

ORIGINAL ARTICLE

# The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology

Bene Bassetti<sup>1\*</sup>, Tania Cerni<sup>2</sup> and Jackie Masterson<sup>3</sup>

<sup>1</sup>School of Education, University of Birmingham, Birmingham B15 2TT, UK, <sup>2</sup>Department of Psychology and Cognitive Science, University of Trento, 38068 Rovereto, Italy and <sup>3</sup>Department of Psychology and Human Development, UCL Institute of Education, London WC1H 0AL, UK

\*Corresponding author. Email: [b.bassetti@bham.ac.uk](mailto:b.bassetti@bham.ac.uk)

(Received 5 May 2021; revised 2 February 2022; accepted 7 February 2022)

## Abstract

The orthographic forms (spellings) of second language (L2) words and sounds affect the pronunciation and awareness of L2 sounds, even after lengthy naturalistic exposure. This study investigated whether instruction could reduce the effects of English orthographic forms on Italian native speakers' pronunciation and awareness of L2 English sounds. Italians perceive, produce, and judge the same sound as a short sound if it is spelled with one letter and as a long sound if it is spelled with a digraph, due to L1 Italian grapheme-phoneme correspondence (GPC) rules whereby double consonant letters represent long consonants. Totally, 100 Italian learners of English were allocated to two conditions (final  $n = 88$ ). The participants in the explicit GPC (EGPC) condition discovered English GPC rules relating to sound length through reflection, explicit teaching, and practice; the participants in the passive exposure condition practiced the same words as the EGPC participants, but with no mention of GPCs. Pre- and postintervention production (delayed word repetition) and phonological awareness (rhyme judgment) tasks revealed no positive effects of the instruction. GPC instruction appears to be ineffective in reducing orthographic effects on L2 phonology. Orthographic effects may be impervious to change, whether by naturalistic exposure or by instruction.

**Keywords:** orthographic effects; second language phonology; pronunciation; phonological awareness; pronunciation instruction; group randomised trial

## Introduction

Recent studies have consistently demonstrated that the orthographic forms (spellings) of speech sounds in a second language (L2) affect L2 speakers' speech production, speech perception, and phonological awareness (for an overview, see Hayes-Harb &

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Barrios, 2021). Orthography then appears to play a major role in shaping L2 phonology. Furthermore, orthographic effects on pronunciation and phonological awareness—explicit knowledge about the phonology of the second language—persist after years of immersion in a naturalistic environment, in spite of much exposure to native input (Bassetti *et al.*, 2020). In the current study, we investigated whether a teaching intervention that specifies the relationship between L2 graphemes (sound spellings) and corresponding sounds can reduce the effects of orthography on L2 speakers' production and awareness of L2 sounds. A group randomized trial design was used to investigate the effects of making L2 learners of English cognizant of the relationship between number of letters in the spelling of an English sound (consonants and vowels) and the length of the corresponding sound, using student reflection, explicit teaching, and practice with spoken and written input.

### Effects of orthographic forms on second language phonology

The way second language words and sounds are spelled—their orthographic forms—can lead to non-nativelike pronunciations, as L2 speakers pronounce words “the way they are written,” often recoding L2 orthographic forms using the orthography-phonology conversion rules of their first language (for a review, see Hayes-Harb & Barrios, 2021). For instance, native Italian (Italian<sub>L1</sub>) learners of English as a Second Language (English<sub>L2</sub>) can add a sound corresponding to a so-called “silent letter,” for example, adding [l] in *walk* (Bassetti & Atkinson, 2015). Italians may be particularly prone to orthographic effects because there is evidence that such effects are stronger among native users of transparent—that is to say regular—writing systems (Escudero, 2015), of which Italian is an example. In particular, when an English consonant is spelled with double letters, Italian<sub>L1</sub> learners of English<sub>L2</sub> often produce it as longer than the same consonant spelled with a single letter. This is because their L1 Italian contrasts short and long consonants (called *singletons* and *geminate*s, respectively), and Italian geminate consonants are spelled with double letters, as in /'fato/-/'fato: (‘‘fate’’ vs. ‘‘fact.’’ the /:/ symbol represents a longer sound), spelled <fato>-<fatto> (Bertinetto & Loporcaro, 2005). Although English does not have contrastive length (Davis, 2011), Italians produce long consonants in English<sub>L2</sub> words with double consonant letters, for instance, producing a longer [t] in *kitty* than in *city* (Bassetti, 2017), and producing homophones with longer or shorter consonants depending on their spelling, for instance producing a longer [n] in *Finnish* than in *finish* (Bassetti *et al.*, 2018).

The effects of number of letters on sound duration are established early in the word learning process, as Italians produce long and short consonants in newly learned spoken words after as little as eight exposures, if the spoken word was presented together with its orthographic form (Cerni *et al.*, 2019). In metalinguistic awareness tasks, Italian<sub>L1</sub> speakers judge the same consonant as two different consonants when spelled with double letters or a single letter, and generally believe that English has long and short consonants (Bassetti *et al.*, 2020). Orthographic forms even appear to affect speech perception, as Italians perceive different consonants in English homophones such as *finish* and *Finnish*, but only when the two target words

are preactivated, for instance by images of a finish-line flag and a Finnish flag (Bassetti et al., 2021). Overall, it appears that Italian<sub>L1</sub> speakers of English<sub>L2</sub> make a contrast between English long and short consonants—a contrast that is not attested in the phonological system of the English language and that affects their L2 perception, production, and awareness.

The orthographic forms of English vowels also affect Italians' production, perception, and judgment of English<sub>L2</sub> vowels, but the effects are less straightforward than with consonants. Neither the Italian nor the English language has contrastive vowel length; however, English tense vowels are longer than lax vowels (as well as having qualitative differences, Roach, 2004), and vowel digraphs often represent long vowels in the English writing system, such as <oo> in *moon* (/mu:n/) and <ea> in *jeans* (/dʒi:nz/) (Carney, 1994). Single vowels as well as the <V\_e> digraph (known as “silent e”) can also represent tense (that is to say, longer) vowels, as in *ski* (/ski:/) and *June* (/dʒu:n/). Italian<sub>L1</sub> speakers of English<sub>L2</sub> often produce a vowel as longer in words spelled with a vowel digraph than in words spelled with a single vowel or a <V\_e> digraph, for instance producing the target vowel [u:] as longer in *moon* and shorter in *June* (Bassetti & Atkinson, 2015; Bassetti et al., 2020) and categorise and perceive different vowels in homophonous words spelled with a digraph, versus with a single letter or <V\_e> digraph (Bassetti et al., 2020; Bassetti et al., 2021). Italians are not aware that <V\_e> is a split digraph that indicates a long vowel, Bassetti & Atkinson, 2015).

Lengthy naturalistic exposure does not appear to reduce such orthographic effects on L2 phonology. A few studies compared the perception, production, and awareness of English<sub>L2</sub> consonant and vowel length in Italian instructed learners of English<sub>L2</sub>, with no or minimal exposure to a native-speaking environment, with that of Italian<sub>L1</sub>-English<sub>L2</sub> sequential bilinguals who had studied English at school and then had been living in an English-speaking environment for a long time (Bassetti et al., 2020; Bassetti et al., 2021; Bassetti et al., 2018; Mairano et al., 2018). Results consistently showed that lengthy naturalistic exposure does not reduce orthographic effects on L2 phonology. The sequential bilinguals produced, perceived, and judged English<sub>L2</sub> consonants and vowels as being short or long depending on their spelling.

Since orthographic forms affect L2 phonology, and lengthy naturalistic exposure does not seem to reduce such effects, the next question is whether a teaching intervention might reduce such orthographic effects, given that phonetic instruction can have positive effects on L2 perception and production.

