Perception and production training effects on production of English lexical schwa by young Spanish learners

Esther Gómez-Lacabex (corresponding author)
University of the Basque Country (UPV/EHU), Spain

Francisco Gallardo-del-Puerto
University of Cantabria (UC), Spain

Jian Gong
School of Foreign Languages, Jiangsu University of Science and Technology, China

abstract

Phonetic training has been found to expedite aural and oral abilities in the L2. While considerable research has been conducted on the effects of perception training on production and of production training on perception, fewer studies have addressed them as separate training regimes in the same experimental setting outside laboratory conditions. This paper examined the effects of two training procedures (one based on production tasks and one based on perception tasks) on the production of English lexical *schwa* by young Spanish learners in their intact EFL classrooms. Both trained groups exhibited significant gains in the post-test and a slight advantage of the production-based trained group was observed. Learners' orosensory awareness, self-perception, and self-feedback were actions included in this protocol which may have contributed to such advantage. Our results demonstrate that guided pronunciation training protocols can be successful in the classroom with young learners to boost production skills.

Key words: *second language speech, perception training, production training, English lexical schwa, L2 pronunciation.*

Introduction

Intervention on L2 learners' speech perception and production abilities has been researched from a phonetic approach concerned with whether training facilitates the development of new phonemic categories, frequently singling out perception and production skills and with developed protocols such as High Variability Phonetic Training (HVPT), based on phonetic tenets of repetition and variability, for example. Along these lines, considerable research demonstrates the effects of perception-based phonetic training protocols on speech perception skills (see review article by Bradlow, 2008, for example). To date, we can also refer to a growing number of training studies exploring the effects of perception training on production skills (see recent meta study by Sakai and Moorman, 2018). While many of these studies owed their design to the premise that speech perception is a precursor to speech production (Flege, 1995), a recent focus on speech production seems to indicate that gestural knowledge (Galantucci, Fowler, & Turvey, 2006) and visual and articulatory knowledge (Hazan et al., 2006) mediate in speech perception. Hence, the effects of production-based training protocols on speech perception and production (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015; Wong, 2013) also become relevant to explore.

Another good number of studies have investigated the administration of pronunciation training or instruction with the purpose of integrating the development of L2 pronunciation in the learning context, outside the controlled environment of the phonetics laboratory. This line of research presents varied methods and results (Lee, Jang, & Plonsky, 2015; Thomson & Derwing, 2015) as it has explored different training regimes (Derwing, Munro, & Wieve, 1998), learning paradigms (Saito, 2012;

Saito, 2013; Mora & Levkina, 2017) in classroom settings, or the integration of technology for pronunciation teaching in Computer Assisted Pronunciation Teaching (CAPT) (Fouz-González, 2015; Rogerson-Revell, 2021).

As the call for integrating pronunciation in the language curriculum is still being made (Murphy, 2018; O'Brien et al., 2018; Pennington, 2021), the present study explores the administration of two differentiated phonetic training protocols in a classroom setting, both digitally designed and delivered within an official program to boost the use of individual laptops in the school with young learners. It explores the merging of more rigid phonetic protocols with pronunciation teaching mediated via technology in a quasi-experimental design, with intact learner groups. More particularly, it explores the effects on young students' production skills of perception training, based on discrimination and identification tasks, and production training, based on listen-and-repeat tasks with self-feedback. This is a dimension which remained unexplored in previous work by the first two authors of the present paper (Gómez-Lacabex & Gallardo-del-Puerto, 2014), which focused on the effects of these two different regimes on perception abilities.

Perception and production training effects within the same experimental design

There is a handful of studies which have investigated the effects of perception and production training regimes in the same experimental design. Some have studied their effects when administered separately (Catford & Pisoni, 1970; Leather, 1996; Cibelli, 2022) and the effects of combining perception and production training have also been explored (Wong, 2013; Herd et al., 2013). These studies' methodological approach can shed light into the possible within- and cross-modal effects of perception and production training and generates interesting pedagogical implications. In a first set of studies, perception and production are carefully monitored into two different trainings as in Catford and Pisoni (1970), which used intense listening techniques for perception and articulatory descriptions and silent mimicry for production training, for example. In a further attempt to isolate both skills in training tasks, Leather (1996) based his perception training on computer-based aural input and used visual feedback for the production based protocol to train Dutch speakers on Chinese lexical tones. More recently, Cibelli (2022) has also examined differences between a perception protocol based on AX discrimination tasks with feedback and a production protocol based on explicit gestural explanations, repetition and visual cues training English speakers learning Hindi coronal stops.

Another small handful of studies has explored a further 'combined' experimental scenario, in which, apart from isolating a perception-based and a production-based training protocol, they have explored the combination of both modalities. Wong (2013) compared three parallel training modes to teach English /i/ and /ɪ/ to Cantonese speakers. One group received High Variability Phonetic Training (HVPT) (perception), another group underwent articulatory training (production) and a third group experienced both perception and articulation training. The study corroborated within-modal effects but did not show full cross-modal effects as the group having received articulatory (production) training did not improve perception skills. The study also showed that the dual skill exhibited greater gains in production skills. In the same vein, Herd, Jongman and Sereno (2013) provided perceptual, articulatory (controlled for perceptual cues) and combined training protocols on Spanish /d/ vs. /ɾ/ for English speakers, also supporting within- and cross-modal effects and, interestingly, also showing that the combined training favoured mainly production skills.

A final set of studies conceptualised in more learning-oriented contexts have sought to adapt the protocols to classroom practice. Aliaga-Garcia and Mora (2009) explored the effects of a multi-task protocol involving both perception and production practices such as identification, discrimination, phonetic transcription, listen-and-repeat and read-aloud on English vowels and VOT for Spanish/Catalan speakers, showing moderately different degrees of perception and production improvement. Gómez-Lacabex (2009) investigated the effects of two distinct trainings on young Spanish learners in their EFL lessons. One was based on perceptual identification and discrimination tasks, limiting oral practice, and another one based on articulatory and visual cues, limiting aural practice. Both groups improved their abilities to identify and produce a schwa in lexical words, the production-based trained group scoring higher in both skills. In subsequent studies, Gómez-Lacabex and Gallardo-del-Puerto (2014, 2020) also studied the effects of two phonetic training procedures based on either production (with self-feedback) or perception skills on perceptual awareness of English lexical schwa. No differences were found between the two experimental groups of young Spanish learners, both exhibiting similar improvement in perception.

