Effects of orthographic forms on second language speech production and phonological awareness, with consideration of speaker-level predictors

Bassetti, Bene; Mairano, Paolo; Masterson, Jackie; Cerni, Tania

DOI:
10.1111/lang.12423

LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Effects of Orthographic Forms on Second Language Speech Production and Phonological Awareness, With Consideration of Speaker-Level Predictors

Bene Bassetti,[a] Paolo Mairano,[b] Jackie Masterson,[c] and Tania Cerni[d]

[a]University of Birmingham, [b]University of Lille, [c]UCL Institute of Education, and [d]University of Trento

Orthographic forms (spellings) can affect pronunciation in a second language (L2); however, it is not known whether the same orthographic form can affect both L2 pronunciation and metalinguistic awareness. To test this, we asked 260 speakers of English—first-language (L1) English speakers, L1 Italian and L2 English sequential bilinguals, and L1 Italian learners of L2 English—to perform word repetition tasks and rhyme judgment tasks for word pairs containing the same consonant or vowel spelled with a letter or a digraph. L1 Italian speakers established a long–short contrast and used

---

The authors are grateful to Emanuela Buizza and Valentina De Iacovo for contributing to the acoustic analysis, Alex Panicacci for contributing to data collection, Elaine Sook Tin Chin and Becky Taylor for contributing to materials preparation, Sheila Verrier for recording spoken materials, and the *Language Learning* reviewers and editors for their insightful comments on earlier versions of this paper. 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. An earlier version of this paper was presented as a poster at the meeting of the Society for the Scientific Study of Reading, Porto, Portugal, in July 2016.

Correspondence concerning this article should be addressed to Bene Bassetti, School of Education, University of Birmingham, Birmingham B15 2TT, UK.
E-mail: b.bassetti@bham.ac.uk

The handling editors for this article were Pavel Trofimovich and Scott Crossley.

DOI: 10.1111/lang.12423

consonant and vowel length contrastively in their L2 English, both in production and in an awareness task. This provides evidence for a direct link between the effects of the same orthographic phenomenon on speech production and on metalinguistic awareness. Results were strengthened by combining experimental and qualitative data in the study of orthographic effects. Finally, the results show that proficiency predicts orthographic effects, and that orthographic effect predictors vary in naturalistic and instructed contexts.

Keywords orthography; speech production; second language; metalinguistic awareness; orthographic effects; phonological awareness; pronunciation

Introduction

Recent research has shown that the orthographic forms of the second language (L2)—the way the words are spelled—can affect how L2 speakers produce and perceive L2 sounds (Bassetti, Escudero, & Hayes-Harb, 2015). However, most research on orthographic effects on L2 phonology consists of small-scale studies of the effects of a specific orthographic form on the production of a specific sound (see below). Also, no studies that we know of link effects on production with effects on metalinguistic awareness, which would provide evidence that orthography affects L2 speakers' phonological representations. Furthermore, given the small-scale nature of the majority of studies in this domain, it has not been possible to investigate systematically factors that may predict the effect of orthography on L2 speech production, such as proficiency, short-term memory, and length of residence in a target language environment, among others.

In the present study we aimed to increase the understanding of orthographic effects on L2 phonology by linking effects on speech production and on metalinguistic awareness, and by providing a first systematic investigation of how speaker-level variables may influence the magnitude of the effects of orthography on L2 phonology. We first discuss the effects of orthography on L2 production and L2 awareness, and then present two experiments in which we investigated the effects of the number of letters in the spelling of an English sound on speech production and awareness among first-language (L1) Italian speakers of L2 English, and speaker-level predictors of these orthographic effects on L2 phonology.

Background Literature

Effects of Orthographic Forms on Second Language Speech Production

There is ample evidence that the orthographic forms of a language affect speech production and perception in L2 users (see reviews in Bassetti, 2008; Bassetti et al., 2015). Although orthographic effects can lead to sound additions, deletions, and substitutions (Bassetti, 2008), the most common effects are sound substitutions, which occur when L2 speakers substitute a target sound with another sound because of the sound's spelling. For instance, speakers of L2 German have been reported to pronounce a word-final <d> as [d] in their L2 German, although in that position <d> represents the voiceless [t] (Young-Scholten, 2002). The produced sound may be an allophone of the target sound, as, for instance, when L1 Spanish speakers of L2 American English produce a [t] or a [d] instead of a flap that is represented orthographically with the letters <t, tt, d, dd> (Vokic, 2011). L2 speakers have even been reported to produce sounds that do not exist in the target language: For instance, L1 English beginner learners of L2 Spanish may produce a [v], which does not exist in Spanish, instead of a [b] that is represented with the letter <v> (Zampini, 1994). The latter is an example of an *inter-orthography effect*, that is, an orthographic effect that arises when a L2 speaker applies L1 rules for the recoding of a unit of writing into a unit of sound in the L2 (Bassetti, 2017). These effects have been attributed to incongruences between L1 and L2 grapheme–phoneme correspondences (Hayes-Harb, Nicol, & Barker, 2010), and indeed sound discrimination in beginner and naive listeners is facilitated by L2 orthographic forms that contain L1–L2 congruent graphemes, and is disrupted by incongruent ones (Escudero, Simon, & Mulak, 2014).