### The effects of phonetic instruction on L2 phonology

L2 pronunciation instruction can positively affect L2 speech perception and production, as shown by narrative reviews (Thomson & Derwing, 2015) and meta-analyses (J. Lee et al., 2015). There are caveats, however. Some studies found no positive effects (Lord, 2005; Saalfeld, 2012), and we do not know whether perception training improves production and vice versa. Finally, positive effects of instruction may be dependent on variables that are not yet fully understood, such as intervention

duration, the sounds targeted, and the tasks used to test outcomes (Saito & Plonsky, 2019; Sakai & Moorman, 2018).

Looking specifically at duration, research indicates that with a novel language, even minimal instruction that involves briefly explaining that a language has a long-short contrast, or drawing attention to sound duration, can lead to naïve listeners being able to distinguish long and short sounds (Hisagi & Strange, 2011; Porretta & Tucker, 2015). Looking at learners, training helps them perceive geminates in Japanese<sub>L2</sub> (Sonu *et al.*, 2013). For instance, perception training with waveform displays improved the production (Motohashi Saigo & Hardison, 2009) and perception (Sadakata & McQueen, 2013) of singleton-geminate contrasts in beginner learners. Indeed, Americans with up to five years of Japanese<sub>L2</sub> learning experience almost doubled their geminate identification accuracy after just 10 sessions of 15 minutes each (Hirata, 2004). However, L2 learners generally do not achieve native-like categorical perception (Sonu *et al.*, 2013) and produce different durations compared with native speakers. Still, L2 learners can be trained to reduce reliance on duration in the L2: even a short period of training can help Chinese learners of English<sub>L2</sub> rely less on vowel duration and more on qualitative differences to distinguish /i/-/i/ (Wang & Munro, 1999).

Looking specifically at whether instruction might reduce orthographic effects on L2 phonology, research is limited and generally negative. Evidence mostly comes from studies of word learning in novel languages (L0) and indicates that brief grapheme-phoneme correspondence (GPC) instruction does not help. Showalter and Hayes-Harb (2015) reported that learners of Arabic<sub>L0</sub> who saw written input in the unfamiliar Arabic script and received short instructions about Arabic GPCs were not facilitated in discriminating a difficult Arabic phonological contrast. Similarly, Hayes-Harb *et al.* (2018) found no improvement when English speakers learning spoken pseudowords in German<sub>L0</sub> were taught that word-final stops are devoiced even if spelled with letters representing voiced consonants (e.g., word-final <g> is produced [k]), and Showalter (2020) reported the same when English speakers learning spoken pseudowords in Russian<sub>L0</sub> received brief written instruction in Russian GPCs. However, it is important to acknowledge that naïve learners of novel languages are faced with novel and/or difficult sounds and/or orthographic symbols, and evidence from these studies cannot be generalized to L2 learners or users.

Two issues were identified, on the basis of the above literature review that need to be addressed. First, while phonetic instruction can reduce the effects of L1 phonology on L2 phonology, it is unclear whether instruction about phonology-orthography correspondences can reduce orthographic effects on L2 phonology. Second, while instruction can help L2 speakers perceive and produce an L2 contrast, it is unclear whether it can stop them perceiving and producing a contrast that is not attested in the L2.

## The present study

The aims of the study were to investigate whether instruction could reduce the effects of orthographic forms on L2 speech production and L2 speech sound awareness. A group randomized trial (GRT) was conducted with eight classes of students.

Half the classes received explicit teaching of the GPCs under analysis. Students in the other four classes practiced written and spoken forms of the words employed in the intervention, but without explicit teaching of GPCs. Outcome was assessed by comparing the two groups' production and awareness of L2 speech sounds at pre- and posttest. Based on the literature review above, we hypothesized that the explicit teaching of orthography-phonology correspondences (number of letters in the spelling of an English language consonant or vowel and the duration of the related sound) would improve performance, as follows:

1. In a consonant and a vowel production task, reduce or eliminate the difference in duration between the same sound (consonants or vowels) when spelled with a digraph or with a single letter.
2. In a consonant awareness task, reduce or eliminate the incorrect judgment of consonants spelled with single or double letters as two different consonants (a singleton and a geminate).
3. In a vowel awareness task, by clarifying that the same long vowel can be spelled with a single letter or a digraph and a single-letter spelling does not necessarily represent a "short" vowel, reduce or eliminate the incorrect judgment of vowels spelled with single letter or digraphs as two different vowels.

The following paragraphs outline the rationale for the instruction used, the tasks employed, and the GRT approach.

We decided to implement an intervention addressing the GPCs of English consonant and vowel digraphs because exposure to English native speech by itself does not eliminate orthographic effects on L2 phonology and because explicit phonetic instruction can impact L2 speech production. Lengthy naturalistic exposure does not appear to reduce the effects of number of letters (single or digraphs) on second language phonology in Italian<sub>L1</sub> speakers of English<sub>L2</sub> (Bassetti et al., 2018; Bassetti et al., 2020; Bassetti et al., 2021). Researchers also found that sequential bilinguals with lengthy naturalistic exposure produced more native-like voice onset times than instructed learners, but did not produce fewer long consonants (Mairano et al., 2018), and argued that the effects of orthography on timing in L2 speech production may be more impervious to change than effects of L1 phonology that are not reinforced by orthographic representation. Although naturalistic exposure cannot eliminate orthographic effects on L2 phonology, such effects might be reduced or eliminated by explicit instruction and drawing learners' attention to GPCs. This is because explicit phonetic instruction can positively affect L2 phonology (Lee et al., 2020; DLee et al., 2015; Saito & Plonsky, 2019; Thomson & Derwing, 2015), including the perception of contrastive length, at least in a novel language (Hisagi & Strange, 2011; Porretta & Tucker, 2015). It is, however, unclear whether the reverse is also true, that is to say whether instruction can make L2 learners notice the absence of a length contrast.

The intervention was designed to include elements of student reflection, explicit teaching, and pair and group work involving both perception and production training, and both spoken and written practice. Perception and production were both included because evidence of cross-modality training effects is not consistent, that

is to say it is unclear whether perception training can result in improvements in production, and vice versa. Flege's speech learning model (SLM-r) predicts that improvements in L2 perception can lead to improvements in L2 production because speech development is perception-based (Flege & Bohn, 2021). There is evidence that perception-based instruction can improve L2 pronunciation more than production-based instruction (B. Lee *et al.*, 2020). Hence, both were included in the present study.

The intervention also involved oral as well as written production because evidence suggests that written tasks may be more effective than oral tasks at improving L2 learners' speech production (Solier *et al.*, 2019). Intervention tasks included reading aloud printed words, repeating native speakers' pronunciations of the same words, and spelling them. The intervention was administered to whole classes of students (not just the study participants) during a normal English language school teaching session. In order to ensure quality and fidelity, the intervention (explicit GPC instruction) and control group (passive exposure) teaching were delivered by an applied linguist with fifteen years' experience of teaching linguistics (Bassetti). The sessions for each of the eight classes lasted one hour, and the intervention focused on clarifying just one concept, that English digraphs do not represent longer sounds than single letters. We had originally planned two separate sessions for consonants and vowels, but decided on a single teaching session for the following reasons: (1) a series of unforeseen school events resulted in teachers lagging behind with the curriculum, which meant that intervention time needed to be minimised; (2) one hour seemed sufficient to understand that a single binary contrast is unattested in the L2 (most previous interventions targeted multiple L2 contrasts that are difficult for participants to perceive and/or produce); (3) just one hour of pronunciation instruction can be effective (e.g., B. Lee *et al.*, 2020), and a recent meta-analysis even showed that a shorter intervention can improve L2 speech production more than a longer one (Sakai & Moorman, 2018), although another meta-analysis (J. Lee, *et al.*, 2015) found longer instruction to be more successful than shorter instruction; (4) the intervention was meant to be suitable for time-pressed high-school teachers, and it was reasoned that if a classroom-based one-hour session can reduce or eliminate orthographic effects on L2 pronunciation, school teachers would be more likely to adopt the intervention. Consonant duration in English is unlikely to impact intelligibility, hence only justifying minimal investment of time and effort (see Plonsky & Oswald, 2014, for a discussion of potential benefits vis-à-vis implementation costs of interventions), still it affects accentedness, and it should be noted that the vast majority of Italian high-school learners consider a native-like pronunciation very important (Bassetti, 2017; Bassetti *et al.*, 2020).