In short, despite the variability in training protocols, duration of the training protocols or the phonological contrasts explored so far, most of the studies reviewed above seem to support that learning in one domain facilitates improvement in the other modality and that no clear significant advantage of one regime type over the other has been observed so far. The present study aims at further exploring perception and production training outside laboratory conditions with younger

Spanish learners technically supported in their EFL school lessons, with the purpose of investigating the effects of such trainings in the reality of the foreign language classroom, which has been less observed in the literature so far and has relevant pedagogical implications.

Research Questions

The present work is a follow up study of Gómez-Lacabex and Gallardo-del-Puerto (2014, 2020), which revealed a facilitative effect of two differentiated trainings for perceptual awareness of English lexical schwa (*schwa* as the nucleus of the unstressed syllable of English content words). It explores the effects of perception and production training on production skills in the same experiment.

Although the English unrounded mid-central vowel, or *schwa,* has traditionally been described as a phoneme (Cruttenden, 1994; Roach, 2000; Yavaş, 2006), its acoustic, articulatory and durational fluctuations are difficult to perceive and produce in the case of learners whose L1 does not include central vowel phonemes, as is the case of Spanish (Delattre 1969; Hualde, 2005). In addition, English and Spanish have been described as exhibiting different prosodic shapes. Fundamental frequency, intensity, duration and vowel reduction are the correlates frequently ascribed to English stress (Fry, 1955; Laver, 1994); in Spanish, vowel reduction has been described to have a minor role as a stress correlate (Ortega-Llebraria & Prieto, 2011; Hualde, 2005). Despite recent research having revealed that Spanish exhibits phonetic vowel reduction processes (Cobb & Simonet, 2015; Delforge, 2008; Ronquest, 2013), these weakening realizations are allophonic. Hence, Spanish speakers are not used to attending to English vowel reduction phonemically (Broś, 2015; Gómez-Lacabex, García Lecumberri & Cooke, 2009) and do not tend to incorporate it in their production (Brown, 1990; Flege & Bohn, 1989; Ikeno et al., 2003; Rallo-Fabra, 2015). In addition, orthographic depth of the extent at which alphabetical languages more or less consistently associate letters and phones (Katz & Frost, 1992) has also given rise to the notion of transparent and opaque languages. The tendency to consistently associate a grapheme to a phone in transparent languages such as Spanish provokes mispronunciations when the target sound, in this case English schwa, exhibits multiple graphemic representations in an opaque system (Rallo-Fabra & Jacob, 2016).

We addressed the following research questions:

1. Do perception and production training impact young learners' production of English lexical schwa?
2. Are there any training effect differences between the perception and production regimes?

Limited work has been conducted regarding the effects of phonetic training on vowel reduction with younger learners in the EFL classroom. In a study by Gómez-Lacabex (2009), the group undergoing production-based training tended to outscore the perception-based trained group in both skills, but the subjects in Gómez-Lacabex and Gallardo-del-Puerto (2014, 2020) showed no training modality effect in their perception skills.

We predict that both protocols will impact these students' production of English schwa, showing within training effects for production and across training effects for perception. Further, we wish to investigate whether there might be differences in the production of lexical schwa after the two differentiated perception and production trainings.

Method

Participants

Seventy-five Spanish EFL 6th graders (aged 12) in three intact groups of 25 students each took part in the experiment. Two of these groups were assigned two types of training: an AURAL group (10 boys and 15 girls) underwent perception tasks while an ORAL group (8 boys and 17 girls) underwent listen-and-repeat practice. The third group was treated as Control (9 boys and 16 girls) and received no training during the experimental period. Instead, this group was enrolled in a project on developing reading strategies on-line. All the groups had started learning English in school at the age of 3. They were being taught three hours of English as a foreign language per week. Additionally, the school was enrolled in a language program that integrated three languages in the school curriculum (most) frequently in fifth and sixth grades: Spanish, Basque (both official in the region) and English. The disciplines being taught in English were English, Arts and Crafts and Physical Education. Science was only taught in English for one term each year. The teachers were non-native speakers of English with C1 proficiency level. The learners' language level was A2/pre-intermediate[i].

Training stimuli

Fifty two-syllable words including a lexical *schwa* were selected in pre-tonic (*a̲larm*) and post-tonic (*sala̲d*) unstressed positions for both trainings (see Appendix A). *Schwa*/full vowel minimal pairs were created by splicing mimicked Spanish full vowel and genuine English *schwa* and pasting them to the same baseline for each informant's word. The items were recorded in a soundproof booth by three balanced (British) English/Spanish bilinguals (two female, one male) who were experienced

EFL teachers living in Spain. These speakers first produced the English words (presented to them orthographically) twice. Next, they said the same words mimicking typical Spanish-accented English pronunciation by trying to colour the schwas toward Spanish-like vowel qualities (e.g., *salad* /sal<u>a</u>d/, *summon* /sam<u>o</u>n/). A native Spanish speaker highly proficient in English and with a phonetic background supervised the recording. *Schwa*/full vowel minimal pairs were created by extracting the mimicked Spanish full vowel and the *schwa* from the second production of each token and inserting these vowels into the same baseline target word using Praat (Boersma & Weenink, 2019). For the sake of clarity, Table 1 displays the acoustic values of some of the schwa vs. full vowel contrasts.

