFI without corrective feedback may be sufficient to promote the development of L2 pronunciation.

2.2.1. Speech perception and production training: relevant research

Speech perception refers to “the cognitive process by which sounds are heard and categorised. This contrasts with speech production, which refers to the output of the cognitive system, mediated by physical control of speech gestures” (Thompson, 2022, p.43). For the purpose of this study, the suprasegmental feature of intonation will be considered and highlighted in the definition. In this way, every time that speech perception is mentioned in the text, this will refer to the cognitive process by which intonation patterns are heard and categorised as fall or rise. This will also be true for speech production, which will refer to the output of the cognitive system, mediated by physical control of speech gestures to produce intonation patters (rising or falling of the voice).

Feedback can also be provided in the context of speech perception and production training. As the present study deals with perception and production training together with the provision of OCF, the description of previous studies examining the effects of different training types with or without feedback is pertinent.

Studies exploring the impact of perception training with or without feedback are numerous. For example, Hattori and Iverson (2010) assessed the effects of perceptual training (i.e., identification, discrimination, and best exemplar tasks) without CF of the English consonants (/r/ and /l/) to L1 Japanese speakers. They discovered that participants who improved in perception were not the same as the ones that improved in production.

Furthermore, Iverson et al. (2012) provided a high-variability perceptual training (HVPT) on 14 English vowels (monophthongs and diphthongs) to L1 French speakers. CF was given in the form of the word “yes!” on the screen followed by a cash register sound if the response was correct, or the word “wrong” on the screen accompanied by two tones with descending pitch if the response was incorrect. If the response was correct, participants also heard the word one more time; however, if the response was incorrect, participants heard the correct word followed by a two-series repetition of the correct word and incorrect word so that they could notice the contrast between these two words. Results reported that participants only improved their perception of these English vowels by contrast with moderate gains in production after the HVPT.

Moreover, Huensch and Tremblay (2015) studied the effect of perceptual training on the perception and production of English palatal codas (/ʃ ʧ ʤ/) in L1 Korean speakers. Participants formed an experimental group, which received perception training on the palatal codas, and a control group, which participated in an unrelated perceptual training. Results showed that the experimental group outperformed the control group in both perception and production of the codas and generalised learning to new words. However, results also indicated that there was a lack of one-to-one relationship between perception and production gains. This may suggest that even so the perception and production systems may be linked, there are certain underlying L2 representations in these systems that may be distinct.

Hwang and Lee (2015) investigated the effect of perception training on the production of English vowels and consonants in Korean speakers. Participants were trained to discriminate multiple sets of English vowel and consonant sound contrasts, and then pre- and post-tests of their recordings of those sounds were evaluated to find connections. The results revealed that the effect of perceptual training on production was not significant and that improvements in the two systems (i.e., perception and production) were not correlated. A similar study but with different results was conducted by Kangatharan et al. (2021) which explored the effect of perception training on the production of English vowels. Analyses showed that the intelligibility of the vowels treated significantly improved after the perception training suggesting that high-variability training can positively impact on the production of L2 speech.

Saito et al. (2022) examined the effects of three types auditory training: auditory-only (F2 discrimination training), phonetic-only training (English [æ] and [ʌ] identification training), and auditory-phonetic training (a combination of both) on the overall auditory processing ability of 98 L1 Japanese learners of English as L2. The results indicated that the phonetic-only group improved in identification of the trained vowels, while the auditory-only and the auditory-phonetic groups improved both their discrimination and identification skills. This suggests that training discrimination of acoustically abstract categories in terms of frequencies would help develop more domain-general abilities and therefore would promote learners' ability on a more domain-specific level (e.g., identification of English [æ] and [ʌ]).

As regards studies examining the effects of production training with feedback, these are scarce and have shown mixed results.

Kartushina et al. (2015) investigated the effect of phonetic production training with visual feedback on the perception and production of Danish speech sounds by native French speakers. Participants were assigned to an experimental group and a control group. The production training consisted of five training sessions of about 50 minutes in which participants produced four Danish vowels and received immediate visual feedback about their articulation together with the pronunciation of the vowel by a native speaker. Results revealed that one hour production training per vowel improves the production of it compared to the control group whose production performance in the post-test remained unchanged. Also, the experimental group showed improvement in perception after training; however, no correlation was established between perception and production scores. These findings may suggest that production training is effective in improving both perception and production of sounds, but this improvement may not progress in the same way within individuals.

Linebaugh and Roche (2015) studied the effects of production training of some English sounds (/æ, ʌ, g, ʤ, ɜ, ɔ/) on both their perception and production by L1 Arabic speakers. CF was not provided during training. Three experimental groups and three control groups were formed based on the pairs of sounds treated (/æ ʌ/, /g ʤ/, and /ɜ ɔ/). The experimental groups received explicit instruction on how to make the sounds and produced the sounds actively during training, whereas the control groups only engaged in listen and repeat activities and identification tasks. The results indicated that explicit articulatory training can lead to improved perceptual ability.

Furthermore, Chang (2023) conducted a study to assess the effect of production training on the perception and production of English tense-lax vowel contrasts (/i ɪ/ and /e ɛ/) with two types of CF (ultrasound biofeedback vs traditional acoustic input-only CF). To this end, Mandarin native speakers formed two experimental groups: ultrasound biofeedback and traditional OCF. After statistical analyses, both groups showed similar significant improvements in production and perception which may suggest that the benefits in production were transferred to perception. Results also showed that no correlation was found between perception and production gains indicating that they work somewhat independently.

Finally, some other studies have compared training in perception, production or a combination of both to determine which one is more favourable.

Aliaga-García and Mora (2009) compared the impact of two training methods (auditory and articulatory) on the full set of English monophthongs (/iː ɪ e ɜː æ ʌ ɑː ɒ ɔː ʊ uː/). 84 bilingual Catalan-Spanish learners of English comprised two experimental groups and one control group. The experimental groups were assigned either to an identification or an articulatory audiovisual high variability phonetic training (HVPT). The identification HVPT group was trained to identify audiovisual cues to recognise the vowel category within a vowel subset, whereas the articulatory HVPT group received training to identify relevant audiovisual cues for accurate vowel articulation. Auditory feedback was provided to both groups either to correct identification or change errors in articulation. The results indicated that both experimental groups improved their vowel perception accuracy (i.e., identification and discrimination); however, the identification group scored higher in identification and showed a lesser degree of error dispersion per vowel. A similar pattern was found for the articulatory audiovisual training. Both groups improved in their production of some vowels, but the articulatory group was more accurate in vowel production. In conclusion, the results showed that training seems to be domain-specific, that is, perception training produces higher gains in perception, and production training in production.

Alshangiti and Evans (2014) contrasted three types of English vowel training: production training (PT), which trained only production, high variability phonetic training (HVPT), which trained only perception, and hybrid training (HT), which trained both perception and production. The analysis of the results confirmed that training appears to be domain-specific: the PT participants largely improved in their production of the English vowel sounds but showed small improvement in their perception, the HVPT group improved in vowel identification but no in production, and the learners who participated in the HTP improved in both perception and production of the vowels. Wong (2013) obtained similar results in her study comparing three training designs: HVPT, explicit articulation training (ET), and HT. The HVPT group improved significantly in perception, the ET improved significantly in production, and the HT improved in both and outperformed significantly the other two training groups.

Furthermore, Sakai (2017) explored how the perception and production modalities are connected by comparing the effects of perception-only and production-only training methods on the English vowels /i/ and /ɪ/. 60 Spanish native speakers constituted four groups: perception-only group, production-only group (sound), production-only group (no sound) and control group. In this way, the perception-only group heard target phonemes but never produced the sounds, the production-only group (sound) saw real-time visual representations of spoken vowels and could only hear their own voices, the production-only group (no sound) saw real-time visual representations of spoken vowels but wore noise-cancelling headphones and listened to white noise. Results showed that perception-only training led only to significant gains in perception, and production-only training led to variable results for production (i.e., the no sound group showed no evidence of change), and medium-sized improvements in perception. These findings suggest that perception and production modalities are connected to some extent since production-training did produce some gains in perception; however, the mechanisms involved in their functioning may differ. Moreover, the results indicated that production training with access to some auditory stimuli seems to be necessary to ensure gains in that modality.

In summary, most experimental and quasi-experimental studies indicate that there is some degree of autonomy between the speech perception and production domains. In view of the fact that one of the research questions is concerned with this relationship, this topic will be explored in the following section.

2.2.2. Speech perception and production: relationship and relevant research

Most of the studies described in the previous section shed light on the intricacies of the connection between the perception and production modalities, and many of them have found that the evidence is not consistent to establish a direct relationship between these two domains.

Other studies which have explored this relationship include Shin and Iverson's (2014) which found no correlations between the perception and production domains of English vowel epenthesis among Koreans learning English.

Moreover, Kartushina and Frauenfelder (2014) assessed L1 Spanish speakers studying L2 French on their perception and production of the vowel sounds /ø œ/ and /e ɛ/. The results showed no correlation between the performance in perception and production among these sounds and no effect of perception on production in mixed-effects regression analyses. Nagle (2018) also studied the relationship between perception and production using the Spanish stops /d p b/ in English native speakers. Varied results were obtained in his analyses where some correlations were established between the perception and production of /d/ and /p/, but no significant relationships of /b/.

In addition, Schertz et al. (2015) examined Korean speakers' perception and production of stop contrasts in L1 Korean and L2 English. In general, the production data values showed to be more homogenous than the perception data.

Also, differences found in perception did not appear to be directly related to differences in production in either the L1 or the L2. In conclusion, it could be argued that it is possible to produce some sounds accurately even when perception is still inaccurate in an L1 and L2, and that both modalities do exhibit separate and autonomous mechanisms worth exploring.

The perception-production studies just described point to the idea that the connection between these two modalities is not as direct as it is usually conceived, and therefore the type of training to be used would be of crucial importance to guarantee concomitant gains in both perception and production. Additionally, individual differences (e.g., anxiety, working memory capacity, developmental and proficiency level, etc.) should be taken into consideration since they seem to play an important role in modulating the effects of both training and the provision of corrective feedback.

2.3. Individual differences: language proficiency, working memory, and their relationship with SLA and corrective feedback

Language proficiency, and working memory capacity (i.e., phonological working memory and central executive ability) were used to determine their role in modulating the effectiveness of the oral corrective feedback provided in the current study. In the following paragraphs, both variables will be described in terms of their connection with corrective feedback in SLA.

In SLA, L2 proficiency usually refers to the level of attainment in an L2 (Pawlak, 2021). This level of attainment can be described with respect to the command of a specific structure (e.g., the use of passive voice, question tags, etc.), skill (e.g., speaking, listening, etc.), or with respect to the general communicative/linguistic competence of a learner (e.g., their syntax, pronunciation, and pragmatics). Proficiency performance can be measured using a wide range of assessment instruments; for example, standardised examinations, classroom test results, course grades, performance on specific tasks, etc. It can also be operationalised in different ways (in terms of accuracy, complexity or fluency; Michel, 2017).

In the current study, proficiency will refer to the general communicate competence of the participants. Proficiency was established using a combination of methods: a) students' year at university, and b) a standardised listening proficiency test from the Oxford Placement Test 2. Using a combination of measures to account for L2 proficiency validates the results (Tremblay, 2011). In this way, the expected level of proficiency of the participants based on their current semester at university (B2-C1) was confirmed by the listening proficiency test administered as pre-test.

The role of L2 proficiency in moderating the effectiveness of different types of corrective feedback (CF) has been broadly examined. In general, studies have shown that lower-proficiency learners would benefit from more explicit types of CF (Ammar & Spada, 2006; Li, 2014; Park et al., 2016) whilst higher-proficiency ones would benefit from any type of CF (Ammar & Spada, 2006; Lin & Hedgcock, 1996; Williams, 2005). These results may suggest that explicit types of CF would be more advantageous for low-proficiency learners as it would make the corrections more evident. On the other hand, as higher-proficiency participants have already achieved some automaticity of their language processing, they require less attention resources to process meaning. This may facilitate their having more cognitive resources at their disposal to be allocated to attend to form, given that less attention resources were needed to process meaning. A probable explanation for this advantage of higher proficiency learners may be found in the Information Processing Model (McLaughlin, 1990).

In conclusion, learners with higher levels of L2 proficiency seem to have more attentional resources available to focus on form as well as meaning, which makes them better equipped to process and benefit from CF (Nassaji, 2015).

Working memory (WM) is another of the multiple factors that moderate the acquisition of an L2 (Darcy et al., 2015; Linck et al., 2014; Wen, 2015; Wen et al., 2017). According to Baddeley's model (2017), WM is a multicomponent system of limited capacity that is responsible for manipulating, storing, and retaining incoming information in the mind as we interact with the world. This model assumes a four-component system comprising the central executive, the visuospatial sketchpad, the phonological loop, and the episodic buffer. The central executive, and the phonological loop (hereafter referred to as phonological working memory (PWM)) have been identified as facilitative components for SLA (Baddeley, 2017; Cowan, 2013; Williams, 2012), and therefore selected as moderating variables in the present study. The following definitions will be used throughout the study.

PWM is a temporary storage system that holds speech and acoustic based information, which fades after two to three seconds if not rehearsed. This component has been linked to the development of first and second language acquisition (Baddeley, 2017), grammatical learning (Kapa & Colombo, 2014; Leseman & Verhagen, 2022; McCormick & Sanz, 2022; Tagarelli et al., 2015), vocabulary learning (Kapa & Colombo, 2014; Service & Simard, 2022), and reading skills (Leeser & Herman, 2022).

The central executive (CE) is described as a domain-general component that is responsible of attentional control over WM processes. Attention control refers to tasks such as focusing, dividing, and switching attention, and integrating WM with long term memory stores (Baddeley, 2003). The CE is implicated in language comprehension and production processes (Alptekin & Erçetin, 2012; Ahmadian, 2012). In addition, the CE is believed to be a good predictor for SLA success in explicit learning conditions because explicit processes are directly related to attention (DeKeyser, 2003; Linck et al., 2014).

Furthermore, some studies have established a relation between learners with higher WM capacity and their ability to notice different CF types and incorporate it in their modified output. Recast and metalinguistic OCF will be considered in the descriptions as only these two types were addressed in the present study.

Regarding the associations between working memory capacity (WMC) and the effectiveness of recast, Kim et al., (2015) found a connection between high WMC (measured with a running span test) and the ability to notice errors (English question structures) in interactive tasks. In this way, learners with high WMC noticed more of the corrected L2 forms, while learners with low WMC noticed less. Similar conclusions were reached by Sagarra (2007). In her study, English learners of Spanish received computer-delivered recasts on their errors after completing sentences with missing Spanish adjectives. The results showed that learners who had modified most of their output following feedback from pre- to post-tests had also scored high in the WM test (reading span task).

Révész (2012) explored the provision of recasts and its effects on oral and written production. WM tests (digit and non-word span tests, and a reading span test) were also used to establish any relationship between WMC and learners' performance. Two main findings were obtained: (a) recasts were considerably effective in improving oral production, but far less effective for written production; (b) participants with high executive control (i.e., central executive) ability (attentional resources and storage capacity as measured by the reading span) achieved more development on the written tests, while those with higher PWM (storage capacity only as measured by the digit and non-word span tests) showed better improvement on the oral test. The first finding is consistent with previous studies which link recasts with the development of procedural (i.e., oral production) rather than declarative knowledge (i.e., written production; Loewen &

Nabei, 2007; R. Ellis, 2007; R. Ellis et al., 2006; Sheen, 2007). The second finding suggests that participants with high central executive ability may have been more adept at allocating conscious attention to recasts and therefore developing metalinguistic, declarative knowledge from the corrections in the feedback. This ability to direct and switch their attention to metalinguistic information contained in the feedback would have allowed high execute control participants to do better on the written tasks as these tests were more declarative knowledge oriented. A similar explanation can be offered for the high PWM learners and their high scores in the oral tests. Learners with high PWM were able to maintain the information in recasts in their short-term memory for a longer period resulting in more chances for the corrections to be incorporated into their long-term memory which, in turn, would facilitate the proceduralisation of emerging L2 knowledge (N. Ellis, 2005). As the oral tests used in this study were more conductive to require the use of procedural knowledge, a high PWM would have been beneficial to do well on those tests.