Unlike most studies of phonetic instruction, we tested the effects not only on production but also on phonemic awareness—the ability to identify and manipulate consonants and vowels, which is a component of phonological awareness. First, it is generally agreed that explicit instruction can improve awareness of L2 phonology (Saito & Plonsky, 2019). Second, we assume, following Bassetti (2008, 2017) that L2 orthographic forms affect phonological representations, that are then reflected in both production and metalinguistic awareness, and indeed we had found that awareness of sound length predicts orthographic effects on duration in speech production (Bassetti *et al.*, 2020). Third, we thought that instruction could affect

performance in phonological awareness possibly more than in a production task, where multiple tasks compete for the L2 speakers' limited mental resources and attention to form is limited by the need to think about content. More in general, phonemic awareness is considered to develop as a consequence of literacy in an alphabetic writing system, hence likely to be affected by an intervention that addresses GPCs.

Finally, a group randomized trial design was selected as the most appropriate method for establishing the effectiveness of the intervention. It was not possible, within the constraints of the participants' school schedules, to carry out fully randomized allocation of participants to teaching conditions. While fully randomized controlled trials have been argued by some to be the ideal design for indicating the effectiveness of educational interventions (Connolly et al., 2017), group randomized trials, where classes or schools are randomly allocated to conditions, are widely employed in situations like the one in the present study, where randomization of individual participants to conditions was not logistically possible (Torgerson & Torgerson, 2008).

## Method

Using a group randomized trial design, we tested 100 learners of English as a Second Language studying in eight classes organized in four pairs (please see *Randomization* section below). One of the two classes in each pair was randomly selected to receive the explicit GPC instruction (EGPC condition), and the other the passive exposure intervention (PE condition). The participants were tested before (pretest,  $n = 100$ ) and after the intervention (posttest,  $n = 88$ ).

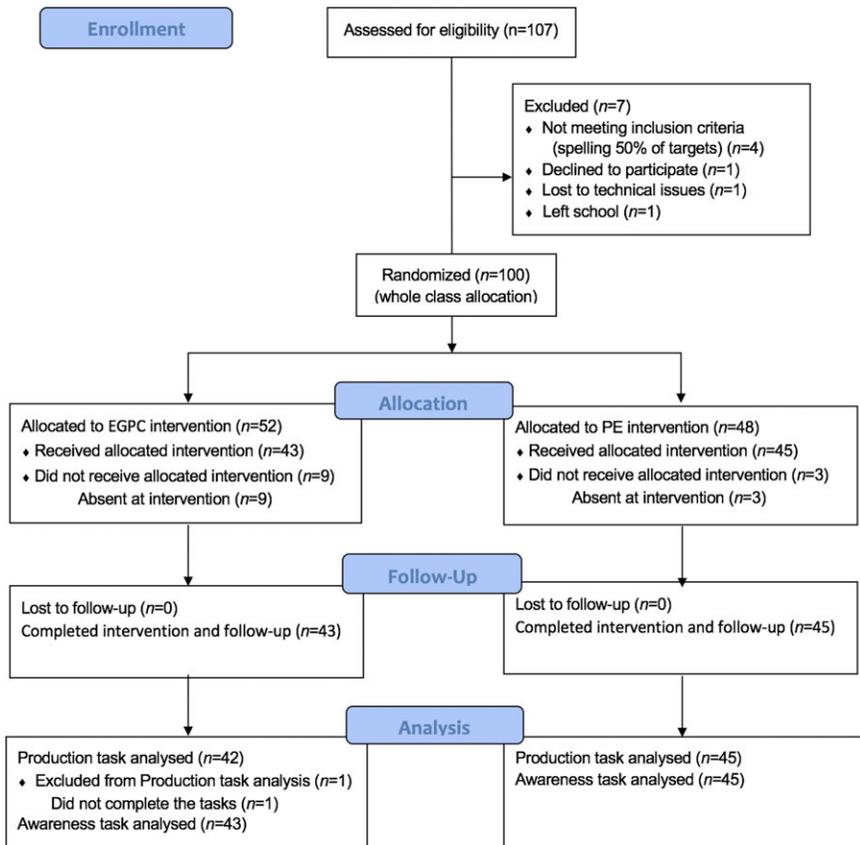
## Participants

Participants were Italian<sub>L1</sub> high-school learners of English<sub>L2</sub>. They were recruited from fourth-year classes of three high schools in Rome, Italy. Six English language teachers at the schools, teaching a total of eight classes, consented for their students to take part in the study. All students in these classes were invited to participate. Fourth-year classes were targeted to ensure a minimum age of 16 years, similar levels of English (as Italian schools follow the same national curriculum), and sufficient amount of time available to take part in the study (fifth-year students need to prepare for graduation exams). All participants could discriminate English tense and lax vowels (see Bassetti et al., 2020). Figure 1 gives an outline of the number of students involved in each stage of the study from initial recruitment to analysis of the results, following CONSORT recommendations.<sup>1</sup>

The final sample of students whose data were included in the analyses comprised 43 participants in the EGPC condition (one of them did not complete the posttest production task) and 45 in the PE condition. Table 1 gives a summary of demographic variables for the participants in the two conditions. The participants all took part in the study of Bassetti et al. (2020) and for that study assessments were administered of short-term memory, self-reported mimicry ability, pronunciation learning strategies, English language proficiency, English reading to listening ratio, attitudes to English language, and desire for native-like pronunciation (please see Bassetti

**Table 1.** Demographic variables for the participants in the explicit GPC (EGPC) and passive exposure (PE) conditions. Standard deviations are in parentheses

Measure	EGPC	PE
N	43	45
Age (years)	16.84 ( $\pm 0.37$ )	16.98 ( $\pm 0.50$ )
English language age of onset of acquisition (years)	5.74 ( $\pm 1.95$ )	5.78 ( $\pm 1.77$ )
English language length of study (years)	10.35 ( $\pm 1.89$ )	10.69 ( $\pm 2.32$ )



**Figure 1.** Flow-chart showing allocation of participants, as per CONSORT recommendations.

*et al.*, 2020 for details of the assessments used and summary scores). The students who received the EGPC and PE interventions were found to be well matched on the variables, and independent *t*-tests confirmed that there were no significant differences.

The sample size was deemed sufficient to reveal a medium effect size (Partial  $\eta^2 = 0.03$ ), with 0.90 power at an alpha-level of .05 (G\*Power 3.1 software, Faul et al., 2009). All participants reported normal or corrected-to-normal vision. None reported reading or listening difficulties, two participants were left-handed. Participation was voluntary and rewarded with book vouchers. Parents of all the participants provided written consent. The study was approved by the University of Warwick Research Ethics Committee (Number: 118/14-15, Project Title: Effects of Orthography on Phonology in Second Language Speakers of English: Pronunciation, Phonological Awareness, Speech Perception and Spelling).