Table 1

F1, F2 and duration values of schwa vowels and full vowels for the testing sample

Item	Schwa vowels			Full vowels		
	F1	F2	duration	F1	F2	duration
about	452	1592	0.035	1033	1754	0.103
ago	548	1711	0.063	916	1222	0.076
agree	421	1416	0.036	916	1222	0.076
alone	565	1881	0.071	905	1810	0.080
allow	590	1907	0.029	991	1851	0.092
amend	795	1764	0.095	984	1712	0.083
basin	536	1845	0.067	532	2377	0.081
Boston	584	1826	0.072	628	1128	0.116
cloven	676	1787	0.068	702	2125	0.113
cousin	559	1749	0.090	493	2335	0.123
felon	738	1854	0.091	615	1196	0.086
fireman	667	1897	0.082	917	1847	0.159
Helen	704	1935	0.093	705	2284	0.098
Ivan	689	1814	0.091	985	1804	0.158
Jackson	615	1759	0.073	692	1234	0.104
pitman	566	1903	0.048	763	1953	0.101
raisin	444	1832	0.065	594	2506	0.086
random	599	1809	0.104	560	1241	0.113
raven	657	1955	0.102	698	2105	0.105
season	600	1880	0.082	712	1517	0.101
seven	693	1880	0.073	723	2070	0.087
starlet	655	1793	0.065	759	2148	0.073
summon	680	1520	0.069	562	976	0.122

woman	620	1697	0.094	830	1733	0.107

Note. These are the production values of the 24 items in the female voice used for training.

All audio files were normalized to the same Root Mean Squared level. The audio input chosen for the testing sessions was one of the female voices. The audio files recorded by the two other speakers were used in the training sessions to avoid speaker-specific adaptation in the post-test, hence, testing generalization to a new voice. Generalization to new stimuli was not incorporated in this study.The items were equally distributed across training sessions and tasks according to speaker voice (male/female). The students also completed a word-familiarity questionnaire prior to the pre-test. They ticked whether the words shown in a list (selected items) were known to them or not. As the list included some proper names, students were asked to tick "yes" if they identified these as such. This information was used to distribute the items equally in each training session by means of the median value of familiar/unfamiliar. It is worth noting that the median value was used so that the items would be evenly distributed into training sessions. Hence, familiarity was not considered categorically in this design.

Training procedures

So as to adapt the two training protocols to the ecology of young students in the classroom, they were designed to blend different techniques used in phonetic training research for each protocol. Perception training included discrimination and identification tasks, while production training used listen-and-repeat techniques along with an articulatory and a visual tip. Sessions were programmed to last approximately 30 minutes and were delivered once a week during the English lesson in consecutive weeks during the school's second term. Prior to the differentiated protocols, both groups received two preparatory sessions so that the students became familiarized with the research team (as suggested by the institution) and in which a research team member with teaching experience presented the *schwa* along with the concepts of syllable, stressed syllable, unstressed syllable and multiple spellings. The modelling of *schwa* was avoided during the delivery of these sessions for the listen-and-repeat group. In addition, this group was provided with two pronunciation tips: one articulatory: 'when you say the underlined vowel, shape your lips for the following consonant" and one visual: 'imagine that the underline rises and crosses the vowel so that it is not pronounced'. These tips were not provided to the perception-based trained group.

As seen in Table 2, the perception training regime (AURAL group) presented 62 trials in every session in the following sequence i) same/different discrimination: students listen to two stimuli and decide whether they are same or different; ii) odd-one-out: students listen to three stimuli and identify the different one; iii) correct/incorrect identification: students listen to one stimulus and decide whether it is in/correctly pronounced: iv) Correct/incorrect identification with confidence rating tasks in four degrees (see Appendix B for captions of these interfaces). Feedback was delivered by flashing a happy smile emoji and providing a cumulative star every time a trial was correct in the first attempt.

Table 2

Distribution of items into AURAL and ORAL tasks and week

week	AURAL tasks	ORAL task

	Same/ different	Oddity	(In)Correct	(In)Correct +Confidence	Listen and repeat +self-feedback
1	12	13	25	12	24 (X2/3)
2	13	12	25	13	24 (X2/3)
3	12	13	25	12	24 (X2/3)
4	13	12	25	13	24 (X2/3)
Total items	50	50	50	50	96 (X2/3)

The regime for the ORAL group was based on imitation practice and self-feedback. Students heard the word on the screen and repeated it in a Matlab interface, which recorded their productions. Self-feedback was included as part of the protocol and the program forced learners to listen to their productions once (see Appendix B for captions of the interface). As we tried to balance the number of trials per week for each protocol, we arranged that they should listen to their recordings once and re-record if not satisfied. Hence, potentially, they could attempt three production trials, with a minimum of 48 trials and a maximum of 72.

Production testing

There were 24 items selected for testing. Word familiarity distribution (Appendix A) and cognate status were controlled. Four words with full-vowel in the unstressed syllable were included as control stimuli (*grandson*, *resin*, *jurat*, *paypal*). Productions were collected in a sound-proof booth two weeks before and after the training period. A custom-designed Maltab program administered the pre-test and post-test interfaces, in which the learners saw the word they had to produce in written form.

Auditory assessment of productions

Productions were presented to a group of eight non-native listeners to be assessed auditorily. They spoke Spanish as a first language and were linguistically and phonetically trained. Six were English Studies postgraduate students and two were qualified EFL language teachers. We relied on this listener profile as previous work has shown that trained non-native listeners can successfully identify Foreign Accent (Gallardo-del-Puerto et al., 2015; Major, 2007). They each were pre-tested (using minimal pairs) on their ability to acoustically distinguish a reduced vowel (*schwa*-like and or deleted vowel) against a full vowel in English lexical words and later received a light training on the task. A new Matlab interface presented the listeners with a three forced-choice task. Judges were asked to listen to each word, paying attention to the unstressed syllable only, which was identified (underlined) for them in their interface (*raven*). They were also individually familiarized with the assessing procedure: i) click on 'weak' if you hear a reduced vowel ii) click on 'strong' if the vowel you hear is not a reduced vowel. Additionally, the choice 'unassessable' was also provided in case a judgement could not be provided due to unintelligible pronunciation, wrong stressing or sound/syllable deletion. Two native speaker voices were added as controls. The program automatically presented the randomized stimuli in 5-sessions blocks, which the listeners took on five different days. This identification task resulted in 14,448 observations (75 students × 24 stimuli + 48 stimuli in 2 native voices× 8 judges).