Furthermore, Goo (2012) conducted a study to determine how WMC moderates the noticing of errors using recasts and metalinguistic feedback. In his study, 95 L1 Korean learners of English were divided into two experimental groups (recast group and metalinguistic group) and a control group. WMC was measured using a reading span task and an operation span task to account for complex executive control capacity and PWM capacity respectively. The results showed that recasts were as effective as metalinguistic feedback to improve the treated linguistic target. Considering that modified output opportunities were discouraged in Goo's study, the participants in the metalinguistic group had fewer chances to engage in negotiation of meaning and thus notice the corrections in the feedback. Also, as recast was provided in a controlled setting, participants in this group were able to pay more attention to the corrections in the recasts than they would have in a much less controlled classroom setting. The results also indicated that the executive control capacity only mediated the noticing of recasts but not the metalinguistic feedback. This may be explained taking into account the nature of the two types of feedback provided. On the one hand, metalinguistic feedback is explicit by nature; in this way, participants needed less cognitive resources to find the modified errors in the feedback. On the other hand, recasts, being implicit by nature, would have required participants to have more cognitive control of attentional resources to attend to form during the interactions. This would explain why the executive control capacity of WM would be more relevant to notice implicit types of feedback.

3. RESEARCH QUESTIONS, HYPOTHESES, AND OBJECTIVES

In view of the above, the present study has established the following research questions, hypotheses, and objectives.

3.1. Research questions

- How effective are recast and metalinguistic oral corrective feedback (OCF) to produce gains in the perception and production of English intonation patterns for wh-questions in L2 learners after training?
- To what extent are the perception and production modalities related?
- Which tone (fall or rise) is more challenging to perceive and produce to Chilean Spanish speakers?
- To what extent do individual differences (L2 proficiency and working memory capacity) moderate the perception and production of English intonation patterns, and the OCF provided?

3.2. Hypotheses

- Both types of OCF (recast and metalinguistic) will lead to positive gains in the perception and production of English tones for wh-questions after training. However, recast will be more beneficial than metalinguistic OCF.
- The modalities of perception and production will show some associations but will work in a relatively independent manner.

- The rising tone will be more challenging to perceive and produce for English wh-questions as this is not associated with the default pattern for this type of sentence.
- High-proficiency learners will benefit from any type of OCF, whilst lower-proficiency ones will only benefit from metalinguistic OCF.
- Phonological working memory (PWM) will relate positively with metalinguistic OCF, whilst central executive capacity (CEC) will relate positively with recast.

3.3. General objectives

- To determine the effect of recast and metalinguistic OCF to produce gains in the perception and production of intonation patterns (fall and rise) for English wh-questions in L2 learners with Spanish as L1 after training.
- To determine the role of individual differences (working memory capacity and L2 proficiency) in moderating the perception and production of English intonation patterns (fall and rise), and the types of OCF provided.

3.4. Specific objectives

- To compare which OCF (recast or metalinguistic) is more effective to treat suprasegmental errors.

- To assess the degree of connection between the modalities of perception and production.
- To determine which tone (fall or rise) is more challenging for Chilean Spanish speakers.
- To determine the extent that L2 proficiency is related to the development of the perception and production modalities and the noticing of OCF.
- To determine the extent that working memory capacity is related to the development of the perception and production modalities and the noticing of OCF.
- To establish associations between the components of working memory (PWM and CEC) and the OCF provided (recast and metalinguistic).

4. METHODS

4.1. Design

This study used a quasi-experimental design. In this way, the study involved the use of pre- and post-tests, randomisation of participants into two experimental groups, but lacked a control group. Quasi-experimental studies "are a subtype of non-experiments that attempt to mimic randomized, true experiments in rigor… and do not require a true control group but may include a comparison group. A comparison group is an additional experimental group that receives a different experimental treatment" (Rogers & Révész, 2020, p. 134).

Selecting this research design was motivated by the research questions and objectives of the current study as these aimed to compare the effects of two types of OCF treatments (independent variable) mediated by dependent and moderating variables.

4.1.1. Research purposes

This study was correlational in nature. Given the number of different variables, a correlational design was used to explore their associations. The purpose of correlational research is to establish the extent to which two or more variables are related focusing on the importance of this relationship rather than on a cause-effect association (Fouché & De Vos, 2011). This type of research design has an explanatory component since determining how one variable may influence another helps explain and understand a particular phenomenon. However, this explanatory component is said to be partial as there may be other variables related to the phenomenon being considered (Hernández-Sampieri et al., 2010).

4.1.2. Variables

The two experimental groups received different treatments in the form of OCF (recast vs metalinguistic) on the perception and production of intonation patterns of wh-questions. The treatments (OCF) constituted independent variables as they were expected to produce some variation or change in the dependent variables (i.e., the perception and production of intonation patterns of wh-questions). Therefore, the dependent variable is the variable that can be influenced by the other variables (Loewen & Plonsky, 2016).

Pre- and post-tests were used to measure the changes in perception and production of intonation patterns of wh-questions after treatment. Furthermore, other tests were given to the participants to inform about their language proficiency and working memory capacity (i.e., PWM and CE). The measures from these tests were used as moderating variables as they mediate the relationship between independent and dependent variables. In other words, language proficiency and working memory capacity of the participants may influence the extent to which these benefit from feedback (Rogers & Révész, 2020).

4.2. Participants

40 L2-English learners with Chilean Spanish as L1 participated in the current study. Participants were third-year students in the course "Entonación de la Lengua Inglesa" [English intonation] from the English Teaching Programme at Universidad de Concepción. Their ages ranged between 21 and 22 years of age. All the linguistics and literature-related courses in the programme were taught in English. Therefore, all the participants had received quantitively similar amount of English language exposure before the tests and training sessions were administered. Furthermore, at the moment of carrying out this study, all participants had previously attended pronunciation workshops in general segmental phonetics during their first year, and were currently receiving instruction in suprasegmental phonetics.

In relation to the expected level of proficiency of the participants, this can be described as B2 (18 participants) or C1 (22 participants) according to the Common European Framework for Languages (2020). The listening section from the Oxford Placement Test 2 was used to measure their proficiency levels (Allan, 1992). All of the participants involved in this study voluntarily agreed to participate and reported not having any speaking or hearing impairment.

Controlling the participants' characteristics: (a) language background (Spanish as L1), (b) similar language experience with the L2 (third-year English Teaching students at the same university who had received previous instruction on segmental phonetics and were currently taking a course on English intonation), and (c) similar L2 proficiency levels (B2-C1) ensure internal validity (Mackey & Gass, 2005).

4.3. Materials

4.3.1. Perception tests

The perception pre- and post-tests consisted of an identification task designed to measure the ability of participants to perceive falling and rising tones of wh-questions of English. Each test included 20 wh-questions with a falling intonation and 20 wh-questions with a rising intonation presented in a random fashion. All wh-questions used at pre- and post-test were different but followed a similar pattern; they were either subject questions (i.e., wh-question word + auxiliary verb + subject) or object questions (i.e., wh-question word + auxiliary verb + subject + main verb + complement) (see Appendix A). Furthermore, the vocabulary selected for the tests were A2-B1 level frequent words either taken from textbooks or online.

The stimuli were recorded using a laptop, noise-cancelling headphones with microphone (circumaural; 15 Hz – 25000 Hz) and the Audacity software (version 3.1.3) at a sampling rate of 44100 Hz in WAV format. One male English native speaker was asked to make the recordings of the stimuli.

Internal validity of the perception tests was achieved in the following way.

The perception tests were carefully designed to ensure representativeness of the measurements obtained (content validity). In this way, to measure perception, the two most frequent intonation patterns of English (fall and rise) were selected (Roach, 2009; Wells, 2009). Furthermore, the selection of wh-questions followed the structure of either a subject or object question, which are the two most common and frequent pattern for questions. In addition, perception pre- and post-tests were directly connected to the constructs to be measured (construct validity), as they were identification tasks of intonation patterns for English wh-questions. Equivalence in terms of length (40 stimuli each), type of wh-questions (subject or object questions), number of intonation patterns used (20 falling tones, 20 rising tones each), and vocabulary level (A2-B1) was also controlled to ensure internal validity.

External validity of the perception tests was ensured using random sampling. In this way, the wh-questions for the perception pre- and post-tests were presented randomly by the TP platform (Rauber et al., 2012).

4.3.2. Production tests

The production pre- and post-tests were designed to measure the ability of the participants to produce English wh-questions with falling and rising tones. In these tests, participants had to read out loud 20 wh-questions with a falling intonation, and 20 wh-questions with a rising intonation. All wh-questions used at pre- and post-tests were different but followed a similar pattern; they were either subject questions or object questions. The 80 wh-questions were randomised to create two fixed lists (see Appendix B). Also, each test included five extra questions for practice.

The 80 wh-questions were taken from English textbooks that cover the levels A2 and B1.

The same internal validity criteria used for the perception pre- and post-tests were applied for the production pre- and post-test. As for external validity, the randomisation of the wh-questions to create two fixed lists was employed.

4.3.3. Listening proficiency test

The listening part of the Oxford Placement Test 2 (Allan, 1992) was used to confirm the participants' level of English proficiency. The criteria used to select this test were as follows: (a) it is a validated instrument, (b) it reports results from A2 to C1 (Oxford University Press, n.d.), (c) the results obtained with this test have shown to be normally distributed (Zoghlami, 2014), and (d) its expected completion time is of approximately 12 minutes.

The test consisted of 100 gapped sentences followed by two possible alternatives. The two options could either be minimal pairs or have a different stress pattern. For instance, "is it ready for typing/taping yet?" or "do you know if this text is copyright/copied right?"

The paper-based version of the test was input into a Google Forms (see Appendix C) for easier administration and grading. Each correct answer was given one point as score, and then transformed into percentage.

This test was chosen since the use of validated or standard tests is also a method to ensure criterion-related validity (Mackey & Gass, 2005), which is related to the extent to which tests used in a research study are comparable to other well-established tests of the construct in question.

4.3.4. Working memory span tasks

The phonological working memory (PWM) and central executive (CE) have been shown to be most directly relevant to first and second language learning and processing (Wen, 2015). For this reason, a non-word repetition task (NWRT) and a listening span task (LST) were used to measure the PWM and CE respectively. The results were used to account for the complexity of this system and their relations with the perception and production of intonation patterns for English wh-questions and the effects of OCF.

Furthermore, Spanish phonotactics were used for both tasks given that working memory ability is dependent on our previous knowledge of the phonological regularities of the L2 (Cheung, 1996; Masoura & Gathercole, 1999), therefore poor results could be attributable to L2 competence rather than to working memory capacity and ability.

4.3.4.1. Non-word repetition task (NWRT)

To measure PWM, a NWRT was administered. 36 Spanish non-words were used in this task taken from the test created by Oportus and Ferreira (2015) who used a 240 non-word list of Brea-Spahn's study (2009) to develop a validated NWRT for Chilean Spanish speakers. Oportus and Ferreira (2015) chose these 36 non-words using four linguistic criteria: phonotactic probability, degree of similarity to real Spanish words, syllabic length, and accent pattern. In the end, 12 non-words of different number of syllables (two, three, and four) formed the final group of 36 words (see Appendix D). Internal validity of this test was achieved by means of competent judges (Hernández Nieto, 2011) who assessed the 36 non-words.

New audios were created for the 36 non-words recorded by a male native speaker of Chilean Spanish following the same characteristics of the original test.

The scoring system used for this task follows Kaushanskaya and Yoo's method (2013), where the score is obtained by calculating the proportion between the correctly produced phonemes and the total number of phonemes in each non-word. For example, with the non-word oridá /o ri 'ða/, if the participant produces /o re 'ða/, only the phoneme /e/ is considered incorrect, and therefore the proportional score of the correct phonemes is 0.8. By contrast, if all the segments are correct, the proportional score is 1 (Archibald & Gathercole, 2006).

4.3.4.2. Listening span task (LST)

To measure the central executive (CE) capacity of WM, a Chilean Spanish adaptation by Véliz et al., (2022) of Zychowicz et al.'s (2017) LST was used.

The task consisted of 54 sentences divided into 9 sets. The number of sentences in each set gradually increased from 2 to 10 (e.g., set one contains two sentences, set two contains three sentences, etc.). The sentences were recorded using the Audacity software (version 3.1.3) by two Chilean Spanish native speakers, one female and one male. All items were grammatically correct, complex sentences of approximately 10 words in length. 27 sentences were lexically correct and the other 27 were lexically altered so that they did not make sense in everyday life. For instance, the sentence: *sonó la alarma y todos corrieron rápidamente hacia la plaza* makes sense. On the other hand, the sentence: *el lápiz me dijo que se compró una casa* is illogical as pencils do not speak. Furthermore, each sentence-final word was a common noun.

The task required participants to simultaneously judge whether each sentence made sense (processing capacity) as well as recall the last word of each sentence (memorisation) for subsequent recollection.

Following Zychowicz et al.'s (2017) scoring procedure, partial scoring was used to assign scores to the participants' answers. In this way, the final score of the LSP of each participant was the total number of correctly remembered words in all sets. The processing task (i.e., judging the logic of sentences) only served as a distractor, therefore its results were not considered in calculating the final score. However, if a participant obtained a score below 80% of correct answers in the processing task, they were excluded from the sample given that they lacked concentration on the task.

Internal validity of this test was achieved by means of competent judges (Hernández Nieto, 2011) who assessed the sentences on their semantic sense (sensical/non-sensical) and the semantic relationship of the sentences in each group. In addition, for each group of sentences, judges indicated how they were thematically consistent with each other. Judges rated the sentences in terms of these dimensions using a scale ranging from 1 (strongly disagree) to 5 (strongly agree). To analyse how judges rated this test, two criteria were used. First, their averages on each dimension were determined. Second, their agreement on the ratings for each dimension across groups of sentences were determined using the Kendall's coefficient of concordance (Wt).

4.3.5. Training material

The audio stimuli for the perception exercises were taken from the 80 wh-questions used in a perception test designed by Naranjo (2020). The perception test used 10 wh-questions with falling intonation, and 10 wh-questions with rising intonation for each session. These questions were selected and adapted from English textbooks (A2 to B1 level). These stimuli were different from the ones used in the pre- and post-test.

For the production practice, participants had to record 20 wh-questions (10 with falling tone, 10 with rising tone) for each session which were presented on a randomised list in a word document (see Appendix E). These stimuli were different from the ones used in the pre- and post-test.

The perception and production training sessions were carefully designed to ensure representativeness of the measurements obtained (content validity). Therefore, the wh-questions used contained the two most frequent intonation patterns (fall and rise) of English (Roach, 2009; Wells, 2009), and followed the structure of either a subject or object question, which are the two most common and frequent patterns for this type of question. Furthermore, the trainings for both perception and production were directly connected to the constructs to be measured (construct validity) as required participants to engage in shorter identification or production tasks of similar characteristics to the pre- and post-tests.