### **Randomization**

In order to have a balance in terms of teacher and school characteristics across the EGPC and PE conditions, allocation of participants was implemented as indicated next. First, the eight classes of students were paired: there were two pairs where both classes in a pair had the same teacher (and same school) and two pairs where both classes in a pair were in the same school but had different teachers. Then, one class within each pair was randomly allocated to the EGPC condition and one to the PE condition. To ensure concealed allocation, randomization was carried out by an independent researcher with no knowledge of the project. As different classes included different numbers of students, 14 participants were absent for the intervention session, and the final sample included 43 participants who received the EGPC intervention and 45 who received the PE intervention.

Apart from the Bassetti, who delivered the intervention, all the researchers were blind to participants' group allocation: the Cerni who administered the pre- and postintervention assessments and coded responses in the awareness task, and the phoneticians who performed acoustic analyses. Students and their teachers were also blinded, as follows. Teachers were informed that all classes would receive a guest session, but were not cognizant of session content and class allocation. Students were not informed in advance about the intervention, which was presented on the day as a guest class and not as part of the project. They were also not told the purpose of the project, other than it being a study of Italian students' English language learning. Students were debriefed at the end of the study.

### **Procedure, materials, and tasks**

Participants were assessed twice, pre- and postintervention, in one-to-one sessions lasting approximately an hour that were delivered by the Cerni. On both occasions, each participant was assessed in a production task (delayed word repetition), then a phonological awareness task (rhyme judgment). The pretest assessment also included a memory task and a spelling-to-dictation task to assess knowledge of the target words in the production task; the posttest assessment also included open questions about metalinguistic awareness. All the tasks were developed by Bassetti et al. (2020) and are described in detail in that study. Task materials are available in two online repositories (<https://osf.io/p3q6d> and [www.iris-database.org](http://www.iris-database.org)).

The production task involved delayed word repetition and tested whether participants produced a sound as longer when spelled with a digraph than when spelled

with a single letter. The phonological awareness task tested whether participants judged the same sound as two different sounds (a short one or a long one) when spelled with a single letter or a digraph. In the production task, participants were asked to listen to a native speaker's production of a sentence while seeing a related picture (to facilitate comprehension), to count backwards (in order to remove traces of native production from memory), and then to listen to a cropped version of the same sentence and produce the missing word three times in a frame. There were 20 C-CC word pairs where the same target consonant was spelled with single or double consonant letters, such as *city-kitty*, and 20 V-VV word pairs where the same vowel was spelled with a single vowel letter or a digraph (double letters or two-letter grapheme), such as *skis-sees*. There were four counterbalanced lists, each one containing 20 pairs (10 consonant pairs and 10 vowel pairs) out of the total 40. Each participant saw the same list at both the pre- and posttest assessments.

The rhyme judgment task involved asking participants to decide whether two words contained the same sounds. The targets were 12 C-CC word pairs, such as *very-cherry*, and 12 V-VV word pairs, such as *rule-cool*. There were also 24 rhyming controls—rhyming words containing one sound that is spelled differently in the two words, such as *tale-mail*—and 24 non-rhyming controls—word pairs whose rimes differed in one consonant or vowel, such as *toy-day*.

The spelling-to-dictation task assessed participants' knowledge of the spelling of target words presented in the production task, so that a spoken word would be entered in the analysis only if the participant knew its spelling.

At the beginning of the school year, all participating classes read and listened to specially prepared texts containing one occurrence of each of the target words necessary for the production task, during a normal teaching session delivered by their English language teacher. Preintervention assessments were conducted between November and March. Then, after being allocated to the EPGC or PE groups (see *Randomization*, above), the eight entire classes received a one-hour teaching session during a normal English language session in their classroom in mid-April. The posttest assessment took place between May and June. The study ended with a debrief session held by the Bassetti for all teachers and students at each school outside of teaching hours, where the project and its findings were presented and discussed. Since the posttest revealed no effects of the intervention, and because teachers were unwilling to give up one hour of teaching at the end of the year, we did not administer a planned delayed posttest and did not deliver the EPGC intervention to the PE group, which we had planned for ethical reasons (see Connolly *et al.*, 2017).

### **Intervention**

Both EPGC and PE groups received a one-hour session, delivered by the same teacher, with similar structure and tasks, and crucially practiced the same English words, with the same amount of spoken and written input and practice. The difference was that the experimental EPGC group learnt about the GPCs of English consonant and vowel digraphs; the PE control group learnt about English word formation. The control condition was included for three reasons: (1) to prevent participants from guessing the purposes of the intervention, (2) to

avoid placebo or demand effects, and (3) to provide all participants with the same amount of exposure to the target English words.

Both EGPC and PE conditions included personal reflection, explicit teaching, and practice, and addressed both perception and production training, using oral and written input, as detailed below. Each learning set started with an activity aimed at reflecting on a phenomenon, followed by an explicit instruction mini-lecture about the phenomenon, followed by a teacher-led group activity, worksheet-based pair work, and a final group production task. The aim was to make students reflect, provide explicit teaching, and then practice.

For each group, a set of 22 PowerPoint slides and a handout were prepared. Preintervention lesson materials and intervention materials are available at <https://osf.io/eubpz> and [www.iris-database.org](http://www.iris-database.org).

#### *Instruction in the two conditions*

*EGPC condition.* Students worked on the relationship between digraphs and sound length, looking first at consonants then at vowels. In both parts, there were three components: reflection (individual reflection and/or group discussion aimed at awareness training), explicit teaching, and practice (worksheet-based pair work and/or teacher-led group production task including speech imitation).

In the first part, titled “Double letters in English orthography and pronunciation,” the first task allowed students to reason on the discrepancy between the homophony of the words *finish* and *Finnish* and their different spellings: students first listened to a recording of [fɪnɪʃ] and spelled it, then listened to two sentences containing the word *finish* and *Finnish*, respectively, and spelled the two words. In the explicit teaching part, the Bassetti explained that double consonant letters do not represent long consonants in English. In the practice part, there was a teacher-led group production task of 40 CC-words of increasing length (3 to 20 letters, e.g., *ill*, *immunohistochemistry*) that used a produce-listen-repeat format whereby students read the word aloud, listened to a native speaker while noticing the difference with their own pronunciation, then repeated the word. Finally, there was group discussion of the girls band name *Girls aloud*, a pun on *girls allowed*.

The second part followed a similar structure to direct participants’ attention to the relationship between vowel spelling and vowel duration: reflection (listening to two sentences containing the words *ship* and *sheep*, respectively, and spelling the two words; then listening to homophones such as *cede-seed* and *rose-rows* while seeing their spellings); explicit teaching of the relation between “long” vowels and vowel digraphs (including <e\_e> or “silent e”); practice (group production of 20 words containing vowel digraphs with produce-listen-repeat format).

Finally, students practiced consonants and vowels with a homophone-matching task: working in pairs, they were asked to strike through homophones from a word cloud containing 22 C-CC and V-VV orthographic minimal word pairs (e.g., *analyst-annalist*, *lode-load*; some student pairs did not complete the task due to time constraints). Then, the students produced the 44 words in a pronounce-listen-repeat group activity.

*PE condition.* Students practiced the same words using the same produce-listen-repeat tasks as the EPGC group, but with no mention of spelling, as students worked

on English derivation and word-formation (in order to create meaningful activities, a small number of words were replaced with others containing the same GPCs, e.g., *annual* instead of *annalist*). Students reflected on the structure of derived words, then were taught about morpheme and suffixes, then practiced the same words as the intervention group in a word-decomposition and a word-building task (e.g., decomposing *rosy* into *rose-y*, matching *ill* with *-ness* to make *illness*), and finally they produced these words in a pronounce-listen-repeat group activity.