Results

Results first present two inter-rater reliability analyses: mean agreement (Table 3) and intra-class correlation (Table 4) to explore the degree of agreement among the judges and discard any possible outlying performance. The two analyses indicated that the group of listeners was homogeneous and reliable in their assessing performance. Production results are presented for overall training effects (Table 5) as average ratings (0-10) for the experimental groups (AURAL and ORAL) and the Control group. Variables were normally distributed, for which ANOVA and t-tests were conducted. The confidence interval of the ANOVA tests was 95%. The alpha level of significance was set at 0.05. Partial eta-squared (η^2) values were used as effect size, for which magnitudes of .01, .06 and .14 are considered as small, medium and large, respectively (Cohen, 1973; Richardson, 2011). One-way ANOVA and Bonferroni inter-group tests revealed that there were no significant differences among the three groups ($F(2.71) = 1.556$, $p = 0.218$, $\eta^2 = .053$), indicating that they were suitable for comparison at the outset of the study.

Inter-rater agreement

Inter-judge agreement was analysed by means of the number of judges who agreed on the most repeated value (mode). The mean agreement (1–8) among the 8 judges for each test was high for all the contexts analysed (Table 3). We incorporated a further control measure by adding two native voices to the stimuli in the pre-test phase, which were also identified fairly accurately by the raters. This indicated that the raters were able to perform it.

Table 3

Means and standard deviations results for all speakers' groups: AURAL and ORAL, Control and Native control speakers

Mean agreement for 8 judges (1 - 8)	Pre-test	S.D.	Post-test	S.D.
AURAL	7.52	0.65	7	1.26
ORAL	6.36	1.63	6.56	1.29
Control	7.04	1.12	6.92	1.41

Overall mean	7	1.28	6.82	1.32
Native speakers	8	0	-	-

Note. S.D. indicates standard deviation. Two native speaker voices were used.

Intra-class correlations (ICC) were also calculated. Modes (most frequent value among the 8 judges being 0-strong, and 1-weak) for each word for each subject were calculated. Average ratings from the 8 judges for each student across all items were then calculated and computed to a 0 to 10 scale [ii] (see Table 4). For pre-test, which included the 2 native voices (n = 76), ICC estimates were based on a mean-rating (k = 8), absolute-agreement, 2-way mixed-effects model: ICC (3, 8) = 0.937, $p < .001$, inter-rater agreement reliability was very high. For post-test, which excluded the 2 native voices (n = 74), ICC estimates were also very high: ICC (3, 8)= 0.957, $p < .001$. We also calculated Cronbach´s alpha values for pre-test (0.974) and post-test ratings (0.984). These results indicated that using the average ratings from the 8 non-native speakers was a reliable measure in this study.

Table 4

Average ratings and standard deviations for all speakers' groups: AURAL and ORAL, Control and Native control speakers and pre-post difference

Average ratings (0-10)	Pre-test	S.D.	Post-test	S.D.	Pre-post difference (paired-samples t-test)
AURAL	2.976	1.37	4.700	2.38	$t(24) = -3.557$ $p < 0.01$
ORAL	3.664	1.64	5.700	1.85	$t(24) = -6.559$ $p < 0.001$
Control	3.77	1.55	3.729	1.84	$t(23) = -0.683$ $p = 0.51$
Overall rating	3.47	1.5	4.72	2.2	
Native	9.6	0	-	-	

Note. S.D. indicates standard deviation. Two native speaker voices were used.

Training effects

Table 5 shows average ratings for the three groups in pre-test, post-test and gain scores. Complementarily, data spread at the two testing phases can be seen in Figure 1. Repeated measures ANOVA, with ‘group’ as between subject factor, ‘time’ as within-subject factor, showed significant ‘time effect [$F(1.71) = 39.361$, $p < .001$, $\eta^2 = .357$] and significant ‘time-by-group’ interaction [$F(2.71) = 10.582$, $p < .001$, $\eta^2 = .230$], indicating that there were differences between pre-test and post-test stages and these were different for the groups. Besides, the effect sizes for these differences were of a large magnitude. Analyses also revealed that the ‘group’ effect was not significant [$F(2.71) = 2.607$, $p = .081$, $\eta^2 = .068$].

For the significant ‘time*group’ interaction, further simple effect analyses (Bonferroni adjustment) showed that AURAL and ORAL groups’ ‘time’ effects were significant ($p < .001$), the Control group’s ‘time’ effect not being significant ($p = .905$). This indicated that the training affected the experimental groups in one way, by experiencing an increase in their score, and the Control group in a different way, by not undergoing a significant change between the two testing phases. Simple effect analysis (Bonferroni adjustment) also showed that, in post-test phase, there was no significant difference between AURAL and ORAL and AURAL and Control ($p>.05$), but there was a significant difference between ORAL and Control ($p < .01$). These results indicate that, despite the fact that both experimental groups significantly realised more reduced vowels in the post-test, the Oral group had a more robust training effect, as its performance in post-test distances significantly from that of the Control group. This was corroborated in the post-test, separate one-way ANOVA analyses which also indicated that there were significant differences with a large effect size between the three groups [$F(2.71) = 5.723$, $p < 0.01$, $\eta^2 = .139$]; post-hoc multiple comparisons (Bonferroni) showed that such significant differences were between groups ORAL and Control ($p < .01$) only, and not between AURAL and ORAL ($p>.05$) or AURAL and Control ($p>.05$) groups.