4.4 Procedure

4.4.1. Recruitment and randomisation of participants

40 participants from the "Entonación de la Lengua Inglesa" [English intonation] course on the English Teaching Programme at Universidad de Concepción voluntarily agreed to participate in this study. All participants were randomly assigned to two experimental groups (EGs) as the treatment given to each group differed in the type of OCF provided (metalinguistic vs recast). Therefore, 20

participants formed the metalinguistic experimental group (MEG), and 20 the recast experimental group (REG). This number of participants per experimental group conforms with the suggested sample size (≥15) of participants for experimental and quasi-experimental studies conducted in controlled settings to guarantee external validity (Fraenkel & Wallen, 2012). As in the present study 20 participants were used per experimental group, the results obtained may have the statistical power to be generalised.

4.4.2. Listening proficiency test administration

After the recruitment and randomisation of the participants, the listening proficiency test was administered. Participants were given the listening proficiency test in one session at the beginning of one of their English intonation classes. A phonetics laboratory at Universidad de Concepción was used to give the participants the test.

First, students received instructions on what they were going to do before taking the test. After that, students were asked to check that their desktop computers, internet connection and noise-cancelling headsets were working properly. Next, the researcher sent the link to the test through email to all the participants.

Once students had clicked on the link, a Google Form of the test was opened. The first page showed the same instructions given by the researcher at the beginning, followed by two example exercises for them to get familiar with the task. Then, they clicked on “next” to go the next page and start with the test. There, participants had to listen to 100 gapped sentences and then choose between two words or phrases as possible answers. The two options could either be minimal pairs or have a different stress pattern.

After about 12 minutes, the test ended. The participants' scores were shown to them at the end of the test. Furthermore, their responses were automatically checked, saved and sent to the researcher's email address once they had finished.

4.4.3. Non-word repetition task administration

A week later, the NWRT was given to all participants individually using a desktop computer running on Windows 10, and a noise-cancelling headset with microphone (circumaural; 15 Hz – 25000 Hz) at a phonetics laboratory at Universidad de Concepción. The tools Microsoft PowerPoint and Audacity were used to show the instructions and play the stimuli and record the participants' repetitions, respectively. Prior to the start of the experiment, participants were informed on the procedure of the task and listened to four trial stimuli to become familiar with it and adjust the volume if necessary. After this, the NWRT began and participants' recordings as well. All instructions were presented in Spanish. For example, participants read 'pseudopalabra 1' along with the instruction 'escuche atentamente y repita inmediatamente después del beep'. Then, a non-word was played for participants to listen to and repeat as close to the original as possible. Each stimulus lasted 2000 ms, and 4000 ms were allowed for each repetition. After each repetition, the slide changed to show the same instructions and play a different non-word for participants to repeat. The order of presentation of the 36 non-words was the same for each of the participants. This order was randomly assigned using random sampling. The expected completion time of each NWRT was eight minutes.

As described above, in order to ensure the validity and reliability of the administration of the test, trial stimuli were presented to the participants to help them become familiar with the task (Mackey & Gass, 2005), stimuli were presented at a regular pace (Williams, 2012), and repetition of the stimulus was requested immediately after listening to it (Juffs y Harrington, 2011).

4.4.4. Pre-tests administration

Two pre-tests were administered to collect information about their baseline ability to perceive and produce intonation patterns (fall and rise) of English wh-questions: the production pre-test, and the perception pre-test.

4.4.4.1. Production pre-test

This test was given to participants in the individual sessions immediately after the NWRT at a phonetics laboratory at Universidad de Concepción. A desktop computer running on Windows 10, and a noise-cancelling headset with microphone (circumaural; 15 Hz – 25000 Hz) were used for this test. Furthermore, Microsoft PowerPoint (version 2307) and Audacity (version 3.1.3) were used to show the wh-questions and record the answers of the participants respectively. Before starting the test, instructions in English were presented to the participants, and five questions were used for practice. The instructions read as follows: “production test. This test was designed to measure your ability to produce questions with different intonations in English. In the following slides you will see questions followed by the symbols (↘) which means a falling tone, and (↗) which means a rising tone. For example: who are you calling? (↘) vs who are you calling? (↗)”. After that, five slides with five different questions were used to practise. The instruction there read: “read the following question out loud using the intonation in parentheses. You can read the question only once.” Following the practice exercises, the test began. The test took between 6 to 8 minutes. After the completion of the test, participants were asked to leave the room for the results to be collected by the researcher.

4.4.4.2. Perception pre-test

In the course of the same week that the NWRT and production pre-test took place, the perception test was administered.

To give this test, a phonetics laboratory at Universidad de Concepción was used. The test was administered in two separate sessions with 20 participants per group. The perception test was presented using the TP platform (Rauber et al., 2012) (see Appendix F). The order of 40 wh-questions was randomised by the software to eliminate response bias. For each stimulus, there were two alternatives to choose from. The participants had to listen to each stimulus and then click on the tone they believed they heard: falling tone or rising tone. The following instructions were written in English in the practice phase as well as the testing phase: “you will listen to questions with different intonation. After you listen to the questions, choose the tone you think you heard (falling tone or rising tone).” The test took between 8 to 12 minutes. After the completion of the test, participants were asked to leave the room for the results to be collected by the researcher.

4.4.5. Training

After the pre-tests, the participants started with the training.

Four training sessions were held to teach and practise the perception and production of the two intonation patterns of English wh-questions selected for this study (fall and rise). Each session lasted approximately 25 minutes and extended over a period of four weeks. The four training sessions were delivered during class time using the platform CANVAS LMS (2022). Each session followed the same structure: (a) explicit instruction on the two intonation patterns followed by example questions, (b) perception exercises of the two patterns, and (c) production exercises of the two patterns. A description of one of the training sessions is provided in the paragraph below.

Participants logged in to CANVAS LMS to get access to the training session. First, they watched a 2-minute video explanation of the pragmatic uses of the two intonation patterns of English wh-questions followed by 10 example questions (five per pattern). Participants were encouraged to repeat the example sentences to practise the intonation patterns. After the video, they listened to 20 audio stimuli and chose between the options “default (fall)” or “encouraging rise”. There were 10 falling and 10 rising tones in random order for each perception task. Finally, the participants recorded 10 wh-questions using the default tone (falling), and 10 using the encouraging rise. Participants were asked to practise for five minutes before recording themselves using Audacity. The recordings had to be uploaded to CANVAS LMS.

4.4.6. Oral corrective feedback provision

A week after the completion of a training session, oral corrective feedback (OCF) was provided.

Participants received individual recorded OCF on their productions of erroneous intonation patterns using the CANVAS LMS platform. The type of feedback that they received varied based on the OCF group the participant belonged to. In this way, a participant in the MEG (metalinguistic experimental group) received metalinguistic corrective feedback, whereas a participant in the REG (recast experimental group) received recast. Scripted OCF was provided to each OCF group. Examples of each scripted OCF are as follows:

MEG: “you will listen to the wh-questions you produced with the incorrect intonation pattern followed by a short explanation of the error. Then, you will listen to the correct way to produce it. Listen carefully. 1: you said: *where does she ↗live?* Here you used a rising tone which we use to show interest in wh-questions. You should have used a falling tone which we use to request information. Your question should have sounded like this. Repeat after you listen: *what's your ↘name?*

REG: "in this audio you will listen to the corrected intonation patterns you were expected to use for the wh-questions. Each wh-question has been numbered to help you find your mistake in the correction sheet attached to this e-mail. As you listen to the audio, use the correction sheet to check the incorrect intonation pattern you used in your recording. Listen carefully. 2. *what do you do on ↘Mondays?* (One second apart) 5. *What did you ↗say to him?* This is the end of your feedback. To have better results it is highly recommended that you also listen to your own incorrect productions and compare them with corrections in this recording.

To make sure that participants actually listened to their OCF, five minutes were allotted for this at the beginning of the following training session.

4.4.7. Post-tests administration

Subsequent to the training with OCF, perception and production post-tests were administered to determine the effectiveness of the treatment given.

Both post-tests followed the same procedure as the pre-tests but employed different wh-questions.

4.4.8. Listening span task administration

As a consequence of the listening span task requiring more time than expected to be available for use, this task was given to a subset of only 20 participants of the experimental groups as the other participants were not available at the time.

The LST was given to the participants in two separate sessions with 10 participants per group. A desktop computer running on Microsoft Windows 10, and a noise-cancelling headset was used for the completion of this task. The software platform PSYCHOJS version 2022.2.4 run by Pavlovia was used to implement the LST.

Before participants started with the task, they were informed of its content and what they were expected to do using a PowerPoint presentation. After that, each participant received an email with a unique code and a link to the LST. Participants had to click on the link and then use the unique code to start the task.

The LST started with information about the objectives of the task, and a description of each phase. The first phase consisted of a trial. Two trial sets of sentences (a set of three sentences, and a set of four sentences) were used to ensure that participants became familiar with the two tasks (i.e., judging the semantic sense after each sentence, and recalling the last word of each sentence). After the trial phase, the actual LST started. In this way, participants had to listen and answer if the sentence made sense to them by clicking on “yes” or “no”. Following the end of each set of sentences, participants were asked to recall and write the last word of each sentence; thus for set one, participants had to recall and write two words, for set two, they had to recall and write three words, etc. A beep was used to indicate that participants had to recollect the last word of each sentence by typing them. The order of recall was free, that is, they could write the words in any order. Participants were instructed to answer as quickly as possible; they had to click on a “continue” button only once they had heard the complete audio of a sentence. Then, they would hear a beep indicating that a new set was about to begin. The administration of the test took about 20 minutes.

4.4.9. Data analysis

The data collected from the tests in the pre- and post-tests were used for the statistical analysis. The variables obtained were a) perception measure (perception of the intonation patterns of wh-questions), b) production measure (production of the intonation patterns of wh-questions), c) listening proficiency measure, d) phonological working memory (PWM; measured with a non-word repetition task) and e) central executive capacity (CE; measured with a listening span task). Furthermore, OCF in the form of f) recast and g) metalinguistic also constituted variables.

In this way, the dependent variables were the perception and production of the intonation patterns of wh-questions (a & b); the independent variables included recast and metalinguistic OCF (f & g); and the listening proficiency (c), PWM and CE (d & e) constituted moderating variables.

Data was analysed using a a between-within analysis of variance (ANOVA), and correlation analyses between the independent, dependent and moderating variables were also explored.

5. RESULTS

The data was gathered from the participants' results obtained in the pre-tests and post-tests measurements of the perception and production of intonation patterns (fall and rise) of English wh-questions, the listening proficiency test (LPT), and the working memory tests (a non-word repetition task and a listening span task). To run all the statistical analysis, the R Statistical Software (Version 4.1.3; R Core Team, 2022) was used.

The scores of perception and production across feedback groups (MEG vs. REG) and time (T1 vs. T2) were compared, which are shown in Table 5.1 and Figure 5.1. Specifically, a between-within analysis of variance (ANOVA) was conducted with perception and production as the dependent variables in separate models. Results from these analyses indicated that scores in perception did not differ when comparing feedback groups, F (1, 38) = 0.01, p = .928, $\eta^2 < .01$, or time, F (1, 38) = 3.28, p = .078, $\eta^2 = .03$. The interaction between feedback group and time was not statistically significant, F (1, 38) = 1.49, p = .230, $\eta^2 = .01$, indicating that there were no differences when comparing T1 vs. T2 in any of the two feedback groups.

Results for production indicated a different pattern of results. There were no differences between feedback groups, F (1, 38) = 0.63, p = .431, $\eta^2 = .01$. However, there were significant differences between T1 (M = 71.88, SD = 17.40) and T2 (M = 89.44, SD = 12.92), F (1, 38) = 41.15, $p < .001$, $\eta^2 < .26$. The interaction between feedback group and time was not statistically significant, F (1,

38) = 1.46, p = .234, η^2 = .01, suggesting that participants in both MEG and REG similarly increased their scores in production from T1 to T2.

Figure 5.1

Results of pre (T1) and post (T2) tests for perception and production of English wh-questions per feedback group (MEG, REG)

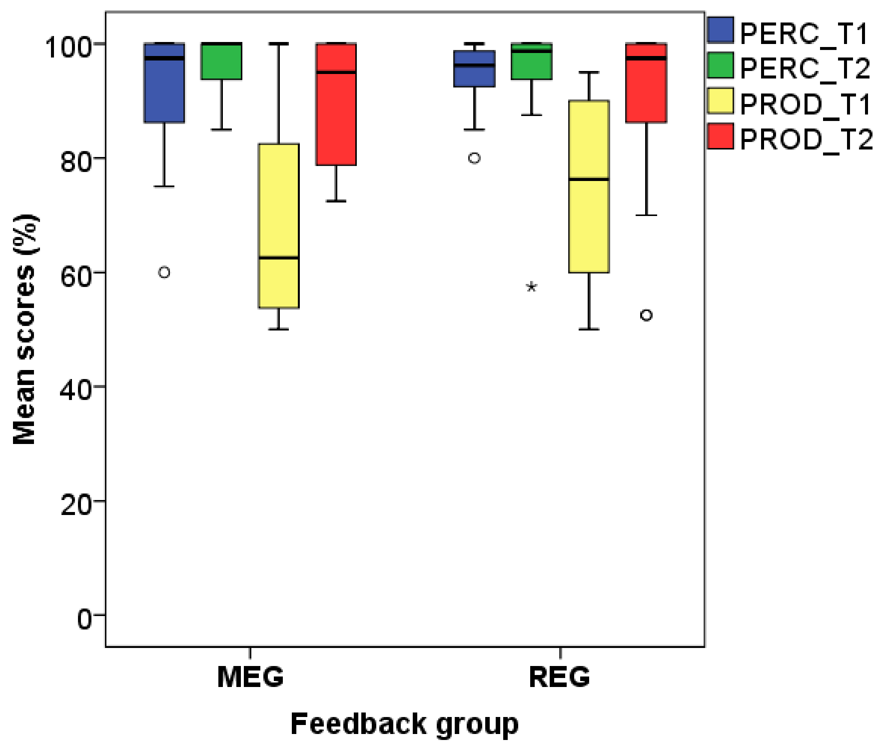


Table 5.1

Perception and production mean and standard deviation scores by feedback group (MEG, REG) and time (T1, T2)

	T1				T2			
	MEG		REG		MEG		REG	
	M	SD	M	SD	M	SD	M	SD
Perception	92.40	10.60	94.40	5.67	96.90	4.51	95.20	9.66
Production	68.60	18.10	75.10	16.50	89.50	10.70	89.40	15.10

The differences between T2 and T1 in perception and production for each participant were computed. Then, proficiency (measured through LPT), phonological working memory (measured through NWRT), and central executive ability (measured through LST) were explored to find whether they were associated with differences in perception and production over time. Also, these associations were examined to discover if they varied across feedback groups using linear regression models including the interaction between feedback group and the variables measurig proficiency, phonological working memory, and central executive ability in different models.

Results from these analyses indicated that proficiency (LPT), r (38) = -.01, p = .944, phonological working memory (NWRT), r (38) = -.23, p = .150, and central executive ability (LST), r (18) = .26, p = .275, were not associated with differences in perception over time. These nonsignificant associations were observed in both MEG and REG feedback groups, as the interactions between feedback group and proficiency, $b = 0.09$, $\beta = 0.39$, $s.e. = 0.48$, $t = 0.19$, $p = .851$, phonological working memory, $b = 2.48$, $\beta = 13.13$, $s.e. = 2.15$, $t = 1.16$, $p = .255$, and central executive ability, $b = 0.49$, $\beta = 1.41$, $s.e. = 0.73$, $t = 0.67$, $p = .510$, were not significant.