## Data analysis

In the acoustic analysis, in order to extract the duration of each target sound produced in the delayed word repetition task, we followed the procedure outlined in Bassetti *et al.* (2020). Praat software was used (Boersma & Weenink, 2016), following standard criteria (Turk *et al.*, 2006). Three expert phoneticians performed the analysis (interclass correlation in 5% of the data from .97 to .98, all  $ps < .001$ )

To summarize, consonant duration was measured as the duration of closure (Esposito & Di Benedetto, 1999), while vowel duration was considered from the onset to the cessation of a clear formant pattern, especially relying on F2. For each target, the average duration was calculated as the mean of three repetitions. The measure used in statistical analysis was a long-sound-to-short-sound ratio calculated from each word pair (CC:C ratio for consonants, VV:V ratio for vowels).

Statistical analyses were run in R version 3.5.0 (R Core Team, 2018) with RStudio version 1.1.453 (RStudio Team, 2018) and the following packages: *doBy* (Højsgaard & Halekoh, 2016), *ggplot2* (Wickham, 2009), *lme4* (Bates, Mächler, *et al.*, 2015), *lmerTest* (Kuznetsova *et al.*, 2017), *Publish* (Gerds & Ozenne, 2018), *RePsychLing* (Bates, Kliegl, *et al.*, 2015), *Rmisc* (Rmisc, 2013), and *scales* (Wickham, 2016). Data preparation and model construction followed similar procedures as in Bassetti *et al.* (2020).

## Delayed word repetition

Out of the total 3,480 word tokens presented in the pretest to the 87 participants included in the analysis, 308 were not produced or not analyzable acoustically, and 376 were eliminated because the participants spelled them incorrectly in the spelling-to-dictation task. Considering the posttest data set, 180 word tokens were not produced or were not analyzable acoustically, while 409 were eliminated due to being spelled incorrectly.

After calculating long-sound-to-short-sound ratios from each word pair, outliers (top and bottom 1% of ratios from each condition and time of testing) were removed (2.13% of the data). In order to compare the same quantity and type of word pairs produced before and after the intervention, we selected from the clean data only C-CC and V-VV pairs that were produced by the same participant at pretest and posttest, discarding those pairs that were produced only on one occasion (111 C-CC trials, 101 V-VV trials). The final data set consisted of 1202 C-CC and 886 V-VV pairs. Ratios were log-transformed due to positive skew.

Two linear mixed models were used to analyze the effects of the teaching intervention on consonant duration ratios and vowel duration ratios using *lmer* function

**Table 2.** Geometric means and 95% CIs for CC:C ratios and consonant awareness accuracy (percent correct) by condition (Explicit GPC, Passive Exposure) and time of testing (pretest, posttest)

		EGPC		PE		
		Mean	95% CIs	Mean	95% CIs	
Delayed word repetition CC:C ratio	Pretest	1.53	[1.48; 1.58]	1.56	[1.51; 1.61]	
	Posttest	1.45	[1.41; 1.50]	1.53	[1.49; 1.58]	
Rhyme judgment Accuracy (%)	Pretest	C-CC pair	64	[60; 68]	54	[49; 58]
		Control pair	74	[71; 77]	71	[68; 73]
	Posttest	C-CC pair	64	[60; 68]	53	[49; 58]
		Control pair	78	[75; 80]	74	[71; 76]

in the *lmerTest* package. The initial model included fixed effects for condition (EGPC versus PE), time of testing (pretest versus posttest) and their interaction. Maximal random structure was initially tested for the two models by including by-participant and by-item random intercepts and slopes. Due to failure of convergence or overfitting, random components were simplified examining the summary of the *rePCA* function in the *RePsychLing* package (Bates, Kliegl, et al., 2015).

Once the random structure was established, significant predictors were selected using the backward elimination procedure, excluding those fixed effects that failed to reach significance in model comparisons through likelihood-ratio tests. In the *Results* section, we report significant findings of the final models.

### Rhyme judgment

Responses to consonant and vowel judgments (correct or incorrect) were analyzed with two logit mixed-effects models with binomial distribution using the function *glmer* in the *lmerTest* package. The initial maximal models included fixed effects for condition (EGPC versus PE), time of testing (pretest versus posttest), type of rhyme (control or C-CC/V-VV rhyme), and their interactions. Rhyming and nonrhyming control items were grouped into a unique control category. As in the models for production duration ratios, maximal random effect structure was initially tested. Fixed and random effect reduction followed the same procedure described above for the delayed word repetition task.

## Results

### Consonants: Duration and awareness

Table 2 presents descriptive results by condition and time of testing in the delayed word repetition task and by condition, time of testing, and type of rhyme in the rhyme judgment task (a ratio of 1 shows that the consonant has the same duration in both words in a pair; a higher ratio shows that the consonant has longer duration in the word spelled with double letters than in the corresponding word spelled with a single letter).

**Table 3.** Results of final models for the CC:C ratio in the delayed word repetition task (A) and for accuracy in the consonant rhyme judgment task (B)

Consonant analyses					
A. Delayed word repetition – CC:C ratio					
Random effects	Variance	St. dev.			
Participants	0.016	0.127			
Word pairs	0.008	0.082			
Residuals	0.057	0.238			
Fixed effects	Estimate	SE	95%CI	<i>t</i>	<i>p</i>
Intercept	0.427	0.025	[0.378; 0.476]	17.08	< .001
Time (Posttest – Pretest)	-0.033	-0.137	[-0.060; -0.006]	-2.38	.018
B. Rhyme judgment – Accuracy					
Random effects	Variance	St. dev.	Corr		
Participants	0.593	0.770			
Type of rhyme Participants	1.297	1.129	-0.60		
Word pairs	0.953	0.976			
Fixed effects	Estimate	SE	95% CI	<i>z</i>	<i>p</i>
Intercept	1.013	0.229	[0.565; 1.462]	4.43	< .001
Type of rhyme (C-CC – control)	-0.724	0.366	[-1.441; -0.007]	-1.98	.048
Condition (EGPC – PE)	0.412	0.150	[0.119; 0.706]	2.75	.006
Time (Posttest – Pretest)	0.255	0.086	[0.086; 0.423]	2.96	.003
Type of rhyme (C-CC – control)*Time (Posttest – Pretest)	0.264	0.138	[-0.534; -0.007]	-1.91	.057

Table 3A presents the parameters of the final model on CC:C duration ratio. The model retained as random effects by-participant and by-word pair intercepts. Even though the CC:C ratio descriptively decreased on the posttest for the EPGC condition, time of testing was the unique fixed effect that improved the model fit. This result suggested that the CC:C ratio decreased at posttest but independently from the intervention. Crucially, the condition  $\times$  time interaction was not significant, showing that the type of intervention had no effect.<sup>2</sup>

Looking at the consonant awareness results, the final model for accuracy (Table 3B) included as random effects the intercepts for participants and word pairs and a by-participant random slope for type of rhyme. Fixed effects that added significant information to the model were the main effect of condition, type of rhyme and time of testing. Overall, accuracy was higher for the EPGC participants than the PE participants, and C-CC rhymes elicited more errors than control rhymes. Regarding the difference between time of testing, the accuracy rate was higher at posttest than at pretest, but independently from the intervention. Interestingly,

**Table 4.** Geometric means and 95% CIs for VV:V ratios and vowel awareness accuracy (percent correct) by condition (explicit GPC, passive exposure) and time of testing (pretest, posttest)