Table 5

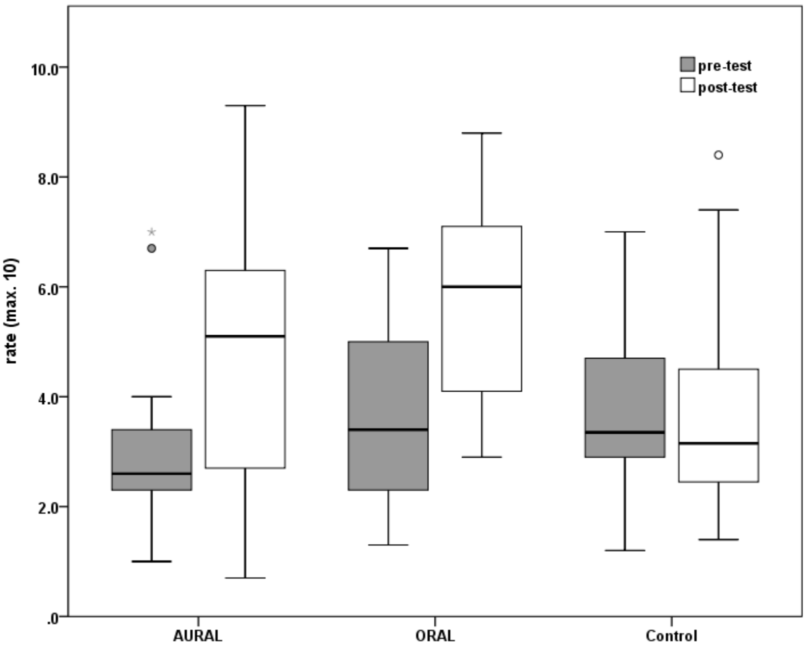
Average ratings and standard deviations for AURAL, ORAL and Control groups, gain scores, for all items (n = 24)

	Pre-test	S.D.	Post-test	S.D.	Gain	S.D.
AURAL	2.98	1.4	4.70	2.4	1.72	2.2
ORAL	3.66	1.6	5.70	1.8	2.04	1.6
Control	3.77	1.6	3.73	1.8	-0.42	1.0

Note. S.D. indicates standard deviation. Table presents average rating for all 24 items (range 0-10).

Figure 1

Average ratings for groups in pre-test and post-test



Note. Average ratings have a maximum of 10 points.

Discussion

This study explored the impact of two differentiated pronunciation trainings on L2 young learners' production of English lexical *schwa* in the classroom. While research in natural settings has shown that younger L2 learners tend to outperform adult L2 learners both in perception and production skills (Tsukada et al., 2005; Jia et al., 2006), research on L2 phonological development in formal learning contexts has found later young learners to exhibit better short and mid-term performance on both perception and production skills than earlier young learners (Fullana, 2006; García Lecumberri & Gallardo-del-Puerto, 2003), suggesting that 'the earlier, the better' may not be warranted in this population. Indeed, the development of L2 pronunciation in younger populations has received less attention in the literature (Thomson & Derwing, 2014); we even know less about young EFL students in formal learning environments, who may have been surveyed as in Tergujeff (2013) in Finland or Calvo Benzies (2016) in Spain, but much less frequently tested on their English pronunciation skills (Fullana, 2006; García Lecumberri & Gallardo-del-Puerto, 2003). Research on phonetic training has also limitedly explored this population (Gómez-Lacabex & Gallardo-del-Puerto, 2014).

Our results indicated that both perception and production trainings led to gains as both experimental groups (AURAL and ORAL) were judged to produce more reduced vowels after their training, with the Control group exhibiting no improvement. First, results for the AURAL group ascribe to those studies which have shown an influence of perception training on production skills (Sakai & Moorman, 2018). In the present experiment, perception training was delivered in the form of a combination of four different discrimination/identification tasks with feedback with images and colours to dynamize the interface, and distributed into 4 sessions to adapt the protocol to the attention span of the learners, which never lasted more than 20 minutes. Second, results from the ORAL group contribute to the less copious literature which has observed that production training can boost production skills (Flege, 1988; Delvaux et al., 2013; Cucciarini et al., 2009). In addition, the group having received production-based training always outscored the perception-based trained group, becoming significantly distant from the Control group. We highlight three characteristics of the out-of-laboratory production-based regime in the present study. First, it was based on auditory self-perception, that is, the learners had the chance to listen to their own performance and, in addition, it presented opportunities for repetition. The protocol may have allowed the learners to develop their proprioception or 'their own awareness of the position of their articulators' (adapted from Trask, 1996) by observing and exploring their own *schwa* (mis)pronunciations. Indeed, self-speech perception has been found to play "a significant role in the planning and execution of future speech production" (Casserly & Pissoni, 2010, p. 642). In line with this, these learners were also given an articulatory tip in the introduction session, which may have further stimulated some orosensory awareness. Second, the learners were encouraged to judge their own pronunciation of *schwa*, which may have activated self-corrective feedback skills (Lee & Lyster, 2015; Saito & Lyster, 2012). The fact that they were able to improve their performance on the basis of their own (mis)pronunciations, suggests that self-feedback may be valuable during pronunciation training. Finally, the possibilities for self-repetition and more management for self-pace may have contributed to a more stress-free and less rigid practice experience. Further research is indeed welcome so as to explore such aspects more experimentally.

This study adds to the phonetic instruction studies which have explored perception and production training protocols within the same experimental design, showing a slight advantage for production training over perception training protocol on production skills. As the perception-production link still receives attention in the literature, recent production-based phonetic training protocols have exploited varied techniques such as intonation contour/soundwave displays (Chun et al., 2015) or spectrograms (Patten & Edmonds, 2015) and vowel formant information (Kartushina et al., 2015). However, when technical interfaces are confronted outside laboratory conditions, the average language practitioner and learner may find it difficult to interpret them. Mimicking and/or imitation techniques (Delvaux et al., 2013), providing articulatory cues (Wong, 2013) or activating production feedback are some of the procedures which were used in this work in the interest of the group's ecological validity. These procedures have shown that production-based protocols can produce higher gains in production than a perceptual training regime, also seen in Gómez-Lacabex (2009), suggesting that perception is necessary but maybe not sufficient for production (Kartushina et al., 2015). As more populations are investigated such as young learners in formal learning contexts for whom production and perception skills are not expected to develop 'early' (Fullana, 2006; García Lecumberri & Gallardo-del-Puerto, 2003), or heritage speakers, whose production can lag behind native-like perception skills (Jun, Knightly, and Au, 2003), we shall continue to unravel the link between L2 speech perception and production.