Similar results were found for production. Results indicated that proficiency (LPT), r (38) = -.16, $p = .329$, phonological working memory (NWRT), r (38) = .10, $p = .531$, and central executive ability (LST), r (18) = -.13, $p = .595$, were not associated with differences in production over time. These nonsignificant associations were observed in both MEG and REG feedback groups, as the interactions between feedback group and proficiency, $b = -0.79$, $\beta = -1.82$, $s.e. = 0.86$, $t = -0.91$, $p = .367$, phonological working memory, $b = -1.43$, $\beta = -4.09$, $s.e. = 4.08$, $t = -0.35$, $p = .729$, and central executive ability, $b = 1.75$, $\beta = 3.36$, $s.e. = 1.09$, $t = 1.61$, $p = .127$, were not significant.

The scores in the outcome test across feedback group (MEG, REG), time (T1, T2), test (perception, production), and tone (rise, fall) using between-within analysis of variance (ANOVA) were compared. The means across these factors are shown in Table 5.2.

Table 5.2

Mean scores across feedback group (MEG, REG), time (T1, T2), test (perception, production), and tone (rise, fall)

			MEG		REG	
			M	SD	M	SD
Perception	T1	Rise	93.50	11.40	96.80	4.67
	T1	Fall	91.20	15.40	92.00	9.92
	T2	Rise	96.20	6.26	96.20	6.04
	T2	Fall	97.50	3.44	94.20	15.70
Production	T1	Rise	57.80	37.50	69.00	31.90
	T1	Fall	79.50	27.30	81.20	22.00
	T2	Rise	85.80	19.80	86.00	20.20
	T2	Fall	93.20	10.80	92.80	13.60

Results from the between-within ANOVA is shown in Table 5.3. These results indicated that only time, test, tone, the interaction between time and test, and the interaction between test and tone predicted the scores in the tests. In other words, scores, on average, were significantly higher at T2 (M = 92.75, SD = 7.74) when compared to T1 (M = 82.63, SD = 10.92). In addition, scores, on average, were significantly higher for perception (M = 94.72, SD = 6.46) when compared with production (M = 80.66, SD = 12.61). Scores, on average, were also significantly higher in the fall tone (M = 90.22, SD = 9.81) when compared with the rise tone (M = 85.16, SD = 12.86). The interaction between time and test indicated that among perception there were no significant differences between T1 (M = 93.40, *SD* = 11.00) and T2 (*M* = 96.10, *SD* = 9.03), *t* (79) = -1.93, *p* = .057; but among production, there were significant differences when comparing T1 (*M* = 71.90, *SD* = 31.10) with T2 (*M* = 89.4, *SD* = 16.70), *t* (79) = -5.19, *p* < .001. Finally, the interaction between test and tone *(rise, fall)* indicated that among perception, there were no significant differences between fall (*M* = 93.80, *SD* = 12.20) and rise (*M* = 95.70, *SD* = 7.49), *t* (79) = -1.42, *p* = .160; but among production, there were significant differences between fall (*M* = 86.70, *SD* = 20.20) and rise (*M* = 74.60, *SD* = 30.30), *t* (79) = 2.86, *p* = .005.

Table 5.3

Analysis of variance

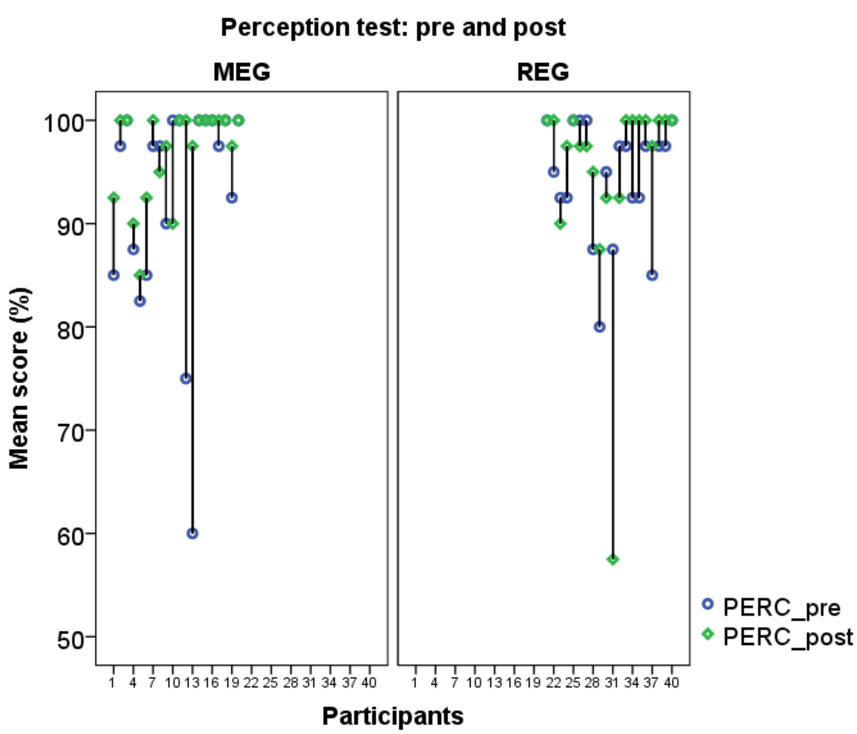
Predictor	Effect
Feedback group (MEG vs. REG)	F (1, 38) = 0.40, p = .531, $\eta^2 < .01$
Time (T1 vs. T2)	F (1, 38) = 56.23, p < .001, $\eta^2 = .07$
Test (Perception vs. Production)	F (1, 38) = 65.04, p < .001, $\eta^2 = .13$
Tone (Rise vs. Fall)	F (1, 38) = 4.19, p = .048, $\eta^2 = .02$
Feedback group × Time	F (1, 38) = 3.60, p = .065, $\eta^2 = .01$
Feedback group × Test	F (1, 38) = 0.74, p = .395, $\eta^2 < .01$
Feedback group × Tone	F (1, 38) = 0.65, p = .424, $\eta^2 < .01$
Time × Test	F (1, 38) = 18.28, p < .001, $\eta^2 = .04$
Time × Tone	F (1, 38) = 0.73, p = .399, $\eta^2 < .01$
Test × Tone	F (1, 38) = 9.33, p = .004, $\eta^2 = .01$
Feedback group × Time × Test	F (1, 38) = 0.19, p = .669, $\eta^2 < .01$
Feedback group × Time × Tone	F (1, 38) = 0.26, p = .616, $\eta^2 < .01$
Feedback group × Test × Tone	F (1, 38) = 0.06, p = .807, $\eta^2 < .01$

Time × Test × Tone	F (1, 38) = 2.31, p = .137, η^2 = .01
Feedback group × Time × Test × Tone	F (1, 38) = 0.31, p = .582, η^2 < .01

5.1. Individual differences

Figure 5.2

Individual results of pre (T1) and post (T2) tests for perception of English wh-questions per feedback group (MEG, REG)

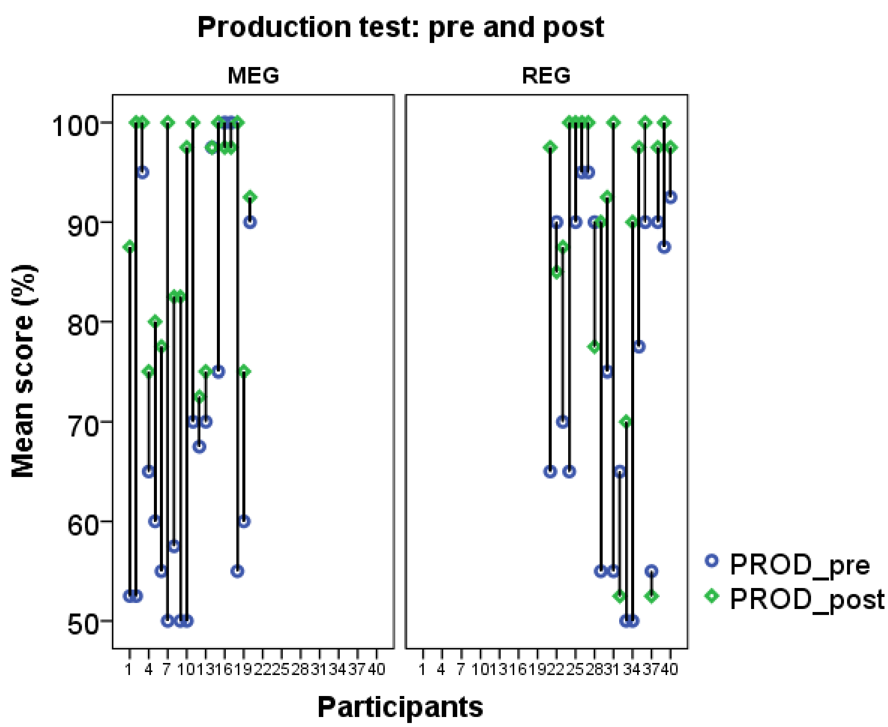


The drop-line graph shows the scores for perception of English wh-questions at pre- and post-test for individual participants. It is clear from the graph that more than half of the participants showed scores that ranged between 90 and 100%

correct already before the training. Thus, there was very little room for improvement. However, there were some cases who did show improvement after training (for example, between participants 10 and 13 in MEG). On the other hand, there were a couple of cases that obtained lower scores in the post-test (for example, participant 31 in REG). This illustrates individual differences in performance which cannot be observed if we only look at the overall means.

Figure 5.3

Individual results of pre (T1) and post (T2) tests for production of English wh-questions per feedback group (MEG, REG)



The drop-line graph shows the scores for the production of English wh-questions at pre- and post-test for individual participants. It is clear from the graph that more than half of the participants showed improvement from pre- to post-test and that their scores were considerably lower than those for perception before training. It is worth noticing that participants with poorer results before training showed larger amounts of improvement at post-test. Also, there were some cases who did not show much improvement at post-test as their production accuracy was already high before training (for example, participants 3 and 20 in MEG, or between participants 25 and 27 in REG). On the other hand, there were a couple of cases that obtained lower scores in the post-test (for example, participants 28 and 32 in REG). This may suggest that training with OCF benefitted more those learners who were less accurate and had more room for improvement.

6. DISCUSSION

The main objective of this study was to determine the effects of recast and metalinguistic oral corrective feedback (OCF) on the perception and production of intonation patterns (fall and rise) of English wh-questions in L2 learners with Chilean Spanish as L1. To achieve this, two experimental groups were formed to ensure equal provision of the two types of corrective feedback on production alone. In this way, the recast group (REG) received recast (repetition of the utterance in its correct version), and the metalinguistic group (MEG) received metalinguistic oral corrective feedback in individual audio files. OCF was provided to all participants in four training sessions. These sessions comprised the same explanatory videos and practice exercises but differed in the type of OCF received. Corrective feedback on perception was only provided through automatic “correct” or “incorrect” messages during the perception practice exercises; therefore, no individualised OCF was given on this modality.

The first research question concerned with the main objective was “how effective are recast and metalinguistic oral corrective feedback (OCF) to produce gains in the perception and production of English intonation patterns for wh-questions in L2 learners after training?” To answer this question, the changes of each speech modality (i.e., perception and production) from pre- to post-tests in each experimental group (i.e., recast and metalinguistic) were compared.

Regarding changes in production of the intonation patterns, the results showed that both recast and metalinguistic OCF were significantly beneficial after training (T1: M = 71.90, SD = 31.10 and T2: M = 89.4, SD = 16.70). This may suggest that both types of OCF can contribute to improving L2 learners' suprasegmental features in a similar way. Also, as OCF was only given on production during training, this modality was expected to have substantial gains in accuracy. These results are in line with previous research in which production training along with OCF produced significant acoustic and articulatory improvements (Dowd et al., 1998; Saito, 2013).

As regards the difference in production gains between the MEG (T1: M = 68.60, SD = 18.10; T2: M = 89.50, SD = 10.70) and REG (T1: M = 75.10, SD = 16.50; T2: M = 89.40, SD = 15.10), the former group showed slightly higher improvement scores than the latter from T1 to T2 (pre- to post-test); however, this difference was not found to be significant. Some studies have indicated that explicit corrective feedback strategies would be more effective than implicit ones to correct pronunciation errors as they are more likely to be noticed by the learner (Gooch et al., 2016; Jalal & Alahmed, 2022; Mackey et al., 2007). As a result, learners would have more chances to modify their output which, in turn, would lead to more chances of uptake; a process that has been deemed necessary for acquisition and learning gains (Loewen, 2004; McDonough, 2005). In addition, as the recast provided in this study was on the explicit end of the scale (i.e., participants knew they were being corrected), similar results to the metalinguistic feedback group may be explained. Furthermore, the positive gains in production after training are consistent with the results found in a meta-analysis where explicit corrective feedback led to greater improvement on immediate post-tests, while implicit corrective feedback led to greater gains on delayed post-tests (Li, 2010). Further research is then needed to determine whether recast or metalinguistic OCF would have longer-time effects in the production of intonation patterns of English wh-questions.

In regard to the effects in the perception of the rising and falling intonation patterns of English wh-questions by the participants, the results showed that neither recast nor metalinguistic OCF had a positive effect in significantly decreasing the errors on perception after training. The reason behind this phenomenon may lie in the fact that participants in both experimental groups were remarkably accurate (ceiling effect) in perceiving the falling and rising tones of English wh-questions in the pre-test, which may be interpreted as participants having less room for improvement in the post-test (T1: M = 93.40, SD = 11.00; T2: M = 96.10, SD = 9.03). Furthermore, as individualised feedback was exclusively given on production, this may have led to gains targeting only this modality. This may suggest that training is modality-specific since in the current study, significant improvement was only found in production of the trained pattern (Alshangiti & Evans, 2014; Chang, 2023; Hwang & Lee, 2015; Kartushina & Frauenfelder, 2014; Wong, 2013). Additionally, some studies where production has been trained showed little evidence of transfer of learning from production to perception (Hattori

& Iverson, 2010), or no correlation between the two modalities (Chang, 2023; Kartushina et al., 2015). Along the same line, a meta-analysis into the connections between perception and production also suggests that production training is not always conducive for gains in perception and vice versa, with mixed results found between the connection between perception and production and their interconnected gains (Sakai & Moorman, 2018).

The first hypothesis is then partially confirmed by these results as both recast and metalinguistic OCF were only beneficial to improve the production of the intonation patterns after training. Given that perception ability of the participants was highly accurate before training (ceiling effect), no improvement in this modality was expected. Furthermore, results showed that recast did not lead to greater gains than metalinguistic OCF to correct pronunciation errors rejecting the second part of the first hypothesis. This may be explained by the fact that the recast given in the current study was on the explicit end of the spectrum, and therefore the provision of the two types of OCF led to similar effects.

The second research question which stemmed from the main aim of the study was "to what extent are the perception and production modalities related?"

As for the link between perception and production abilities, results in the current study seem to be closer to the account that describes an indirect link between these two abilities. Participants in this study exhibited high accuracy in the perception of the English wh-question intonation patterns before training, yet this did not transfer to accurate production of the same patterns. Scores in production, although improved after training, remained lower than in perception (see Table 5.1). These results contradict one of the hypotheses of the Speech Learning Model (SLM) (Flege, 2016) which argues that perception precedes production and consequently accurate perception should lead to accurate production. Nonetheless, plenty of studies have found that the connection between the perception and production modalities may not be as consistent and direct as it has been usually accepted (Aliaga-García & Mora, 2009; Alshangiti & Evans, 2014; Chang, 2023; Hattori & Iverson, 2010; Huensch & Tremblay, 2015; Hwang & Lee, 2015; Iverson et al., 2012; Kartushina & Frauenfelder, 2014; Kartushina et al., 2015; Nagle, 2018; Schertz et al., 2015; Shin & Iverson, 2014; Wong, 2013).