		EGPC		PE		
		Mean	95% CIs	Mean	95% CIs	
Delayed word repetition VV:V ratio	Pretest	1.07	[1.04;1.10]	1.05	[1.02;1.09]	
	Posttest	1.09	[1.06;1.13]	1.03	[1.00;1.06]	
Rhyme judgment Accuracy (%)	Pretest	V-VV rhyme	63	[59; 67]	56	[52; 60]
		Control rhyme	82	[79; 84]	80	[77; 82]
	Posttest	V-VV rhyme	66	[62; 70]	59	[55; 63]
		Control rhyme	85	[82; 87]	84	[82; 86]

the interaction between type of rhyme and time of testing added information to the model, approaching significance. This interaction indicated that at posttest EPGC and PE participants improved in accuracy with control pairs (pairwise comparisons on estimated marginal means:  $z_{pre-post} = -3.27$ ,  $p = .001$ ), but did not improve with target C-CC pairs ( $z_{pre-post} = 0.09$ ,  $p = .932$ ). Both EPGC and PE participants were more accurate with control pairs than C-CC pairs at both pretest ( $z_{control-target} = 1.98$ ,  $p = .048$ ) and posttest ( $z_{control-target} = 2.50$ ,  $p = .012$ )<sup>3</sup>.

### Vowels: Duration and awareness

Table 4 presents descriptive results by condition and time of testing in the delayed word repetition task and by condition, time of testing and type of rhyme in the rhyme judgment task.

As reported in Table 5A, the final model for VV:V ratio retained as random effect the by-word pair random intercept. As fixed effect, only condition proved significant, showing that the VV:V ratio was higher in the EPGC participants than the PE participants. No condition x time interaction was retained in the model, showing that the type of intervention had no effect.

Looking at the vowel awareness results, the final model (Table 5B) that best explained the likelihood of a correct answer retained as random effects by-participant and by-word pair intercepts and the by-participant random slope for Type of rhyme. The fixed effects that added information to the model were Type of rhyme and Time of testing. Again, responses to control pairs were more likely to be correct than responses to V-VV rhymes, independently from the time of testing. Furthermore, posttest responses were more accurate than pretest responses, but independently from the intervention.

## Discussion

We aimed to see whether an explicit intervention that targeted the pronunciation of English single and double consonant and vowel graphemes would reduce

**Table 5.** Results of final models for the W:V ratio in the delayed word repetition task (A), and for accuracy in the vowel rhyme judgment task (B)

Vowel analyses					
A. Delayed word repetition – W:V ratio					
Random effects	Variance	St. dev.			
Word pairs	0.012	0.108			
Residuals	0.037	0.123			
Fixed effects	Estimate	SE	95%CI	t	p
Intercept	0.038	0.026	[-0.013; 0.089]	1.45	ns
Condition (EGPC – PE)	0.045	0.013	[0.019; 0.070]	3.41	<.001
B. Rhyme judgment – Accuracy					
Random effects	Variance	St. dev.	Corr		
Participants	0.545	0.738			
Type of rhyme Participants	0.176	0.420	0.34		
Word pairs	1.198	1.094			
Fixed effects	Estimate	SE	95%CI	z	p
Intercept	1.985	0.246	[1.501; 2.470]	8.03	<.001
Type of rhyme (C-CC – control)	-1.551	0.398	[-2.331; -0.770]	-3.89	<.001
Time (Posttest – Pretest)	0.259	0.067	[0.128; 0.394]	3.39	<.001

orthographic effects in Italian learners of L2 English. Previous literature has demonstrated that Italian learners of English produce and perceive consonants and vowels as longer in duration when they are spelled with double letters than single letters, in line with Italian GPCs. Results of the current study revealed that the intervention did not eliminate or even reduce the orthographic effects on either sound duration in speech production or awareness of sound length. Indeed, on the postintervention test, students in the explicit GPC (EGPC) intervention condition did not differ from participants in the passive exposure (PE) condition who had practiced the same words for the same amount of time with no GPC instruction.

Looking at consonants first, on the posttest, the EPGC and PE participants did not differ in a measure of gemination in speech production—geminate:singleton ratio—or in a measure of awareness that English does not make consonant length contrasts—accuracy in a rhyme judgement task. Regardless of time of testing, the geminate:singleton ratio was around 1.5 in both EGPC and PE groups, which is in line with previous studies of gemination in Italian high school learners of English (Bassetti, 2017; Bassetti *et al.*, 2018). Also, on the posttest, the percentage of correct responses to CC-C rhymes was around 60%, compared with around 75% for control rhymes, in line with pretest results previously reported (Bassetti *et al.*, 2020).

The speech production task results confirm that even after a targeted GPC training intervention, Italians produce geminates in English words in correspondence

with double consonant letters. Such results confirm and extend previous findings that orthography-induced consonant gemination in Italian<sub>L1</sub> speakers of English<sub>L2</sub> is persistent. Previous studies found that it is not eliminated by extensive exposure in a target-language environment (Bassetti et al., 2018; Bassetti et al., 2020; Bassetti et al., 2021), although the same exposure can reduce other non-native like timing aspects of L2 speech production that are not reinforced by orthography (Mairano et al., 2018).

Furthermore, the metalinguistic awareness task results extend such findings—that orthographic effects on L2 phonology are persistent—from speech production to metalinguistic awareness. The absence of effects of training on metalinguistic awareness is perhaps more surprising than the absence of effects on speech production. Since a short GPC training intervention can positively impact speech perception and production in a novel language, it should have impacted at least participants' awareness that English has no consonantal length. If there is no improvement in a metalinguistic awareness task, it is unlikely that there could be effects on production tasks, where a number of variables compete for limited cognitive resources, and indeed research shows that phonetic training interventions impact controlled production more than spontaneous production because learners can focus on pronunciation more (Liu, 2011; Saito, 2012).

The reasons for the persistence of orthography-induced L2 gemination are currently unknown, but Bassetti (2022) proposed a number of nonmutually-exclusive explanations, arguing that gemination, established in the mental representation of an English<sub>L2</sub> word during word learning (Cerni et al., 2019), could be reinforced by a series of factors:

- (1) Further encounters with the word's orthographic form, which Italians recode as containing a geminate because their set of English<sub>L2</sub> GPC rules features a <CC>-/C:/ rule;
- (2) Orthography-induced gemination in the spoken word produced by Italian and other English<sub>L2</sub> speakers whose L1 has gemination; and in the L2 speaker's own output, including overt speech, inner speech, and silent reading;
- (3) Gemination in the L2 listener's intake of English<sub>L1</sub> production—Italians illusorily perceive a geminate in a spoken word if their own mental representation of the word contains a geminate (Bassetti et al., 2021);
- (4) Lack of negative evidence—nontarget-like consonant duration does not result in communication breakdown, hence not resulting in interlocutor feedback or noticing the gap between one's own and others' production;
- (5) Incorrectly applying the generally successful strategy of ignoring one's own perception of two L2 sounds as similar when the L2 spelling represents the sounds as different ones (for instance, when perceiving two English<sub>L2</sub> vowels as the same vowel, see Escudero et al., 2008);
- (6) Other corroborating evidence: nonlexical gemination in English native speech (as in *top pick* vs *topic*); gemination in English loanwords to Italian (although the relationship between loanwords and L2 phonology is unclear).

These explanations suggest that gemination production is due to gemination perception, following the assumption that perception and production coevolve, as proposed by the SLM-r model (Flege & Bohn, 2021). Future research could test their validity.