Bassetti and colleagues, in a series of studies, investigated sound substitutions caused by the number of letters (one or two) in the spelling of L2 consonants and vowels, and found that the number of letters in a sound's spelling affected the duration of the sound in L2 speech production, due to the influence of L2 speakers' native grapheme–phoneme correspondences on the recoding of English orthographic forms. Participants were native speakers of Italian, a language that contrasts singleton (short) consonants and geminate (long) consonants, and represents this contrast orthographically using a single consonant letter for a short consonant and double consonant letters for a geminate (Laver, 1994). Bassetti (2017) found that L1 Italian speakers of L2 English produced English consonants as shorter when spelled with a single letter and longer when spelled with double letters, for instance producing a shorter [t] in *city* than in *kitty*. Similar effects were found with vowels, as Italians produced the same English long vowel as longer when spelled with double vowel letters, for instance producing a longer [iː] in *seen* than in *scene* (Bassetti & Atkinson, 2015). Results were confirmed cross-orthographically with L2 speakers of English who were native speakers of Japanese, a language that has contrastive length for both consonants and vowels, but is written with scripts other than the Roman alphabet (Sokolović-Perović, Bassetti, & Dillon, 2019). Another study (Bassetti, Sokolović-Perović, Mairano, & Cerni, 2018) confirmed that this long–short contrast is a genuine phonological contrast: In their L2 English, native speakers of Italian produced English homophonic words as minimal pairs distinguished by a long or short sound, for instance, producing *finish* as [ˈfɪnɪʃ], with a singleton [n], and *Finnish* as [ˈfɪʃːɪʃ], with a geminate [nː].

The literature on orthographic effects on native phonology explains such effects in two ways. One approach argues that orthography and phonology are coactivated in text-literate people (Grainger & Ferrand, 1996; Ziegler & Ferrand, 1998). This assumes that a word has an orthographic representation and a phonological representation that are linked and that are coactivated during speech production and perception. The other approach argues that there is only one representation (Muneaux & Ziegler, 2004; Taft, 2006). In text-literate speakers, the phonological representations of words are influenced by their orthographic forms, which leads to orthographic effects in speech perception and production. However, orthographic effects in L2 speakers may differ from effects in native speakers, both quantitatively, in being more frequent or stronger, and qualitatively, because effects caused by the interaction of L1 and L2 orthographies can only occur in L2 speakers (Bassetti, 2008). Bassetti (2006) argued that L2 phonological representations are likely to be influenced by L2 phonological forms reinterpreted according to L1 phonology, and L2 orthographic forms reinterpreted according to L1 orthography–phonology correspondences. If this is true, then the same orthographic effects found in L2 speakers' speech production should also be found in their L2 awareness. We tested this in the present study by examining whether the long–short contrast reported by Bassetti and colleagues for production might also be found in the metalinguistic awareness of L1 Italian speakers of L2 English.

Effects of Orthographic Forms on Phonological Awareness

Phonological awareness is the awareness of, and ability to identify and manipulate, the sounds of language: phonemes, onsets and rimes, and syllables. It has traditionally been assessed with tasks requiring, for example, the counting, segmenting, or blending of sounds in words or pseudowords. There has been evidence since the 1970s that orthographic forms affect performance in phonological awareness tasks in the native language. The effect of orthography has been demonstrated in studies showing that so-called silent letters lead native speakers to count additional phonemes in phoneme-counting tasks. For instance, L1 English literate children count one more phoneme in *pitch*/pɪtʃ/than in *rich*rɪtʃ/because of the extra letter <t> in the former (Ehri & Wilce, 1980), and L1 English adults count one more phoneme in *debt* than in *dot* (Derwing, 1992).

In spite of evidence that orthographic forms affect both metalinguistic awareness in native speakers and speech production in L2 speakers (see above), there has been limited research on orthographic effects on metalinguistic awareness in L2 speakers. Such research may contribute to our understanding of orthographic effects on L2 speech production. For instance, Detey and Nespoulous (2008) found that L2 written forms led L1 Japanese speakers of L2 French to count an extra syllable in French pseudowords containing consonant clusters that are not legal in Japanese, because they added an epenthetic vowel, whereas those who learned the pseudowords without written forms did not. These orthographic effects on phonological awareness nicely reflect Young-Scholten and colleagues' finding that naive learners of Polish produced an extra syllable by adding an epenthetic vowel when they learned the written form of L2 words, but simplified the consonant cluster when they learned only the spoken form of L2 words (Young-Scholten, Akita, & Cross, 1999).