The second hypothesis that the modalities of perception and production would show some associations but would work in a relatively independent manner is then confirmed by these results.

With reference to the third research question “which tone (fall or rise) is more challenging to perceive and produce to Chilean Spanish speakers?”, the interaction between the modalities (perception and production), the tones (rise and fall) and the time (T1 and T2) were explored.

The results showed that for perception, both falling (M = 93.80, SD = 12.20) and rising (M = 95.70, SD = 7.49) tones showed a high accuracy rate after training, which may suggest that participants had a robust consolidation of the rising and falling tones for that modality. As for production after training, however, results indicated that the falling tone (M = 86.70, *SD* = 20.20) seems to present less problems than the rising tone (M = 74.60, *SD* = 30.30). Thus, participants may have been more prone to use a falling tone for wh-questions as they had been explicitly instructed to always use the default falling tone for this type of sentence in English until this training. This already consolidated intonation pattern for English wh-questions could have resulted in some sort of interference upon using the requested rising intonation and reveals the lack of variety of patterns taught for this type of questions.

The third hypothesis that the rising tone would be more challenging to perceive and produce for English wh-questions as this is not associated with the default pattern for this type of sentence is then partially confirmed. In this way, the results may indicate that the production of rising intonation for wh-questions would be more challenging than the production of falling intonation for the same sentence type. Also, the results may suggest that the learning and development of the production modality would require more time and effort than the learning and development of the perception modality for English intonation patterns of wh-questions in learners with Chilean Spanish as L1.

A secondary objective of this study was to determine the role of individual differences (working memory capacity and L2 proficiency) in moderating the perception and production of English intonation patterns (fall and rise), and the types of OCF provided. To this end, measures in L2 proficiency (measured with a listening proficiency test), and in two components of working memory capacity (WMC): phonological working memory (measured with a non-word repetition task), and the central executive (measured with a listening span task) were obtained.

A fourth research question derived from this objective: “to what extent do individual differences (L2 proficiency and working memory capacity) moderate the perception and production of English intonation patterns, and the OCF provided?”

The statistical analysis did not reveal any relation of these factors, the perception or production of the pattern under study, and the types of OCF provided. If individual differences are taken into account, this lack of relation between these measures may be explained by the wide variability in the participants’ scores for each of these variables. In other words, a participant with high results in language proficiency may have low scores in the perception or production measures, illustrating that the high listening capacity does not modulate highly accurate perception nor production of the intonation patterns for this type of question. The same account may be used to interpret the absence of relation between phonological working memory (PWM) and the central executive (CE) and the perception and production abilities for this group of learners after receiving OCF.

Despite not finding a statistical relation between language proficiency level, the modalities of production and perception and the types of OCF provided, the fact that all participants had a high language proficiency level (M = 80.0, SD = 6.50) and that the gains in production from T1 to T2 were found to be significant regardless of the OCF type (recast and metalinguistic) may indicate that the degree of explicitness is not relevant when correcting advanced users of an L2 (Ammar & Spada, 2006; Nassaji, 2015) or that high proficiency learners would benefit from any type of OCF (Ammar & Spada, 2006; Lin & Hedgcock, 1996; Williams, 2005). Furthermore, as the participants in the current study showed a high proficiency level (B2: 18 participants; C1: 22 participants), it remains to be seen the extent to which lower-proficiency learners (A1, A2, B1) would benefit from this type of OCF for this type of pronunciation learning activity. Moreover, as only the students' year at university and a listening proficiency test were used to account for the participants' English level, a more comprehensive type of assessment (e.g., a four-skills language test) would be suggested for future research.

The fourth hypothesis that high-proficiency learners would benefit from any type of OCF, whilst low-proficiency ones would only benefit from metalinguistic OCF is then again partially confirmed. As the results showed, all participants of this study were high-proficiency learners whose production of the intonation patterns for English wh-questions significantly improved after the provision of either type of OCF (recast or metalinguistic).

In relation to the interaction between working memory capacity (PWM and CE ability) and the degree to which the OCF provided was noticed and used as uptake, the statistical analyses revealed that these two were not associated. However, it is important to note that the scores of the PWM capacity (measured with a non-word repetition task) were overall remarkably high (M = 98.4, SD: 1.4), which would explain their ability to respond positively to the two types of OCF. However, it is important to note that PWM was measured following Oportus and Ferreira’s (2015) model where non-words are repeated right after they are presented in a randomised way. For this reason, it would be interesting to determine if the use of a more challenging non-word repetition task (i.e., items presented in sets of ascending size) could produce different PWM results.

In terms of the results of the CE test (listening span task), these were heterogenous, and no associations were established between these and changes in perception and production. It is worth mentioning that scores in perception before training were already close to ceiling effect while production scores were around 20% lower, being production ability the only one which improved after training (see Table 5.1). These results are consistent with those of Révész’s (2012) who found a relationship between high PWM measures and improvement in oral production, but not so between CE ability and oral production. She suggests that as PWM is responsible for storing and maintaining acoustic information in short-term memory, this would lead participants to have more chances to notice oral feedback and make corrections. In other words, this may imply that PWM capacity alone would be sufficient to predict the gains in production of L2 speech of participants after the provision of oral corrective feedback.

Another reason for the CE ability not to be associated with the noticing of recasts or metalinguistic OCF may be explained by the nature of these two types of feedback. Despite recast being usually described as an implicit type of OCF, the recast provided in this study lay at a more explicit end of the spectrum since participants knew that if feedback was given, they were being corrected, and that corrections dealt with intonation problems. In this way, differences in the participants' CE ability would not entail much of an advantage as the attentional control needed to focus, find, and integrate the corrective feedback was minimal. Furthermore, since the conditions under the two types of feedback provided constituted controlled settings, both recast and metalinguistic feedback may have been as equally salient for participants to easily notice the corrections and adjust their output. This last point would suggest all corrective feedback strategies seem to be effective when provided in laboratory settings, in contrast to classroom settings where explicit corrective feedback techniques are usually more effective than their implicit equivalent (Lyster et al., 2013; Nicholas et al., 2001). Therefore, it is reasonable for both recast and metalinguistic OCF to show similar positive effects.

The fifth hypothesis that PWM would relate positively with metalinguistic OCF, whilst the CE would relate positively with recast was then only confirmed to some extent. The remarkably high scores of the PWM of the participants may suggest that PWM relates positively with explicit types of OCF (explicit recast and metalinguistic OCF) when the learning and development of the production modality is concerned. Establishing associations between CE ability and recast (implicit type) was not possible since the recast used in this study was rather explicit in nature.

7. CONCLUSIONS AND PEDAGOGIGAL IMPLICATIONS

Production training with individualised OCF showed to be significantly beneficial to improve the production of intonation patterns of English wh-questions. Perception ability was highly accurate before training which may not have left much room for improvement. The fact that production scores were low and perception ones were high before training may suggest that perception and production modalities are not strictly connected, and that somewhat different mechanisms are involved in their functioning.

Furthermore, it may seem that L2 learners with a high language proficiency, high perception ability and high PWM would benefit from any type of OCF strategy (metalinguistic or recast) when suprasegmental errors are concerned.

In addition, given that both recast and metalinguistic OCF led to gains in L2 production, it would be recommended that English teachers at tertiary level make use of an eclectic approach (i.e., explicit and implicit OCF strategies) to deal with suprasegmental errors in production.

Considering that L2 suprasegmental acquisition and development is considered to be gradual and slow in nature (Saito, 2018), the use of delayed OCF by means of recordings may be efficient in providing learners with the sufficient time to self-paced practice in a private, stress-free environment as well as the opportunity for them to develop self-monitoring skills.

Also, as only four training sessions of explicit instruction together with OCF proved to be effective to obtain positive results, it is recommended that university teachers use this strategy in the classroom. Seeing that both modalities have been found to exhibit separate and autonomous mechanisms, training should target speech perception and production.

Finally, if delayed (recordings) OCF is to be used, teachers may decide to focus only on the most frequent or neutral (default) suprasegmental features of English, given that the provision of feedback can be time-consuming and laborious.

8. LIMITATIONS AND FUTURE DIRECTIONS

The current study has some limitations that may restrict the generalisability of the results.

Firstly, the lack of a control group may lead us to believe that the effects observed with the production gains are not truly due to the intervention itself. However, given that the participants' characteristics (L1 background, experience with the L2, similar L2 proficiency levels) were controlled, sampling randomisation was employed, and there is enough evidence in the literature which confirms that the provision of OCF is associated with positive outcomes, the use of only comparison groups was chosen to compare which OCF type (recast vs metalinguistic) was more effective. Despite this, it would be recommended to include a control group to make conclusions more robust.

Secondly, pilot studies for the perception and production pre- and post-tests should have been conducted to identify potential problems with their study design. Therefore, the ceiling-effect scores in perception obtained by both experimental groups may have been attributed to either their enhanced ability to perceive the intonation patterns or to the fact that the instrument was not challenging enough. As no pilot studies for these tests were run, it remains unclear if adjustments were needed.

Thirdly, L2 proficiency was measured considering the students' year at university and the scores of the listening component of a standardised proficiency test

(Oxford Placement Test 2). The decision to select only one component (listening) from this test was taken on the grounds that little time was available. It would be advisable then to give participants a comprehensive (four-skills) standardised test if time permits.

Finally, the non-word repetition task followed a simple administration procedure where non-words had to be repeated immediately after their presentation in a randomised way. This design may have been too simple leading to high scores in phonological working memory. Therefore, it would be important to determine if the use of a more challenging non-word repetition task (i.e., items presented in sets of ascending size) could produce different PWM results. Furthermore, the construction of the listening span task exceeded its due date and was finished when the participants were not available any longer. For this reason, only a subset of 20 participants were around to take the listening span task.

REFERENCES

Ahmadian, M.J. (2012). The relationship between working memory and oral production under task-based careful online planning condition. *TESOL Quarterly 46* (1), 165–175. https://www.jstor.org/stable/41576034

Aliaga-García, C. and Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkings, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 2–31). Cambridge Scholars Publishing.

Allan, D. (1992). *Oxford Placement Test 2*. Oxford University Press.

Allen, D., & Mills, A. (2016). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research, 20*(4), 498–513. https://doi.org/10.1177/1362168814561902

Alptekin, C. and Erçetin, G. (2012). The effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly 45,* 235–266. https://doi.org/10.5054/tq.2011.247705

Alshangiti, W., & Evans, B. G. (2014). Investigating the domain-specificity of phonetic training for second-language learning: Comparing the effects of production and perception training on the acquisition of English vowels by Arabic learners of English. *Proceedings of the 10th International Seminar on Speech Production*, 9–12. http://www.issp2014.uni-koeln.de/wpcontent/uploads/2014/Proceedings_ISSP_revised.pdf

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, *28*(4), 543–574. https://doi.org/10.1017/S0272263106060268

Archibald, L., & Gathercole, S. (2006). Nonword repetition: a comparison of tests. *Journal of Speech, Language and Hearing Research, 49*, 970-983. https://doi.org/10.1044/1092-4388(2006/070)

Audacity Team. (2022). *Audacity(R): Free Audio Editor and Recorder* (Version 3.1.3) [Computer software]. https://audacityteam.org/

Azizi, M., & Nemati, M. (2018). Motivating the Unmotivated: Making Teacher Corrective Feedback Work. *Issues in Language Teaching (ILT), 7*(1), 87-110. https://doi.org/10.22054/ilt.2019.42762.401

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189-208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D. (2017). Modularity, working memory and language acquisition. *Second Language Research, 33*(3), 299–311. https://www.jstor.org/stable/26375886

Bozorgian, H., & E. Shamsi. (2020). Computer-assisted pronunciation training on Iranian EFL learners' use of suprasegmental features: A case study. *Computer-Assisted Language Learning Electronic Journal, 21*(1), 93–113.

Brandl, A., González, C., & Bustin, A. (2020). The development of intonation in L2 Spanish: A perceptual study. In A. Morales-Front, M. J. Ferreira, R. P. Leow, & C. Sanz (Eds.), *Hispanic linguistics: Current issues and new directions* (pp. 12–31). John Benjamins Publishing Company. https://doi.org/10.1075/ihll.26.01bra

Brea-Spahn, M. R. (2009). *Spanish-Specific Patterns and Nonword Repetition Performance in English Language Learners*. [Doctoral dissertation, University of South Florida] https://digitalcommons.usf.edu/etd/1875

Bryfonski, L., & Ma, X. (2019). Effects of Implicit versus Explicit Feedback on Mandarin Tone Acquisition in a SCMC Learning Environment. *Studies in Second Language Acquisition, 42*(1), 61-88. https://doi.org/10.1017/S0272263119000317

Cabrelli, J. (2023). Language attrition and L3/Ln. In J. Cabrelli, A. Chaouch-Orozco, J. González Alonso, S. Pereira Soares, E. Puig-Mayenco, & J. Rothman (Eds.), *The Cambridge handbook of third language acquisition* (pp. 317-353. Cambridge University Press.

Cardinali, R. F., & Barbeito, M. C. (2018). Developing intonation skills in English: A systemic functional linguistics perspective. *Global Journal of Foreign Language Teaching, 8*(1), 11-20. https://doi.org/10.18844/gjflt.v8i1.3222

Chang Y.S. (2023). Effects of production training with ultrasound biofeedback on production and perception of second-language English tense-lax vowel contrasts. *Journal of Speech, Language, and Hearing Research, 66*(5):1479-1495. https://doi.org/10.1044/2023_JSLHR-22-00587

Chela-Flores, B. (2003). Optimizing the teaching of English suprasegmentals. *Lenguas Modernas,* 28-29, 255-273. https://lenguasmodernas.uchile.cl/index.php/LM/article/view/45457

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867-873. https://doi.org/10.1037/0012-1649.32.5.867

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. *Companion Volume*. Council of Europe.

Cowan, N. (2013). Working memory and attention in language use. In J. Guandouzi, F. Loncke, & M. J. Williams (Eds.), *The Handbook of Psycholinguistics and Cognitive Processes* (pp. 75-97). Psychology Press.

Crystal, D. (2003). *A Dictionary of Linguistics & Phonetics*. Blackwell.

Darabad, A. M. (2014). Corrective feedback interventions and EFL learners' pronunciation: A case of–s or–es ending words. *International Journal of Learning and Development, 4*(1), 40-58. https://doi.org/10.5296/ijld.v4i1.5010

Darcy, I., Park H. & Yang, C-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences, 40*, 63-72. https://doi.org/10.1016/j.lindif.2015.04.005

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty, & M. H. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 313-348). Blackwell.

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language learning, 48*(3), 393-410.

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

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited.* Routledge.

Dowd, A., Smith, J., & Wolfe, J. (1998). Learning to Pronounce Vowel Sounds in a Foreign Language using Acoustic Measurements of the Vocal Tract as Feedback in Real Time. *Language and Speech, 41*(1), 1–20. https://doi.org/10.1177/002383099804100101

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*(2), 305–352. https://doi.org/10.1017/S027226310505014X

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 339–360). Oxford University Press.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28*(3), 339–368. https://doi.org/10.1017/S0272263106060141

Flege, J. (2016, June 10-12). *The role of phonetic category formation in second language speech acquisition* [Conference session]. New Sounds, Aarhus University, Denmark.

Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation, 3*(1), 34–56. https://doi.org/10.1075/jslp.3.1.02foo

Fouché, C. B., & De Vos, A. S. (2011). Formal formulations. In A.S. De Vos, H. Strydom, C.B., Fouché & C.S.L. Delport (Eds), *Research at grass roots: for the social sciences and human service professions* (pp. 89-100). Van Schaik Publishers.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to Design and Evaluate Research in Education* (8th ed.). McGraw-Hills.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*(3), 445–474. https://doi.org/10.1017/S0272263112000149

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of Recasts and Prompts on Second Language Pronunciation Development: Teaching English /r/ to Korean Adult EFL Learners. *System, 60*, 117-127. https://doi.org/10.1016/j.system.2016.06.007

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation, 2*(1), 56-92.

Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. *Proceedings of the 4th pronunciation in second language learning and teaching conference*, 194-206.

Gutiérrez, F. (2012). Trouble Spots in the Learning of English Intonation by Spanish Speakers. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 219-230). Springer.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. Mounton & Co. N. V. Publishers.

Hamawand, Z., & Hasan AL-JAF, K. A. (2023). Patterns of Intonation in Central Kurdish A Study in Cognitive Phonology. *Journal of University of Human Development, 9*(3), 31–36. https://doi.org/10.21928/juhd.v9n3y2023.pp31-36

Hanh, P. T., & Tho, P. X. (2018). Oral Corrective Feedback in EFL/ESL Classrooms: Classification models. *VNU Journal of Foreign Studies, 34*(5). https://doi.org/10.25073/2525-2445/vnufs.4301

Hattori, Y. (2023). L2 Japanese Proficiency and Working Memory Capacity. *Buckeye East Asian Linguistics, 7*, 69-76. http://hdl.handle.net/1811/103529

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English /r/-/l/ perception and production by adult Japanese speakers. In *Interspeech Workshop on Second Language Studies: Acquisition, Learning, Education and Technology.*

Henriksen, N. C. (2009). Wh-question intonation in Peninsular Spanish: Multiple contours and the effect of task type. *Journal of Portuguese Linguistics, 8*(1), 45-73. http://doi.org/10.5334/jpl.121

Hernández-Sampieri, R., Fernández, C. y Baptista, M. del P. (2010). *Metodología de la investigación*. McGraw-Hill.

Hernández Nieto, R. (2011). *Instrumentos de recolección de datos en ciencias sociales y ciencias biomédicas*. Universidad de Los Andes.

Hidalgo, A., & Quilis, M. (2012). *La voz del lenguaje: Fonética y fonología del español*. Tirant Humanidades.

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics, 52*, 105–120. https://doi.org/10.1016/j.wocn.2015.06.007

Hwang, H., & Lee, H. (2015). The effect of high variability phonetic training on the production of English vowels and consonants. *Proceedings of the 18th*

International Congress of Phonetic Sciences, 1041–1045. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0466.pdf

Instructure. (2022). *CANVAS LMS (Learning Management System).* https://www.instructure.com/canvas

Iverson, P., Pinet, M., & Evans, B. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(1), 145-160. https://doi.org/10.1017/S0142716411000300

Jalal, B. R., & Alahmed, K. I. (2022). The effect of using recast and explicit corrective feedback on improving English pronunciation of Iraqi intermediate school students. *Journal of Positive School Psychology, 6*(10), 1264-1272. https://journalppw.com/index.php/jpsp/article/view/13337/8662

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137-166. https://doi.org/10.1017/S0261444810000509

Kangatharan, J., Giannakopoulou, A., & Uther, M. (2021). The effect of HVP training in vowel perception on bilingual speech production. *Research in English Language Teaching, 2*(1), 15–32. http://hdl.handle.net/10547/624837

Kapa L. L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language, 7*, 237–252. https://doi.org/10.1016%2Fj.jml.2014.07.004

Karimi, L., & Esfandiari, N. (2016). The effect of recast vs. explicit corrective feedback on Iranian EFL learners' stress patterns learning. *Theory and practice in language studies, 6*(6), 1166-1174. http://dx.doi.org/10.17507/tpls.0606.05

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5*, 1-17. https://doi.org/10.3389/fpsyg.2014.01246

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. *Journal of the Acoustical Society of America, 138*(2), 817-832. https://doi.org/10.1121/1.4926561

Kaushanskaya, M., & Yoo, J. (2013). Phonological short-term and working memory in bilinguals' native and second language. *Applied Psycholinguistics, 34*, 1005-1037. https://doi.org/10.1017/S0142716412000100

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition, 37,* 549–581. https://doi.org/10.1017/S0272263114000618

Ladd, D. R. (1996*). Intonational Phonology*. Cambridge University Press.

Lee, A., & Thomson, R. (2022). What to research and how to research in instructed second language pronunciation. In L. Gurzynski-Weiss, & Y. Kim (Eds.), Instructed Second Language Acquisition Research Methods (pp. 233-256). John Benjamins B.V.

Leeser, M., & Herman, E. (2022). Methodological Issues in Research on Working Memory and L2 Reading Comprehension. In J. Schwieter & Z. Wen, *The Cambridge Handbook of Working Memory and Language* (pp.613-634). Cambridge University Press.

Leseman, P., & Verhagen, J. (2022). Working Memory and L2 Grammar Development in Children. In J. Schwieter & Z. Wen, *The Cambridge Handbook of Working Memory and Language* (pp. 550-572). Cambridge University Press.

Levis, J. M. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation.* Cambridge Applied Linguistics.

Levis, J. M., & Muller Levis, G. (2018). Teaching High-Value Pronunciation Features: Contrastive Stress for Intermediate Learners. *CATESOL Journal, 30(1),* 139-160.

Li, S. (2009). The differential effects of implicit and explicit feedback on second language (L2) learners at different proficiency levels. *Applied Language Learning, 19*, 53–79.

Li, S. (2010). The Effectiveness of Corrective Feedback in SLA: A Meta-Analysis. *Language Learning, 60*, 309-365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Li, S. (2014). The interface between feedback type, L2 proficiency, and the nature of the linguistic target. *Language Teaching Research, 18*(3), 373–396. https://doi.org/10.1177/1362168813510384

Li, S. (2018). Corrective feedback in L2 speech production. In J. Liontas, T. International Association & M. DelliCarpini (Eds.), The TESOL Encyclopedia of English Language Teaching (pp. 1-9). https://doi.org/10.1002/9781118784235.eelt0247

Lin, Y. H., & Hedgcock, J. (1996). Negative feedback incorporation among high-proficiency and low proficiency Chinese-speaking learners of Spanish. *Language Learning, 46*(4), 567–611. https://doi.org/10.1111/j.1467-1770.1996.tb01353.x

Linck, J., Osthus, P., Koeth, J., & Bunting, M. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*, 861–883. https://doi.org/10.3758/s13423-013-0565-2

Linebaugh, G., & Roche, T. B. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language and Learning, 9*(1), A1-A17. https://journal.aall.org.au/index.php/jall/article/view/326

Llinares, A. & Lyster, R. (2014). The influence of context on patterns of corrective feedback and learner uptake: a comparison of CLIL and immersion classrooms. *The Language Learning Journal, 42*(2), 181-194. https://doi.org/10.1080/09571736.2014.889509

Loewen, S. (2004). Uptake in Incidental Focus on Form in Meaning-Focused ESL Lessons. *Language Learning, 54*, 153-188. https://doi.org/10.1111/j.1467-9922.2004.00251.x

Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 361–377). Oxford University Press.

Loewen, S., & Plonsky, L. (2016). *An A-Z of applied linguistics research methods*. Palgrave.

Lu, Y., & Kim, M. (2016). Prosody Transfer in Second Language Acquisition: Tonal Alignment in the Production of English Pitch Accent by Mandarin Native Speakers. *Tsing Hua Journal of Chinese Studies, 46*(4), 785-816. https://doi.org/10.6503/THJCS.2016.46%284%29.04

Luquin, M., & Roothooft, H. (2019). The effects of oral corrective feedback and language anxiety on pronunciation development. *Elia: Estudios de Lingüística Inglesa Aplicada, 19*, 41-69. https://idus.us.es/handle/11441/99215

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*, 37-66. https://doi.org/10.1017/S0272263197001034

Lyster, R., Saito, K. & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1-40. https://doi.org/10.1017/S0261444812000365

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity*. Language Learning, 60*, 501–533.

Mackey, A., Al-Khalil, M., Atanassova, G., Hama, M., Logan-Terry, A., & Nakatsukasa, K. (2007). Teachers' intentions and learners' perceptions about corrective feedback in the L2 classroom. *International Journal of Innovation in Language Learning and Teaching, 1*(1), 129-152. https://doi.org/10.2167/illt047.0

Mackey, A. & Gass, S. M. (2005). *Second Language Research: Methodology and Design.* Lawrence Erlbaum Assoc.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal, 82*(3), 338–356. https://doi.org/10.1111/j.1540-4781.1998.tb01211.x

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34*, 383-388. https://doi.org/10.1080/002075999399738

McCormick, T., & Sanz, C. (2022). Working Memory and L2 Grammar Learning among Adults. In J. Schwieter & Z. Wen, *The Cambridge Handbook of Working Memory and Language* (pp. 573-592). Cambridge University Press.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition, 27*, 79-103. https://doi.org/10.1017/S0272263105050047

McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11*(2), 113-128. https://doi.org/10.1093/applin/11.2.113

Mennen, I. (2015). Beyond Segments: Towards a L2 Intonation Learning Theory. In E. Delais-Roussarie, M. Avanzi & S. Herment (Eds.), *Prosody and Language in*

Contact. Prosody, Phonology and Phonetics (pp. 171-188). Springer. https://doi.org/10.1007/978-3-662-45168-7_9

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen & M. Sato, (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50–68). Routledge.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. R. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention* (pp. 339–364). Erlbaum.

Munro, M., & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning, 49*, 285-310. https://doi.org/10.1111/0023-8333.49.s1.8

Nagle, C.L. (2018). Examining the Temporal Structure of the Perception–Production Link in Second Language Acquisition: A Longitudinal Study. *Language Learning, 68*, 234-270. https://doi.org/10.1111/lang.12275

Naranjo. S. (2020). *L2 learners' perception of falling intonation in English wh- questions* [Unpublished master’s thesis]. Universidad de Concepción.

Nassaji, H. (2013). Interactional feedback: Insights from theory and research. In A. Benati, M. Arche & C. Laval (Eds.), *The grammar dimension in instructed second language learning: Theory, research and practice* (pp. 103-123). Bloomsbury Publishing.

Nassaji, H. (2015). *Interactional feedback dimension in instructed second language learning*. Bloomsbury Publishing.

Navarro Tomás, T. (1966). *Estudios de fonología española*. Las Americas Publishing Company.

Navarro Tomás, T. (2004). *Manual de pronunciación española* (28th ed.). Consejo Superior de Investigaciones Científicas.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51*(4), 719–758. https://doi.org/10.1111/0023-8333.00172

Oportus, R. & Ferreira, A. (2015). *La memoria operativa fonológica en aprendices de español como lengua extranjera: propuesta para el diseño de un instrumento*. [Unpublished master's thesis]. Universidad de Concepción.

Ortega-Llebaria, M., & Colantoni, L. (2014). L2 English intonation. *Studies in Second Language Acquisition, 36*(02), 331–353. https://doi.org/10.1017/S0272263114000011

Ortiz, H., Fuentes, M., & Astruc, L. (2010). Chilean Spanish Intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of Intonation of the Spanish Language* (pp. 255-283). LINCOM publishers.

Park, E. S., Song, S., & Shin, Y. K. (2016). To what extent do learners benefit from indirect written corrective feedback? A study targeting learners of different proficiency and heritage language status. *Language Teaching Research, 20*(6), 678-699. https://doi.org/10.1177/1362168815609617

Park, E., Klieve, H., Tsurutani, C., & Harte, W. (2017). International students' accented English - Communication difficulties and developed strategies. *Cogent Education, 4*(1), 1–15. https://doi.org/10.1080/2331186X.2017.1314651

Parlak, Ö., & Ziegler, N. (2016). 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. https://doi.org/10.1017/S0272263116000310

Pawlak, M. (2021). Corrective Feedback, Developmental Readiness, and Language Proficiency. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge Handbook of Corrective Feedback in Second Language Learning and Teaching* (pp. 733-753). Cambridge University Press. https://doi.org/10.1017/9781108589789.035

Peng, D. (2016). English Intonation Teaching in a Perspective of Communicative Functions. *International Conference on Arts, Design and Contemporary Education (ICADCE), 64,* 1450–1453. https://doi.org/10.2991/icadce-16.2016.356

Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in ns–nns interaction. *Studies in Second Language Acquisition*, *25*(1), 99–126. https://doi.org/10.1017/S0272263103000044

Geçkin, V. (2020, October 28-31). *Does L1 transfer persist in end-state L2 English?* [Paper presentation]. GlobELT & GLOBETS 2020, online. https://globets.org/globeltconference/files/GLOBELT_GLOBETS_2020_Proceedings.pdf#page=94

Prieto, P. (2015). Intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(4), 371-381. https://doi.org/10.1002/wcs.1352

Puga, K., Fuchs, R., Setter, J. & Mok, P. (2017). The perception of English intonation patterns by German L2 speakers of English. *Interspeech, 20*(24), 3241-3245. http://doi.org/10.21437/Interspeech.2017-1279

Quesada, L. & Romero, J. (2018, September 19-21). *The improvement of Spanish/Catalan EFL students' prosody by means of explicit rhythm instruction* [Paper presentation]. The International Symposium on Applied Phonetics (ISAPh), Aizuwakamatsu, Japan. https://www.isca-speech.org/archive/pdfs/isaph_2018/quesadavazquez18_isaph.pdf

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Rauber, A., Rato, A., Kluge, D., & Santos, G. R. (2012). *TP* (Version. 3.1) [Computer software]. http://www.worken.com.br/tp/tp_install.html

Rebolledo, M. (2021). La entonación de las oraciones interrogativas en hablantes normotípicos y pacientes afásicos no fluentes: primer acercamiento. *Revista de Lingüística Teórica y Aplicada, 59*(1), 183-204. http://dx.doi.org/10.29393/rla59-8eomr10008.

Révész, A. (2012). Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures. *Language Learning, 62*, 93-132. https://doi.org/10.1111/j.1467-9922.2011.00690.x

Roach, P. (2009). *English phonetics and phonology* (4th ed.). Cambridge University Press.

Rogers, B., Rao, R., & Burner, M. (2021). Exploring extended focus and meaning in Chilean Spanish intonational plateau contours. In J. Colomina-Almiñana & S. Sessarego (Eds.), *Language Patterns in Spanish and Beyond: Structure, Context and Development* (pp.199-222). Routledge.

Rogers, J., & Révész, A. (2020). Experimental and quasi-experimental designs. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 133-143). Routledge.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 229–248). Oxford University Press.