Looking at the findings for vowels, on the posttest, the EPGC and PE participants did not differ in either VV:V ratio or accuracy in the rhyme judgment task. Regardless of time of testing, the VV:V ratio was higher than one in both groups, although it was below the 1.10 ratio for vowels previously reported in sequential bilinguals (Bassetti *et al.*, 2018) and the 1.14 ratio previously reported in high-school learners (Bassetti & Atkinson, 2015). This could be because some learners do not distinguish English tense and lax vowels, and therefore do not use vowel duration to distinguish them. The percentage of correct responses to V-VV rhymes was lower than for control rhymes at both pretest and posttest, but unlike the case with consonants, EPGC and PE participants displayed very similar levels of improvement in V-VV rhymes and control rhymes. This shows that the type of instruction had no effect, but all participants had increased their knowledge of the pronunciation of the specific words being tested, including both targets and controls.

While participants, regardless of type of intervention, showed no change in the percentage of correct responses to rhymes containing a C-word and a CC-word or in the production of CC consonants, there was an overall effect of time of testing on the other types of words tested in the metalinguistic awareness task, as both EPGC and PE participants showed a small but significant increase in the percentage of correct responses to control pairs and to V-VV pairs. This means that participants overall improved their pronunciation of the words they produced and reflected about in these tasks, which can be explained partly because participants improved their English vocabulary and pronunciation over the few months' duration of this project, and probably more because they performed the same task with the same words twice.

The findings indicate that a targeted intervention cannot eliminate or even reduce orthographic effects on the production of geminates in English, and on the awareness that English does not have consonant length contrasts. The findings resonate with other studies that demonstrate pervasive effects of early learned GPCs in monolingual English-speaking children (Henbest & Apel, 2017). In a demonstration of the long-lasting influence of explicit GPC (phonics) instruction, Thompson *et al.* (2009) carried out a study with English-speaking adults who in the first year of school had received either phonics instruction or whole-language (no or minimal GPC) instruction. The two groups of adults had comparable levels of sight word reading ability and word-naming times, but in reading nonwords and low-frequency words, the phonics group produced much higher levels of regular decoding (and fewer rime-consistent) responses. Adults in this group were also much better at providing letter sounds and in phonemic awareness tasks (phoneme counting) than the whole-language group. In our study, we found that we were not able to eliminate orthography-induced effects on second language phonology. We should perhaps not be surprised, in light of the previous evidence showing pervasive and long-lasting effects of early learned orthography-phonology connections.

It appears that, although even minimal training can facilitate the perception of consonant length in a novel language (Hisagi & Strange, 2011; Porretta & Tucker, 2015), a one-hour teaching intervention cannot counteract orthographic effects and

might not help unlearn the singleton-geminate distinction (even though it can counteract L1 phonological effects and help learn the singleton-geminate distinction).

It is, however, too early to rule out the possibility that interventions impact orthographic effects on L2 phonology, due to a number of limitations of this study. First of all, participants were learners with 10 years' experience of studying English in an instructed context. Though this was programmatic, as a reaction toward the excessive amount of research on orthographic effects in novel language learners, training may facilitate beginning learners more than advanced learners, whose non-target-like phonological representations of English words and knowledge of English GPCs have been established over a decade. Indeed there is evidence that phonetic training interventions can be more successful at the beginner level than at later stages (B. Lee et al., 2020; Sakai & Moorman, 2018). Future research could then try a GPC training intervention in beginner L2 learners (as opposed to the present experienced L2 learners and users).

A second limitation of the study involves the fact that a single researcher delivered the instruction to all the groups who took part. To be completely confident that the intervention does not work, it would be necessary for the intervention to be delivered by multiple instructors to avoid the results being influenced by the competence of a single instructor.

Arguably the main limitation of the study was the limited amount of time dedicated to the intervention. Originally, two one-hour sessions had been planned, one for consonants and one for vowels. However, this became impossible due to schools' opposition when a series of strikes and holidays meant that no more time could be devoted to the experiment without damaging the students' ability to complete the program of the year. Yet, evidence of the effects of intervention duration is inconsistent, as two meta-analyses reported opposite findings, as outlined above. Future research could then investigate effects of a longer intervention, possibly using the same instructional materials prepared in this study.

This study is the first, to the best of our knowledge, to have included an intervention aimed at reducing orthographic effects on L2 phonology. The sample size was larger than in most typical studies of phonetic interventions; it was strengthened by having a group that was exposed to the same words for the same amount of time, but without the explicit GPC teaching, and it involved randomized allocation of matched classrooms that ensured control for the effects of language teacher and teaching. Finally, the blinding of the researchers who delivered postintervention assessments, and who conducted acoustic and statistical analyses, contributed to the soundness of the analyses.

The methodological contributions of the study include the fact that the null results reported address the concern, expressed by Torgerson and Torgerson (2013) and others, regarding the bias toward publication of intervention studies reporting significant effects. The study also showed that delayed word repetition and rhyme judgment tasks, developed by Bassetti (2017) and Bassetti et al. (2020), can reliably measure effects of number of letters on both L2 production and phonological awareness. All the tasks used in the study, together with the intervention materials, are available to researchers and teachers (see *Procedure, materials, and tasks* section).

## Conclusions and pedagogical implications

The aim of this study was to test the effects of explicit teaching of orthography-phonology correspondences on production and awareness of consonants and vowels in second language learners. The results showed that the intervention did not reduce orthographic effects on L2 speech production or awareness. The nonsignificant results cannot be attributed to sample size, which was larger than is typical of L2 pronunciation teaching intervention studies, or to the duration of the intervention, which was similar to previous successful pronunciation teaching interventions such as those of B. Lee and colleagues (2020). It appears that explicit instruction in orthography-phonology correspondences does not reduce the impact of orthography on second language pronunciation and phonological awareness. These findings confirm that orthographic effects on L2 phonology, once established, are resistant to change. However, more research is needed, and future studies could increase the duration of the intervention, or try other pedagogical techniques, and target beginner students.

**Acknowledgments.** The authors are grateful to Paolo Mairano for managing and performing spoken materials recording and acoustic analyses; David Torgerson for advice on intervention set-up and data analysis; June McCready for advice on pronunciation teaching; Emanuela Buizza and Valentina De Iacovo for contributing to the acoustic analysis; Sheila Verrier for proofreading teaching materials and recording spoken materials. This work was supported by a Leverhulme Trust Research Grant [grant number: RPG 2013 180] awarded to Bassetti and Masterson. The sponsor had no involvement in the conduct of the research or the preparation of the article.

**Competing Interests Declaration.** The authors declare that they have no competing interests.

## Notes

1 The CONSORT (Consolidated Standards of Reporting Trials) Statement is a set of recommendations for reporting randomized trials, including a flow diagram template ([www.consort-statement.org/consort-statement/flow-diagram](http://www.consort-statement.org/consort-statement/flow-diagram)).

2 Since randomization to the EGPC and PE intervention conditions was carried out on the basis of whole classes, we also analyzed the delayed repetition data using two separate *t*-tests, one for the consonant results and one for the vowel results. The data comprised the mean score across participants in each class for the ratio change from pretest to posttest. The results revealed that the difference between the EPGC and PE conditions was not significant for either consonants or vowels ( $t(3) = 1.41$ ,  $p = .254$ , and  $t(3) = 0.68$ ,  $p = .545$ , respectively).

3 As in the case of the delayed repetition data, we analyzed the awareness data in terms of whole classes using two separate *t*-tests, one for consonants and one for vowels. Data comprised the mean score across participants in each class for accuracy rate on target stimuli change from pretest to posttest. The results revealed that the difference between the EPGC and PE conditions was not significant for either consonants or vowels ( $t(3) = -0.24$ ,  $p = .828$ , and  $t(3) = -0.36$ ,  $p = .740$ , respectively).