It seems important, then, to further establish whether effects of orthography are found in both speech production and metalinguistic awareness in L2 speakers. In two separate studies, Bassetti tested effects on production (Bassetti, 2006) and awareness (Bassetti, 2007), with different participants in each study. The effect of number of letters was examined in phonological awareness tasks and in production (employing identical Chinese triphthongs in both tasks) in adult learners of L2 Chinese. Romanized Chinese, which is used to teach L2 learners, represents Chinese triphthongs with two or three letters depending on the syllable onset (e.g., [jou] is spelled <you> in onsetless syllables and <iu> after a consonant). When the triphthong was spelled with two letters, but not when it was spelled with three letters, the L2 Chinese learners omitted one vowel both in production (producing the triphthong as a diphthong) and in phoneme awareness tasks (counting and segmenting two vowels). However, because the learners were different in the two experiments, no firm conclusions could be drawn about the relationship between production and phonological awareness. In the present study, we investigated the effect of number of letters on sound length in phonological awareness and speech production in learners who participated in both tasks, and we tested whether orthographic effects on phonological awareness can predict orthographic effects on speech production. If the effect is found in both speech production and metalinguistic awareness tasks, this would be evidence that orthography affects phonological representations in L2 speakers' minds.

Predictors of Orthographic Effects on L2 Phonology

Research into orthographic effects on L2 phonology suffers from another limitation, namely, the almost complete lack of research investigating speaker-related variables that may interact with orthographic effects to affect phonological acquisition. Researchers have investigated a number of factors that may affect L2 pronunciation, generally conceptualized in terms of foreign accent or nonnativelikeness (Colantoni, Steele, & Escudero, 2015; Derwing & Munro, 2015), including age of onset of acquisition, L2 proficiency, amount of instruction, amount of exposure, motivation, aptitude, and memory. However, it is not known which of these variables may predict orthographic effects on L2 phonology.

Age of onset of acquisition is considered a main factor as it correlates with foreign accent, and nativelike pronunciation is more likely in early than late L2 learners (Abrahamsson & Hyltenstam, 2009). However, adult L2 learners can attain nativelikeness (Piske, MacKay, & Flege, 2001), and early acquisition does not always result in nativelike pronunciation (Abrahamsson & Hyltenstam, 2009). Research on length of residence, length of study, and relative amount of use of L2 versus L1 has produced contradictory results: For instance, longer length of residence was linked to more nativelike production of [ɹ] in L1 Japanese speakers of L2 English in Flege, Munro, and MacKay (1995) but not in Larson-Hall (2006). Furthermore, length of residence may not be a good measure of the amount of L2 input, as L2 speakers may avoid the L2 (Piske et al., 2001) or be exposed to nonnative input (Derwing & Munro, 2013).

Formal instruction may not be important for the acquisition of L2 phonology: Flege and Fletcher (1992) found that length of instruction only accounted for 5% of variance in L2 English foreign accent ratings, and Suter (1976) even found a negative correlation between amount of classroom training and L2 pronunciation accuracy. Moreover, L2 proficiency (TOEFL iBT® level) correlates with nativelikeness in intonation and rhythm, but not in segmental patterns (Iwashita, Brown, McNamara, & O'Hagan, 2008).

Motivation, and particularly integrative motivation (Shively, 2008), has positive effects (Moyer, 1999) on L2 pronunciation, and the importance attributed to nativelike pronunciation correlates with nativelikeness (Suter, 1976; although Elliott, 1995, found no link). There is agreement that two subcomponents of language aptitude play a role in L2 pronunciation, namely, phonemic coding ability—the ability to discriminate unfamiliar sounds and to retrieve them from memory—and the ability to mimic sounds (see Piske et al., 2001, for a review). Similarly, there is consistent evidence of the positive effects of greater short-term memory. In one study (Darcy, Park, & Yang, 2015), short-term memory in the L2 was the strongest correlate of nativelikeness. Crucially for the present study, L2 English speakers with lower phonological short-term memory rely on vowel duration to identify English vowels more than those with higher phonological short-term memory, probably because the latter are more able to use qualitative differences (Aliaga-Garcia, Mora, & Cerviño-Povedano, 2011).

Looking specifically at orthographic effects on L2 phonology, researchers found a correlation with amount of exposure to L2 orthographic input (Young-Scholten, 2002) but no link with length of residence (Young-Scholten & Langer, 2015). However, overall the little evidence available is inconclusive due to small sample sizes and lack of replication. The present study investigated whether some of the variables that affect L2 pronunciation nativelikeness may impact orthographic effects on L2 speech production and metalinguistic awareness.