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*, 499-529. https://doi.org/10.1111/lang.12015

Saito, K. (2015). Variables affecting the effects of recasts on L2 pronunciation development. *Language Teaching Research, 19*(3), 276–300. https://doi.org/10.1177/1362168814541753

Saito, K. (2018). Advanced Second Language Segmental and Suprasegmental Acquisition. The Handbook of Advanced Proficiency in Second Language Acquisition, 282–303. https://doi.org/10.1002/9781119261650.ch15

Saito, K., Petrova, K., Suzukida, Y., Kachlicka, M., & Tierney, A. (2022). Training auditory processing promotes second language speech acquisition. *Journal of Experimental Psychology: Human Perception and Performance, 48*(12), 1410–1426. https://doi.org/10.1037/xhp0001042

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. https://doi.org/10.1111/lang.12345

Saito, Y., & Saito, K. (2016). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research, 21*(5), 589-608. https://doi.org/10.1177/1362168816643111

Saito, K., & Wu, X. (2014). Communicative focus on form and second language suprasegmental learning: Teaching Cantonese Learners to Perceive Mandarin Tones. *Studies in Second Language Acquisition, 36*(4), 647-680. https://doi.org/10.1017/S0272263114000114

Sakai, M. (2017). (Dis)Connecting perception and production: Training native speakers of Spanish on the English /i/-/ɪ/ distinction. [Doctoral dissertation, Georgetown University] https://repository.library.georgetown.edu/bitstream/handle/10822/1042879/Sakai_georgetown_0076D_13518.pdf?sequence=1&isAllowed=y

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187–224. https://doi.org/10.1017/S0142716417000418

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics, 52*, 183–204. https://doi.org/10.1016/j.wocn.2015.07.003

Scheuer, S., & Horgues, C. (2021). Corrective feedback and unintelligibility: Do they work in tandem during tandem interactions? In A. Kirkova-Naskova, A. Henderson & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 224–251). John Benjamins. https://doi.org/10.1075/aals.19.10sch

Service, E. (1992). Phonology, working memory and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45*(1), 21-50. https://doi.org/10.1080/14640749208401314

Service, E., & Simard, D. (2022). How Measures of Working Memory Relate to L2 Vocabulary. In J. Schwieter & Z. Wen, *The Cambridge Handbook of Working Memory and Language* (pp. 529-549). Cambridge University Press.

Sheen, Y. (2007). The effects of corrective feedback, language aptitude and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition: A Collection of Empirical Studies* (pp. 301–322). Oxford University.

Sheen, Y. (2013). What should teachers do about learner errors? *The European Journal of Applied Linguistics and TEFL, 2*(2), 5-28.

Shin, D.-J., & Iverson, P. (2014). An Experimental Study of Vowel Epenthesis among Korean Learners of English. Phonetics and Speech Sciences. *The Korean Society of Speech Sciences, 6*(2), 163-174. https://doi.org/10.13064/ksss.2014.6.2.163

Solon, M., & Long, A. Y. (2018). Acquisition of Phonetics and Phonology Abroad. In C. Sanz, & A. Morales-Front (Eds.), *The Routledge handbook of study abroad research and practice* (pp. 71-85). Routledge.

Sosa, J. M. (2003). Wh-questions in Spanish: Meaning and configuration variability. *Catalan Journal of Linguistics, 2*, 229-247. http://doi.org/10.5565/rev/catjl.51

Tan, R. S. K. (2016). How do We Stress? Lexical Stress in Malaysian and British English. In "Author", *English in Malaysia: Current Use and Status* (pp. 65-85). Brill.

Tagarelli, K. M., Borges Mota, M. & Rebuschat, P. (2015). Working memory, learning context, and the acquisition of L2 Syntax. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing: Theory, research and commentary* (pp. 224–247). Multilingual Matters.

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: a meta-analysis. *Studies in Second Language Acquisition, 41*(2), 363-387. https://doi.org/10.1017/S0272263118000311

Tench, P. (2015). *The Intonation Systems of English*. Bloomsbury Publishing.

Thompson, R. (2022). Perception in Pronunciation Training. In J. Levis, T. Derwing & S. Sonsaat-Hegelheimer (Eds.), *Second Language Pronunciation: Bridging the Gap between Research and Teaching* (pp. 42-60). Wiley Blackwell.

Tremblay, A. (2011). Proficiency assessment standard in second language acquisition research "clozing" the gap. *Studies in Second Language Acquisition, 33*(3), 339–372. https://doi.org/10.1017/S0272263111000015

Urrutia, H. (2007). La naturaleza del acento en español: nuevos datos y perspectivas. *Revista de lingüística teórica y aplicada, 45*(2), 135-142. https://dx.doi.org/10.4067/S0718-48832007000200010

Valenzuela, M. (2013). A comparative analysis of intonation between Spanish and English speakers in tag questions, wh-questions, inverted questions, and repetition questions. *Revista Brasileira de Lingüística Aplicada, 13*(4), 1061-1083. https://doi.org/10.1590/S1984-63982013005000021

Véliz, M., Pereira, Y., Rojas, C., Cerda, K., & Biedroń, A. (2022). *On the effects of implicit and explicit corrective feedback on perception and production of English sentence accent and (post)nuclear intonation as mediated by L2 foreign language aptitude,* *working memory, and musical ability* [Manuscript in preparation]. Universidad de Talca.

Wells, J. (2009). *English Intonation. An Introduction*. Cambridge University Press.

Wen, Z. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. Wen, B. Mailce, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 41–62). Multilingual Matters.

Wen, Z., Biedroń, A., & Skehan P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*(1), 1-31. https://doi.org/10.1017/S0261444816000276

Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 671-691). Erlbaum.

Williams, J.N. (2012). Working memory and SLA. In S. Gass, & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 427-441). Routledge/Taylor & Francis.

Wong, J.W.S. (2013). The effects of perceptual and/or productive training on the perception and production of English vowels /ɪ/ and /iː/ by Cantonese ESL learners. *Proc. Interspeech 2013*, 2113-2117. http://dx.doi.org/10.21437/Interspeech.2013-501

Xasanov, M. & Mardieva, N. (2023). The aspects and components of intonation. *O'ZBEKISTONDA FANLARARO INNOVATSIYALAR VA ILMIY TADQIQOTLAR JURNALI, 2*(18), 488–494. https://doi.org/10.5281/zenodo.7852756

Yang, L. (2016, March 25-28). *Optimizing Pronunciation and Prosody Teaching in Second Language Learning* [Paper presentation]. The International Symposium on Applied Phonetics (ISAPh), Nagoya, Japan. https://www.isca-speech.org/archive/pdfs/isaph_2016/yang16_isaph.pdf

Yilmaz, Y. (2013). Relative Effects of Explicit and Implicit Feedback: The Role of Working Memory Capacity and Language Analytic Ability. *Applied Linguistics, 34*(3), 344–368, https://doi.org/10.1093/applin/ams044

Yurtbasi, M. (2017). Correcting English learner's suprasegmental errors. *Global Journal of Foreign Language Teaching. 7*(4), 126-131. http://dx.doi.org/10.18844/gjflt.v7i4.3000

Zoghlami, N. (2014). Testing L2 listening proficiency: reviewing standardized test within a competence-based framework. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 Proficiency: Perspectives from SLA* (pp. 191-207). Multiligual Matters.

Zohrabi, M., & Behboudnia, N. (2017). The Effect of Explicit and Implicit Corrective Feedback on Segmental Word-Level Pronunciation Errors. *Applied Research on English Language, 6*(2), 237-266. https://doi.org/10.22108/are.2017.21348

Zychowicz, K., Biedroń, A., & Pawlak, M. (2017). Polish Listening SPAN: a new tool for measuring verbal working memory. *Studies in Second Language Learning and Teaching, 7*(4), 601-618. http://doi.org/10.14746/ssllt.2017.7.4.3

APPENDICES

APPENDIX A: LIST OF WH-QUESTIONS USED FOR PERCEPTION TESTS

WH-QUESTIONS FOR PERCEPTION TEST: FALLING TONE

1. WHO’S YOUR TEACHER? ↘
2. WHERE’S YOUR ROOM? ↘
3. HOW’S YOUR FAMILY? ↘
4. WHO’S YOUR MENTOR? ↘
5. WHEN IS YOUR BIRTHDAY? ↘
6. WHAT DID YOU SAY? ↘
7. WHERE DO YOU LIVE? ↘
8. HOW DID SHE TRAVEL? ↘
9. WHERE DO YOU WORK? ↘
10. WHY DID YOU LIE? ↘
11. WHERE DOES THE BOY SLEEP? ↘
12. WHEN DOES THE MEETING START? ↘
13. WHEN DO THE PEOPLE ARRIVE? ↘
14. WHO DOES THE BOY KICK? ↘
15. WHO DOES THE LADY SEE? ↘
16. WHAT DOES THE BOY PUSH? ↘
17. WHEN DOES THE MOVIE END? ↘
18. WHEN DO THE PEOPLE LEAVE? ↘
19. WHO DOES THE BOY ADMIRE? ↘
20. WHY DOES THE GIRL SMILE? ↘
21. HOW DID SHE TRAVEL THE COUNTRY? ↘
22. HOW DID YOU BURN THAT STEAK? ↘
23. WHEN DID THEY WITNESS THE ACCIDENT? ↘
24. WHAT DO YOU CALL YOUR FATHER? ↘
25. WHO DID YOU TELL YOUR STORY? ↘
26. WHAT DO YOU CALL YOUR MOTHER? ↘
27. WHAT DID YOU GET YOUR FRIEND? ↘
28. WHO DID YOU TELL THE TRUTH? ↘
29. WHY DO YOU BRING YOUR PET? ↘
30. WHY DID YOU STEAL THOSE FLOWERS? ↘
31. WHAT ARE YOU DOING WITH THAT KNIFE? ↘
32. WHY ARE YOU SPEAKING TO THAT PERSON? ↘
33. WHERE ARE YOU STUDYING WITH YOUR CLASSMATES? ↘
34. WHERE ARE YOU LEAVING WITH THAT RING? ↘
35. HOW ARE THEY DEALING WITH THE DIVORCE? ↘
36. WHERE ARE YOU GOING IN THAT DRESS? ↘
37. WHERE ARE YOU GETTING OFF THE TRAIN? ↘
38. WHEN ARE YOU VISITING WITH YOUR KIDS? ↘
39. HOW ARE YOU FEELING ABOUT THE IDEA? ↘
40. HOW ARE THEY DEALING WITH THE PROBLEM? ↘

WH-QUESTIONS FOR PERCEPTION TEST: RISING TONE

1. WHAT'S YOUR ADDRESS? ↗
2. WHAT'S YOUR NAME? ↗
3. WHERE'S THE PAIN? ↗
4. HOW'S THE WEATHER? ↗
5. WHEN IS THE PARTY? ↗
6. WHAT DID YOU SHOUT? ↗
7. HOW DID HE DIE? ↗
8. WHO DID YOU MARRY? ↗
9. WHY DID THEY CALL? ↗
10. WHO DID YOU INVITE? ↗
11. WHAT DOES THE MAN THROW? ↗
12. HOW DOES THE LADY FEEL? ↗
13. WHO DOES THE MAN HOLD? ↗
14. WHY DOES THE MAN DANCE? ↗
15. WHY DOES THE BOY SING? ↗
16. WHERE DOES THE LADY SIT? ↗
17. HOW DOES THE LADY WALK? ↗
18. WHO DOES THE MAN LOVE? ↗
19. WHO DOES THE LADY DESPISE? ↗
20. WHY DOES THE BOY COMPLAIN? ↗
21. HOW DO YOU SPELL THAT WORD? ↗
22. HOW DID SHE TRAVEL THE WORLD? ↗
23. WHEN DID HE OPEN THE PRESENT? ↗
24. WHY DID YOU BUY THAT CAR? ↗
25. WHERE DID YOU TAKE THOSE PHOTOS? ↗
26. WHAT DID HE SEND YOUR SISTER? ↗
27. WHY DID YOU BUY THAT HOUSE? ↗
28. WHY DO YOU BRING YOUR SON? ↗
29. WHY DID YOU STEAL THOSE COINS? ↗
30. WHERE DID YOU FIND THOSE SHELLS? ↗
31. WHAT ARE YOU PUTTING IN THAT BAG? ↗
32. WHAT ARE YOU DRINKING IN THAT MUG? ↗
33. WHY ARE YOU LISTENING TO THAT WOMAN? ↗
34. WHERE ARE YOU STAYING WITH YOUR BOYFRIEND? ↗
35. WHEN ARE YOU PLAYING WITH THE BAND? ↗
36. WHAT ARE THEY DRAWING WITH THAT PENCIL? ↗
37. WHERE ARE YOU GETTING ON THE BUS? ↗
38. WHEN ARE THEY COMPETING IN THE CONTEST? ↗
39. WHEN ARE THEY COMING TO THE CITY? ↗
40. HOW ARE YOU WORKING WITH THE TEAM? ↗

APPENDIX B: LIST OF WH-QUESTIONS USED FOR PRODUCTION TESTS

WH-QUESTIONS FOR PRODUCTION TEST 1:

1. WHAT'S YOUR NUMBER? ↗
2. WHERE'S THE STATION? ↗
3. HOW'S THE FOOD? ↗
4. WHO'S YOUR BOSS? ↘
5. WHEN IS THE TRIP? ↘
6. WHAT DID YOU WRITE? ↗
7. WHERE DO YOU GO? ↘
8. HOW DID SHE KNOW? ↗
9. WHO DID YOU HELP? ↘
10. WHY DID YOU KNOCK? ↘
11. WHAT DOES THE MAN PAINT? ↗
12. WHERE DOES THE BOY HIDE? ↘
13. HOW DOES THE LADY MOVE? ↘
14. WHEN DOES THE MEETING BEGIN? ↘
15. WHEN DOES THE STORE OPEN? ↗
16. WHO DOES THE BOY LIKE? ↘
17. WHO DOES THE MAN ADORE? ↘
18. WHO DOES THE LADY CHOOSE? ↗
19. WHY DOES THE MAN RUN? ↗
20. WHY DOES THE BOY SCREAM? ↗
21. HOW DID YOU MAKE THAT MESS? ↗
22. HOW DID SHE PASS THE TEST? ↘
23. WHEN DID HE OPEN THE DOOR? ↗
24. WHAT DO YOU CALL YOUR GRANDMA? ↘
25. WHAT DID YOU GET YOUR COUSIN? ↗
26. WHO DID YOU TELL THE SECRET? ↘
27. WHY DID YOU BUY THAT BIKE? ↗