## References

- Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191–206). Multilingual Matters. <https://doi.org/10.21832/9781847691118-013>
- Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1835–1842. <https://doi.org/10.1037/xlm0000417>

- Bassetti, B. (2022). *Effects of orthography on second language phonology*. [manuscript in preparation]
- Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics*, *36*(2), 67–91. <https://doi.org/10.1017/S0142716414000435>
- Bassetti, B., Mairano, P., Masterson, J., & Cerni, T. (2020). Effects of orthographic forms on L2 speech production and phonological awareness, with consideration of speaker-level predictors. *Language Learning*, *70*(4), 1218–1256. <https://doi.org/10.1111/lang.12423>
- Bassetti, B., Masterson, J., Cerni, T., & Mairano, P. (2021). Orthographic forms affect speech perception in a second language: Consonant and vowel length in L2 English. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(12), 1583–1603. <https://doi.org/10.1037/xhp0000949>
- Bassetti, B., Sokolović-Perović, M., Mairano, P., & Cerni, T. (2018). Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs. *Language and Speech*, *61*(4), 577–597. <https://doi.org/10.1177/0023830918780141>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv preprint arXiv:1506.04967.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(2), 131–151. <https://doi.org/10.1017/S0025100305002148>
- Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer*. (Version 6.0.19) <http://www.praat.org/>
- Carney, E. (1994). *A survey of English spelling*. Routledge.
- Cerni, T., Bassetti, B., & Masterson, J. (2019). Effects of orthographic forms on the acquisition of novel spoken words in a second language. *Frontiers in Communication*, *4*(31). <https://doi.org/10.3389/fcomm.2019.00031>
- Connolly, P., Biggart, A., Miller, S., O'Hare, L., & Thurston, A. (2017). *Using randomised controlled trials in education*. Sage.
- Davis, S. (2011). Geminate. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 2. Suprasegmental and prosodic phonology, pp. 873–897). Wiley-Blackwell. <https://doi.org/10.1002/9781444335262.wbctp0037>
- Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, *36*(01), 7–22. <https://doi.org/10.1017/S014271641400040X>
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360. <https://doi.org/10.1016/j.wocn.2007.11.002>
- Espósito, A., & Di Benedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *Journal of the Acoustical Society of America*, *106*(4), 2051–2062. <https://doi.org/10.1121/1.428056>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.
- Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In Wayland, R. (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge: CUP.
- Gerds, T. A., & Ozanne, B. (2018). *Publish: Format output of various routines in a suitable way for reports and publication*. R package version 2018.04.17. <https://CRAN.R-project.org/package=Publish>
- Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*. Advance online publication. <https://doi.org/10.1017/S0261444820000658>
- Hayes-Harb, R., Brown, K., & Smith, B. L. (2018). Orthographic input and the acquisition of German final devoicing by native speakers of English. *Language and Speech*, *61*(4), 547–564. <https://doi.org/https://doi.org/10.1177/0023830917710048>
- Henbest, V.S. & Apel, K. (2017). Effective word reading instruction: What does the evidence tell us? *Communication Disorders Quarterly*, *39*(1), 303–311. <https://doi.org/10.1177%2F1525740116685183>
- Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, *116*(4), 2384–2394.
- Hisagi, M., & Strange, W. (2011). Perception of Japanese temporally-cued contrasts by American English listeners. *Language and Speech*, *54*(2), 241–264.

- Højsgaard, S., & Halekoh, U. (2016). *doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities*. R package version 4.5-15. <https://CRAN.R-project.org/package=doBy>
- Kuznetsova, A., Brockhoff, P. B., Christensen, R., & Haubo B. (2017). *lmerTest: Tests in linear mixed effects models*. <https://CRAN.R-project.org/package=lmerTest>
- Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, *88*, 102–185.
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366.
- Liu, Q. (2011). Factors influencing pronunciation accuracy: L1 negative transfer, task variables and individual aptitude. *English Language Teaching*, *4*(4), 115–120.
- Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, *88*(3), 557–567.
- Mairano, P., Bassetti, B., Sokolović-Perović, M., & Cerni, T. (2018). Effects of L1 orthography and L1 phonology on L2 English pronunciation. *Revue Française de Linguistique Appliquée*, *23*(1), 45–57. <https://doi.org/10.3917/rfla.231.0045>
- Motohashi Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology*, *13*(2), 29–47.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912.
- Porretta, V. J., & Tucker, B. V. (2015). Perception of non-native consonant length contrast: The role of attention in phonetic processing. *Second Language Research*, *31*(2), 239–265.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rmisc, H. R. (2013). Rmisc: Ryan Miscellaneous. R package version, 1.5.
- Roach, P. (2004). British English: Received pronunciation. *Journal of the International Phonetic Association*, *34*(21), 239–245.
- RStudio Team. (2018). *RStudio: Integrated development for R*. Boston, MA. <http://www.rstudio.com>
- Saalfeld, A. K. (2012). Teaching L2 Spanish stress. *Foreign Language Annals*, *45*(2), 283–303. <https://doi.org/10.1111/j.1944-9720.2012.01191.x>
- Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *Journal of the Acoustical Society of America*, *134*(2), 1324–1335. <https://doi.org/https://doi.org/10.1121/1.4812767>
- 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., & 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>
- 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>
- Showalter, C. E. (2020). Russian phono-lexical acquisition and orthographic input: Naive learners, experienced learners, and interventions. *Studies in Second Language Acquisition*, *42*(2), 255–277. <https://doi.org/https://doi.org/10.1017/S0272263119000585>
- Showalter, C. E., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, *36*(1), 23–42. <https://doi.org/10.1017/S0142716414000411>
- Solier, C., Perret, C., Baqué, L., & Soum-Favaro, C. (2019). Written training tasks are better than oral training tasks at improving L2 learners’ speech production. *Applied Psycholinguistics*, *40*(6), 1455–1480. <https://doi.org/10.1017/S014271641900033X>
- Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R., & Sagisaka, Y. (2013). Non-native perception and learning of the phonemic length contrast in spoken Japanese: Training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics*, *22*(4), 373–398. <https://doi.org/https://doi.org/10.1007/s10831-013-9107-1>

- Thompson, G.B., Connelly, V., Fletcher-Flinn, C.M., & Hodson S. L.** (2009). The nature of skilled adult reading varies with type of instruction in childhood. *Memory & Cognition*, *37*(2), 223–234. <https://doi.org/10.3758/MC.37.2.223>
- Thomson, R. I., & Derwing, T. M.** (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344.
- Torgerson, C. J., & Torgerson, D. J.** (2013). *Randomised trials in education: An introductory handbook*. Education Endowment Foundation.
- Torgerson, D. J., & Torgerson, C. J.** (2008). *Designing randomised trials in health, education and the Social Sciences: An introduction*. Palgrave Macmillan.
- Turk, A., Nakai, S., & Sugahara, M.** (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schließer (Eds.), *Methods in empirical prosody research* (pp. 1–28). de Gruyter. <https://doi.org/10.1515/9783110914641.1>
- Wang, X., & Munro, M. J.** (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. *Proceedings of the 14th international congress of phonetic sciences* (Vol. 3, pp. 125–128). University of California.
- Wickham, H.** (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. <http://ggplot2.org>
- Wickham, H.** (2016). *scales: Scale Functions for Visualization. R package version 0.4.1*. <https://CRAN.R-project.org/package=scales>

---

**Cite this article:** Bassetti, B., Cerni, T., and Masterson, J. (2022). The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology. *Applied Psycholinguistics*. <https://doi.org/10.1017/S014271642200008X>