While we avoided voice specification effects by including different speakers in testing and in training, we must acknowledge that we did not test generalization effects, with new voices or new words, for example, nor were we able to conduct delayed-post testing to explore durability effects. Furthermore, this study highlighted training perception and production training effects unimodally. It remains to be explored whether integrated training in which both perception and production skills are combined (Wong, 2013; Herd et al., 2013) produces greater gains in either the production (or perception) of schwa with this population.

Conclusions and pedagogical implications

The study suggests interesting pedagogical implications. First, it acknowledges the effects of pronunciation intervention with younger learners in the L2 classroom, indicating that technically supported practice is also possible in this learning environment. Second, the production-based training administered in this study produced slightly greater gains. Some of the characteristics of this protocol such as the role of the learner in the practice, who listened to their own voice and provided self-feedback, may be exploited along with peer, tutor or computer feedback. Finally, the technological management which each student experienced also deserves exploration as, to date, the supply of individual educational technical equipment is not a universal asset in education, which, along with the appealing effects of technology among younger generations, may have contributed to a more motivating learning experience.

References

Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.04, retrieved 28 September 2019 from http://www.praat.org/

Bradlow, A. R. (2008). Training non-native language sound patterns: lessons from training Japanese adults on the English. (pp.287-308). In J. G. H., Edwards, & M. L., Zampini, (Eds.), *Phonology and second language acquisition* (Vol. 36). John Benjamins Publishing.

Brown, R.L., (1990). A Maverick vowel: The notorious /ə/ (schwa). *Hispania*, *73* (4), 1158-1161.

Calvo Benzies, Y. J. (2016). *The teaching and learning of English pronunciation in Spain. An analysis and appraisal of students' and teachers' views and teaching materials*. Unpublished doctoral dissertation. Universidad de Santiago de Compostela, Spain. http://hdl.handle.net/10347/13862

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 629-647.

Catford, J. C., & Pisoni, D. (1970). Auditory versus articulatory training in exotic sounds. *The Modern Language Journal, 54*, 477-481.

Chun, D. M., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation*, *1*(1), 86-114. https://doi.org/10.1075/jslp.1.1.04chu.

Cibelli, E. (2022). Articulatory and perceptual cues to non-native phoneme perception: Cross-modal training for early learners. *Second Language Research*, *38*(1), 117-147. https://doi.org/10.1177/0267658320921217

Cobb, K. & Simonet, M. (2015). Adult Second Language Learning of Spanish Vowels. *Hispania, 98*(1), 47-60.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and psychological measurement*, *33*(1), 107-112.

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics*, *6*(4), 295-323.

Delforge, A. M. (2008). Unstressed vowel reduction in Andean Spanish. In *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. (pp. 107-124).

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2013, August). Production training in second language acquisition: a comparison between objective measures and subjective judgments. Proceedings of XIVth *Interpeech*. Lyon, France. (pp. 2375-2379).

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language learning*, *48*(3), 393-410. https://doi.org/10.1111/0023-8333.00047.

Flege, J. E. (1988). Using Visual Information to Train Foreign-Language Vowel Production. *Language Learning*, *38*(3), 365-407.

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

Flege, J. E., and Bohn, O.-S. (1989). An Instrumental study of vowel reduction and stress placement in Spanish accented English. *Studies in Second Language Acquisition, 11*, 35-62. https://doi.org/10.1017/S0272263100007828.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765-768.

Fouz-González, J. (2015). Trends and directions in computer-assisted pronunciation training. In Mompean, J. A., & Fouz-González, J. (Eds.). (2015). *Investigating English pronunciation: Trends and directions*. (pp. 314-342).Springer.

Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (Ed.), Age and rate of foreign language learning (pp. 41-64). Multilingual Matters

Galantucci, B., Fowler, C.A., & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, *13*(3), 361.

Gallardo-del-Puerto, F., García Lecumberri, M. L., & Gómez-Lacabex, E. (2015). The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges. *International Journal of Applied Linguistics*, *25*(2), 202-224. https://doi.org/10.1111/ijal.12063.

García Lecumberri, M. L., & Gallardo-del-Puerto, F. (2003). English FL sounds in school learners of different ages. In M. P. García Mayo & M. L. García Lecumberri (Eds.), Age and the acquisition of English as a foreign language (pp. 115-135). Multilingual Matters.

Gómez-Lacabex, E. (2009). *Perception and Production of Vowel reduction in Second Language Acquisition*. Serie Tesis Doctorales. University of the Basque Country.

Gómez-Lacabex, E., & Gallardo-del-Puerto, F. (2014). Two Phonetic-Training Procedures for Young Learners: Investigating Instructional Effects on Perceptual Awareness. *Canadian Modern Language Review*, *70*(4). 500-531. https://doi.org/10.3138/cmlr.2324.

Gómez-Lacabex, E., & Gallardo-del-Puerto, F. (2020). Explicit phonetic instruction vs. implicit attention to native exposure: phonological awareness of English schwa in CLIL. *International Review of Applied Linguistics in Language Teaching*, *58*(4), 419-442. https://doi.org/10.1515/iral-2017-0079.

Gómez-Lacabex, E., Garcia Lecumberri, M.L., & Cooke, M. (2009). Training and generalization effects of English Vowel reduction for Spanish Learners. In: M. A. Watkins, A. S. Rauber, and B. O. Baptista (Eds.), *Recent Research in Second language Phonetics/Phonology*: *Perception and Production*. (pp. 32-42) New Castle upon Tyne: Cambridge Scholars Publishing.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740-1751.