28. WHY DO YOU BRING YOUR CHILDREN? ↗

29. WHY DID YOU STEAL THOSE TOYS? ↘

30. WHERE DID YOU TAKE THOSE PICTURES? ↘

31. WHAT ARE YOU PUTTING IN THAT BOWL? ↘

32. WHAT ARE YOU DRINKING IN THAT GLASS? ↗

33. WHY ARE YOU LISTENING TO THAT MUSIC? ↗

34. WHERE ARE YOU GOING IN THAT SUIT? ↘

35. WHERE ARE YOU GETTING OFF THE PLANE? ↗

36. WHERE ARE YOU STAYING WITH YOUR GIRLFRIEND? ↘

37. WHEN ARE YOU VISITING WITH YOUR GANG? ↗

38. WHEN ARE YOU PLAYING WITH THE DOGS? ↗

39. HOW ARE YOU FEELING ABOUT THE FUTURE? ↘

40. HOW ARE THEY DEALING WITH THE WAR? ↘

WH-QUESTIONS FOR PRODUCTION TEST 2

1. WHAT’S YOUR SIGN? ↗

2. WHERE’S YOUR OFFICE? ↘

3. HOW’S YOUR HUSBAND? ↗

4. WHO’S YOUR COLLEAGUE? ↘

5. WHEN IS YOUR FLIGHT? ↗

6. WHAT DID YOU EAT? ↘

7. WHERE DO YOU STOP? ↘

8. HOW DID HE WIN? ↗

9. WHO DID YOU KILL? ↘

10. WHY DID THEY LOSE? ↗

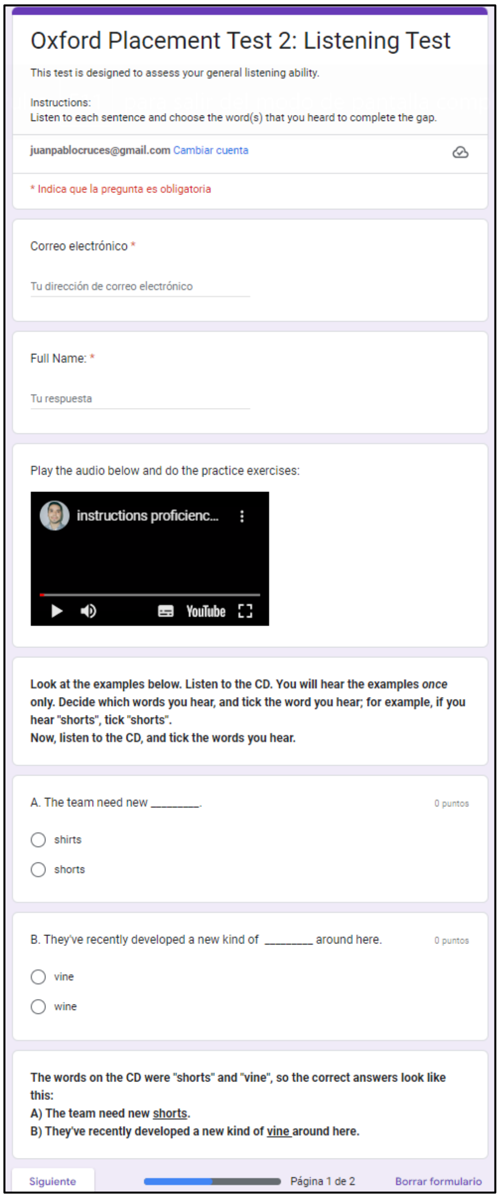
11. WHAT DOES THE BOY WANT? ↗

12. WHERE DOES THE LADY READ? ↘

13. HOW DOES THE LADY STAND? ↗

14. WHEN DOES THE MOVIE AIR? ↘
15. WHEN DOES THE STORE CLOSE? ↗
16. WHO DOES THE BOY HATE? ↗
17. WHO DOES THE MAN TRUST? ↘
18. WHO DOES THE LADY SAVE? ↘
19. WHY DOES THE GIRL CRY? ↗
20. WHY DOES THE BOY LAUGH? ↘
21. HOW DID YOU MAKE THAT CAKE? ↘
22. HOW DID SHE FAIL THAT COURSE? ↗
23. WHEN DID THEY WITNESS THE ATTACK? ↘
24. WHAT DO YOU CALL YOUR CAT? ↗
25. WHAT DID HE SEND YOUR BROTHER? ↘
26. WHO DID YOU TELL YOUR PASSWORD? ↗
27. WHY DID YOU BUY THAT PHONE? ↘
28. WHY DO YOU BRING YOUR PARENTS? ↗
29. WHY DID YOU STEAL THOSE NOTES? ↘
30. WHERE DID YOU FIND THOSE LAMPS? ↗
31. WHAT ARE YOU DOING WITH THAT FORK? ↗
32. WHAT ARE THEY DRAWING WITH THAT PEN? ↘
33. WHY ARE YOU SPEAKING TO THAT IDIOT? ↗
34. WHERE ARE YOU GOING IN THAT OUTFIT? ↘
35. WHERE ARE YOU GETTING ON THE BOAT? ↗
36. WHERE ARE YOU LEAVING WITH THAT MONEY? ↘
37. WHEN ARE THEY COMPETING IN THE RACE? ↗
38. WHEN ARE THEY COMING TO THE CLASSROOM? ↘
39. HOW ARE YOU WORKING WITH THE STUDENTS? ↗
40. HOW ARE THEY DEALING WITH THE COSTS? ↘

APPENDIX C: LISTENING PROFICIENCY TEST – OXFORD PLACEMENT TEST 2



Oxford Placement Test 2: Listening Test

This test is designed to assess your general listening ability.

Instructions:
Listen to each sentence and choose the word(s) that you heard to complete the gap.

juanpablocruces@gmail.com Cambiar cuenta

* Indica que la pregunta es obligatoria

Correo electrónico *

Tu dirección de correo electrónico

Full Name: *

Tu respuesta

Play the audio below and do the practice exercises:

instructions proficienc...

YouTube

Look at the examples below. Listen to the CD. You will hear the examples *once* only. Decide which words you hear, and tick the word you hear; for example, if you hear "shorts", tick "shorts".
Now, listen to the CD, and tick the words you hear.

A. The team need new ________. — 0 puntos

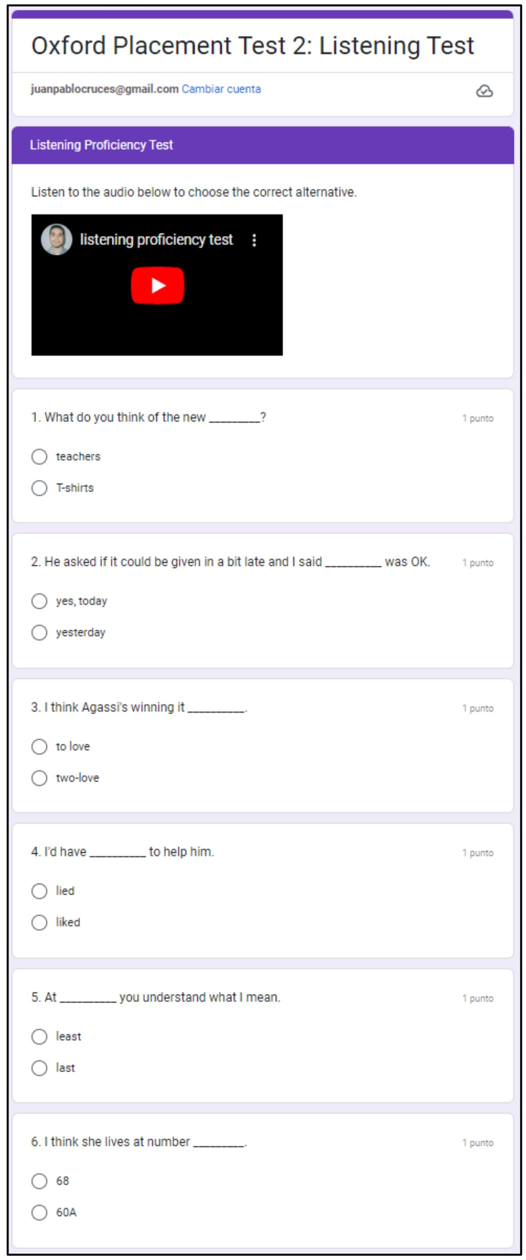
- shirts
- shorts

B. They've recently developed a new kind of ________ around here. — 0 puntos

- vine
- wine

The words on the CD were "shorts" and "vine", so the correct answers look like this:
A) The team need new shorts.
B) They've recently developed a new kind of vine around here.

Siguiente — Página 1 de 2 — Borrar formulario



Oxford Placement Test 2: Listening Test

juanpablocruces@gmail.com Cambiar cuenta

Listening Proficiency Test

Listen to the audio below to choose the correct alternative.

listening proficiency test

1. What do you think of the new ________? — 1 punto

- teachers
- T-shirts

2. He asked if it could be given in a bit late and I said ________ was OK. — 1 punto

- yes, today
- yesterday

3. I think Agassi's winning it ________. — 1 punto

- to love
- two-love

4. I'd have ________ to help him. — 1 punto

- lied
- liked

5. At ________ you understand what I mean. — 1 punto

- least
- last

6. I think she lives at number ________. — 1 punto

- 68
- 60A

Note. Adapted from the listening component of the Oxford Placement Test 2 paper-based version by Oxford University Press. Access the test on the following link: https://forms.gle/WpPb52aJRwWZU67u7

APPENDIX D: LIST OF 36 SPANISH NON-WORDS

Grupo final de estímulos de pseudopalabras					
Pseudopalabras (con acento gráfico)	*N° de sílabas*	*Grado de similitud a palabras reales*	*Patrón acentual*	*Probabilidad fonotáctica esperada*	*Calificación de similitud a palabras reales*
daquiá	2	Bajo	Aguda	-6,03	1,59
ditriá	2	Bajo	Aguda	-6,65	1,88
clastiá	2	Bajo	Aguda	-8,71	1,88
questá	2	Alto	Aguda	-5,63	4,00
fliró	2	Alto	Aguda	-7,11	4,06
armér	2	Alto	Aguda	-5,00	4,38
leisquebé	3	Bajo	Aguda	-11,56	1,53
ebliñá	3	Bajo	Aguda	-8,56	1,59
dasaniá	3	Bajo	Aguda	-7,46	1,59
oridá	3	Alto	Aguda	-6,72	3,00
jomortá	3	Alto	Aguda	-9,09	3,35
tacrimá	3	Alto	Aguda	-8,96	3,41
bisesisblá	4	Bajo	Aguda	-9,36	1,12
miniecrilá	4	Bajo	Aguda	-10,13	1,18
quietolafó	4	Bajo	Aguda	-9,03	1,24
entraresná	4	Alto	Aguda	-8,16	2,24
matrodendá	4	Alto	Aguda	-9,22	2,59
elenisó	4	Alto	Aguda	-5,89	3,06
chínso	2	Bajo	Grave	-5,01	2,00
cáxtar	2	Bajo	Grave	-7,57	2,29
sióga	2	Bajo	Grave	-5,45	2,47
córur	2	Alto	Grave	-6,97	5,35
dráso	2	Alto	Grave	-5,02	5,53
cónstra	2	Alto	Grave	-7,08	6,71
girnústa	3	Bajo	Grave	-9,63	1,71
lerpébar	3	Bajo	Grave	-10,39	1,76
biebáplio	3	Bajo	Grave	-9,48	1,82
autébo	3	Alto	Grave	-6,10	4,18
pablóña	3	Alto	Grave	-8,33	4,59
porguénia	3	Alto	Grave	-8,13	5,53
pasneisódo	4	Bajo	Grave	-10,66	1,18
bieñobléndil	4	Bajo	Grave	-11,85	1,18
predalóño	4	Bajo	Grave	-8,70	1,47
cobrimáno	4	Alto	Grave	-8,21	3,65
fentraráste	4	Alto	Grave	-9,56	4,29
onuríso	4	Alto	Grave	-6,81	4,29

Note. From “La memoria operativa fonológica en aprendices de español como lengua extranjera: propuesta para el diseño de un instrumento,” by Oportus and Ferreira, 2015.

APPENDIX E: WH-QUESTIONS FOR PRODUCTION TRAINING

PRODUCTION TRAINING 1:

1. WHO IS FIXING YOUR LAPTOP? ↘
2. WHAT DID YOU BORROW FROM THE LIBRARY? ↗
3. WHERE DOES THE BUS STOP? ↗
4. HOW DO YOU FOLD THE SHEETS? ↘
5. WHAT ARE YOU SHARING WITH THE GROUP? ↘
6. WHO IS OUR NEXT GUEST? ↗
7. WHERE DID SHE PARK HER TAXI? ↘
8. WHAT LANGUAGES DO YOU SPEAK? ↗
9. HOW DID THE FIRE START? ↘
10. WHO IS SETTING THE TABLE? ↗
11. WHY IS THE KID WASHING HIS HANDS? ↘
12. WHY ARE YOU CHANGING JOBS? ↘
13. WHEN DID YOU RUN THE RACE? ↗
14. HOW IS SHE LIVING HER LIFE? ↘
15. WHEN DO PEOPLE SING CAROLS? ↘
16. WHAT DID THE TEACHER ASK? (show interest)
17. WHAT DID THE TEACHER REPLY? (request information)
18. WHAT DO THE LYRICS MEAN? (show interest)
19. WHERE DO YOU GET INSPIRATION? (sound respectful)
20. WHEN ARE YOU SEEING YOUR PATIENT? (sound kind)

PRODUCTION TRAINING 2:

1. WHO DID YOU GIVE THE GIFT? ↘
2. WHEN DOES THE ALBUM DROP? ↗
3. WHAT'S YOUR EMAIL? ↗
4. WHY DID YOU HIT YOUR COUSIN? ↘
5. WHERE ARE YOU HIDING THE MONEY? ↗
6. HOW DOES THE MEAL TASTE? ↘
7. WHAT DOES THE BIBLE SAY? ↗
8. HOW DID THE PLANE CRASH? ↘
9. WHICH DID SHE PICK? ↗
10. WHEN ARE THEY COMING HOME? ↗
11. WHERE DID YOU LEAVE YOUR HEART? ↘
12. WHO ARE YOUR CLIENTS? ↗
13. WHY DOES THE UNIVERSE EXIST? ↘
14. HOW DID GLOBAL WARMING HAPPEN? ↘
15. WHEN DID EINSTEIN PUBLISH HIS THEORY? ↗
16. WHAT ARE YOU CLEANING WITH THAT THING? (request information)
17. HOW DOES YOUR BACK FEEL? (show sympathy)
18. HOW'S YOUR DAY GOING? (sound respectful)
19. WHERE'S YOUR OFFICE? (request information)
20. WHAT CAN I DO TO HELP? (sound kind)

PRODUCTION TRAINING 3:

1. WHERE ARE YOU HAVING LUNCH? ↘
2. WHAT DID YOU LOSE? ↗
3. WHY DOES THE EARTH SPIN? ↗
4. HOW DO YOU PREPARE COOKIES? ↘
5. WHO IS SHE TAKING TO THE PARTY? ↘
6. WHERE DID YOU HURT YOUR LEG? ↗
7. WHY DO BEES MAKE HONEY? ↘
8. HOW MANY TOWNS DO YOU KNOW? ↗
9. WHO IS DECORATING THE TREE? ↘
10. WHERE DID THE MONKEY SIT? ↗
11. WHY ARE YOU CUTTING THE PAPER? ↘
12. WHEN DID YOU LIGHT THE CANDLES? ↘
13. WHICH ROAD DID YOU TAKE? ↗
14. HOW DO YOU FIGHT THE VIRUS? ↘
15. WHY DON'T YOU CONSULT WITH A DOCTOR? ↗
16. HOW ARE YOU EARNING THIS MONEY? (request information)
17. WHO IS LEADING THE WAY? (confirm information)
18. WHEN DID THEY ATTEND THE WEDDING? (show interest)
19. HOW DO YOU FRY AN EGG? (request information)
20. WHAT ERRORS DID YOU SPOT? (sound respectful)

PRODUCTION TRAINING 4:

1. WHERE DID YOU READ THE LETTER? ↘
2. WHEN DO YOU ADD THE BEANS? ↗
3. HOW DID THE DAISIES GROW? ↘
4.WHICH TEAM IS WINNING THE MATCH? ↗
5. HOW DID SHE ACHIEVE HER GOALS? ↘
6. WHICH CAN YOU AFFORD? ↗
7. WHY DID YOU APPLAUD? ↘
8. WHERE DID YOU MAKE COPIES? ↗
9. WHAT IS YOUR SIZE? ↘
10. WHEN DO YOU CHEW GUM? ↗
11. WHY ARE YOU TIPPING THE WAITER? ↘
12. HOW DOES YOUR DAD SNORE? ↘
13. WHICH GAME DO YOU HAVE? ↗
14.WHO DID YOU ASK TO COME? ↘
15. WHAT IS YOUR DREAM? ↗
16. WHY IS HE PUNCHING THE PILLOWS? (surprised)
17. WHO IS DUSTING THE SHELVES? (request information)
18. WHEN DID YOU PACK YOUR SUITCASE? (show interest)
19. HOW DID YOU BLOW YOUR NOSE? (confirm information)
20. WHERE DO I PLUG IN THIS CABLE? (request information)

APPENDIX F: EXAMPLE OF PERCEPTION TEST ON TP PLATFORM