Herd, W., Jongman, A., & Sereno, J. (2013). Perceptual and production training of intervocalic/d, ɾ, r/ in American English learners of Spanish. *The Journal of the Acoustical Society of America*, *133*(6), 4247-4255. https://doi.org/10.1121/1.4802902

Hualde, J. I. (2005). *The Sounds of Spanish*. Cambridge University Press.

Ikeno, A., Pellom, B., Cer, D., Thornton, A., Brenier, J. M., Jurafsky, D., ... & Byrne, W. (2003). Issues in recognition of Spanish-accented spontaneous English. In *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*. pp. 63–66, Tokyo, Japan.

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related difference vary with amount of L2 exposure. *Journal of the Acoustical Society of America, 119*, 1118-1130. https://doi.org/10.1121/1.2151806.

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

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Leather, J. H. (1996). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In James, A. and Leather, J. (Eds.), *Studies on Language Acquisition [SOLA]*. Volume 13: Second-Language Speech. Berlin: De Gruyter Mouton. 75-101.

Lee, A. H., and Lyster, R. (2015). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*, 1-30. https://doi.org/10.1017/S0272263115000194.

Lee, J., Jang, J., & Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366. https://doi.org/10.1093/applin/amu040.

Major, R. C. (2007) Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition 29*, 4, 539–56. https://doi.org/10.1017/S0272263107070428.

Mora, J. C. & Levkina, M. (2017). Training the production of English L2 vowel contrasts through collaborative map tasks. 9th Annual Conference *Pronunciation in Second Language Learning & Teaching* (PSLLT), September 1-2, 2017, University of Utah, Salt Lake City, Utah, USA.

Murphy, J. M. (2018). Teacher training in the teaching of pronunciation. In Kang, O., Thomson, R., & Murphy, J. (Eds.), *The Routledge Handbook of Contemporary English Pronunciation*. New York, NY: Routledge. (pp. 298-319).

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J., Munro, M.J, Foote, J.A. & Levis, G. M (2018). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, *4*(2), 182-207. https://doi.org/10.1075/jslp.17001.obr.

Oh, J. S., Jun, S.-A., Knightly, L. M., & Au, T. K. M. (2003). Holding on to childhood language memory. *Cognition, 86*, B53-B64.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, *54*(1), 73-97. https://doi.org/10.1177/0023830910388014.

Patten, I., & Edmonds, L. A. (2015). Effect of training Japanese L1 speakers in the production of American English/r/using spectrographic visual feedback. *Computer Assisted Language Learning*, *28*(3), 241-259. https://doi.org/10.1080/09588221.2013.839570.

Pennington, M. C. (2021). Teaching Pronunciation: The State of the Art 2021. *RELC Journal*, *52*(1), 3-21. https://doi.org/10.1177/00336882211002283.

Rallo-Fabra, L. (2015). Can nonnative speakers reduce English vowels in a native-like fashion? Evidence from L1-Spanish L2-English bilinguals. *Phonetica*, *72*(2-3), 162-181. https://doi.org/10.1159/000430920.

Rallo-Fabra, L. and Jacob, K. (2016). ‘When ‘politician’ becomes ‘policía’ and ‘border’ becomes ‘brother’: effects of orthography and cognate status on Spanish-Catalan EFL learners’ pronunciation’, poster presented at the *8th International Conference on Second-Language Speech. New Sounds 2016.* Aarhus.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135-147.

Rogerson-Revell, P. M. (2021). Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. *RELC Journal*, *52*(1), 189-205. https://doi.org/10.1177/0033688220977406

Ronquest, R. (2013). An acoustic examination of unstressed vowel reduction in Heritage Spanish. In *Selected proceedings of the 15th Hispanic linguistics symposium*. 151-171.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *Tesol Quarterly*, *46*(4), 842-854. https://doi.org/10.1002/tesq.67.

Saito, K. (2013). Reexamining Effects of Form-Focused Instruction on L2 Pronunciation Development. The role of Explicit Phonetic Information. *Studies in Second language Acquisition 35*, 1-29. https://doi.org/10.1017/S0272263112000666.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language Learning*, *62*(2), 595-633. https://doi.org/10.1111/j.1467-9922.2011.00639.x.

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.

Tergujeff, E. (2013). Learner Perspective on English Pronunciation Teaching in an EFL context. *Research in Language 11*(1), 81-95.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344. https://doi.org/10.1093/applin/amu076.

Trask, R. L. (2004). *A dictionary of phonetics and phonology*. Routledge.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263-290. https://doi.org/10.1016/j.wocn.2004.10.002.

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. Proceedings of XIVth *Interpeech*. Lyon, France (pp. 2113-2117).

Appendix A: Lists of items used in the experiment

List of words used in the training (50):

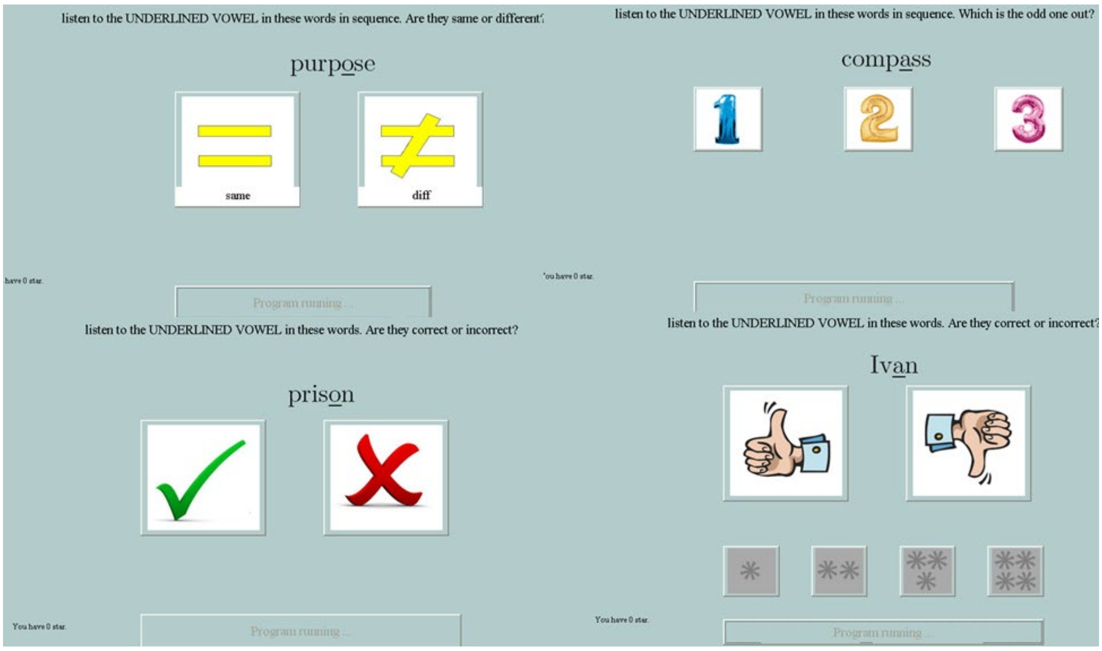
abolish, about, across, afraid, again, ago, agree, ahead, alarm, allow, alone, along, amend, baboon, ballad, balloon, basin, Boston, cloven, colours, command, compass, cousin, craven, datum, felon, fireman, heaven, Helen, Ivan, Jackson, melon, oven, phantom, pilot, pitman, prison, purpose, raisin, random, raven, sailors, salad, salmon, season, seven, starlet, summon, violet, woman.

Lists of words used for testing (24) distributed as more and less familiar:

More familiar words (n = 12): *about, ago, alone, Boston, cousin, fireman, Ivan, Helen, Jackson, season, seven, woman*

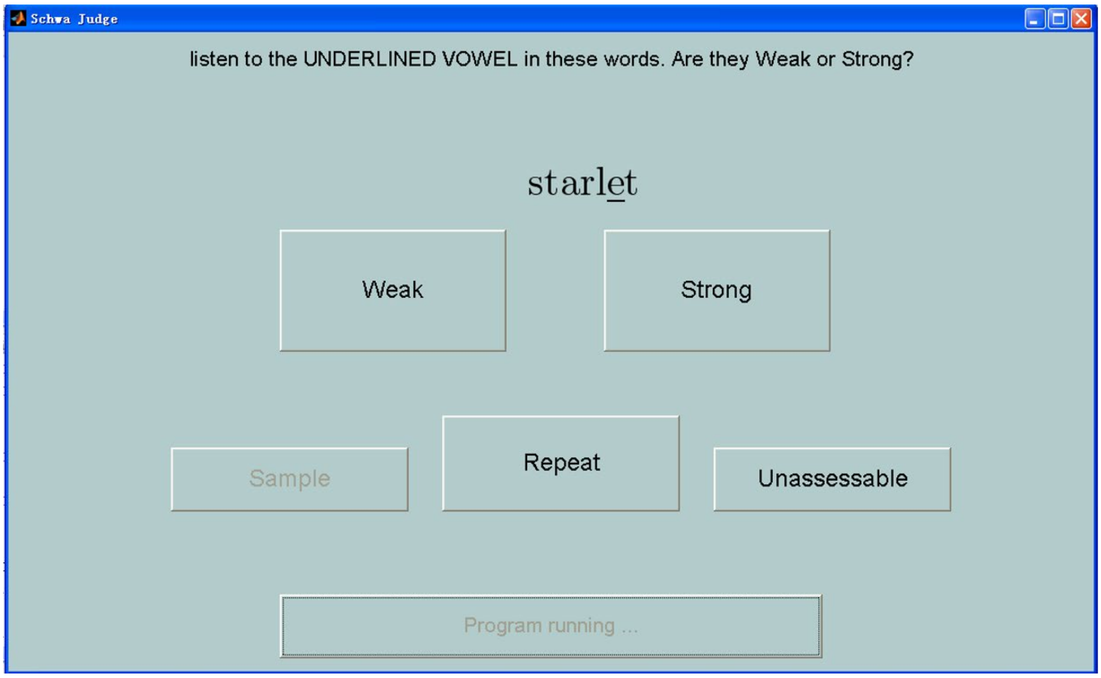
Less familiar words (n = 12): *agree, allow, amend, basin, cloven, felon, pitman, raisin, random, raven, starlet, summon*

Appendix B: screen shots for tasks for AURAL and ORAL groups sessions and production assessment



Screen shot 1: Interfaces for: i) Same/different discrimination task ii) oddity discrimination task, iii) Correct/incorrect identification task and iv) Correct/incorrect identification task with confidence rating for AURAL group.

Screen shot 2: Listen and-repeat interface for ORAL group.



Screen shot 3: production assessment interface for listeners.

[i] As established by the Cambridge Key English Test (KET), which the school delivered during the course.

[ii] We used a mean rating, rather than a binary value, and multiplied by ten to make results more interpretable for the reader.

Esther Gómez-Lacabex (corresponding author)
University of the Basque Country (UPV/EHU), Spain
Department of English and German Philology and Translation and Interpretation.
Faculty of Letters
Paseo de la Universidad 5,
01006 Vitoria-Gasteiz, Spain
tel: 00 34 945013963
e-mail: esther.glacabex@ehu.eus

Francisco Gallardo-del-Puerto
University of Cantabria (UC), Spain
Department of Philology. Edificio de Filología.
Avenida de Los Castros s/n, 39005, Santander, Spain
e-mail: francisco.gallardo@unican.es

Jian Gong
School of Foreign Languages, Jiangsu University of Science and Technology, 2 Mengxi Road, 212003 Zhenjiang, China
habaogj@hotmail.com