The Present Study

There were two main aims in the present study:

1. to measure the effects of L2 orthographic forms in L2 speakers who took part in both speech production and awareness tasks, and to test whether effects on awareness can predict effects on production; and
2. to investigate which speaker-level factors may predict the magnitude of orthographic effects on L2 speech production.

To address these aims, we tested the effect of number of letters on sound duration in L1 Italian speakers of English as a L2 and in English native speakers, using a speech production task and a metalinguistic awareness task (rhyme judgment). We also assessed the effect of a number of relevant participant-related variables using standardized tests, ad hoc tasks, and questionnaires.

With regard to the first aim, there has been, to the best of our knowledge, no previous research directly linking the effect of orthography on speech production and on metalinguistic awareness in L2 speakers. Our study drew on recent findings showing that speakers of L2 English whose L1 has contrastive length often produce the same English sound as short when spelled with one letter and long when spelled with two letters (Bassetti, 2017; Bassetti & Atkinson, 2015; Sokolović-Perović et al., 2019). This short–long alternation is evidence of a length contrast in the L2 phonological systems of these speakers, who have two phonological categories (short and long) corresponding to the same English phonological category, and produce English homophonic pairs—such as *finish* and *Finnish*—as minimal pairs distinguished by a short and a long consonant (Bassetti et al., 2018). The present study then tested the effects of number of letters on sound length in L2 speakers who took part in tasks tapping speech production and metalinguistic awareness.

Participants were two groups of L1 Italian speakers of L2 English: The first group comprised high-school pupils studying in Italy and the second immigrants to the United Kingdom. These participants allowed us to test whether naturalistic exposure reduces orthographic effects on L2 phonology, although previous studies conducted with subgroups of the participants from the present study found no differences (Bassetti et al., 2018; Mairano, Bassetti, Sokolović-Perović, & Cerni, 2018).

Participants performed four tasks. We used two delayed word repetition tasks to test the effect of number of letters on the duration of consonants and vowels in speech production. We predicted that L1 Italian speakers of L2 English would produce a sound as long or short depending on its spelling, producing for instance a short [t] in *city* and a long [tː] in *kitty*. We used two English rhyme judgment tasks to investigate whether a speaker would accept or reject an English rhyme where the consonant or vowel is spelled with one letter in one word and with a digraph in another word (digraphs include double letters, as in *Finnish* and *cheese*, and, for vowels only, graphemes consisting of two different letters, such as <ea> in *defeat*). We predicted that the L1 Italian participants would reject English rhymes containing the identical sound spelled with different numbers of letters, such as *scholar* and *dollar*. We further investigated responses by asking participants to explain their answers in an open format, and predicted that open answers would confirm that long and short sounds are different phonological categories for L2 speakers. Finally, we predicted that level of accuracy in the English rhyme judgment tasks would predict the orthographic effect in Italians' L2 English production, because those who consider short and long sounds as different phonological categories (therefore rejecting such rhymes) are more likely to produce short and long sounds in their L2 English.

The second main aim of the study was to investigate which factors may predict the magnitude of orthographic effects in L2 speech production. Based on the findings of research on predictors of accentedness and nativelikeness in L2 pronunciation, we predicted that the following factors might predict orthographic effects: age of onset of acquisition, length of residence in a L2 environment, proportion of written input out of total (written and spoken) input, phonological short-term memory, mimicry ability for dialects and foreign languages, motivation, and the importance attributed to nativelike pronunciation.

For sequential bilinguals we also investigated the proportion of L2 use out of total (L1 and L2) language use (reading, listening, and interaction), integrative motivation, L2 emotionality as a measure of attitude toward the language (Dewaele, 2010), and English language proficiency. For instructed L2 learners, we also measured English language proficiency using a proficiency test, amount of exposure to native input (length of study with native-speaking teachers, length of study abroad), desire to learn pronunciation, pronunciation learning strategies, and motivation for pronunciation learning.

Method

Design

We investigated the effect of number of letters in the spelling of an English consonant or vowel on sound length in speech production (delayed word repetition tasks) and on accuracy in phonological awareness tasks (rhyme judgment tasks). The within-group factor was target sound spelling, with two levels: single letter or digraph. Language background was a between-group factor with three levels: English native speakers, L1 Italian and L2 English sequential bilinguals, and L1 Italian learners of L2 English. We used a spelling task to confirm participants' knowledge of the spellings of the target sounds. We used open questions to obtain explanations from participants regarding their responses in the phonological awareness tasks. We also measured a number of variables that may predict orthographic effects, using memory tasks, proficiency tests, and questionnaires.

Participants

There were 260 participants: 80 English native speakers, 80 L1 Italian and L2 English sequential bilinguals, and 100 L1 Italian learners of L2 English. Participants had no reported visual, hearing, language, or reading difficulties; 7% were left-handed. The learners were younger ($M = 16.91$ years, $SD = 0.45$) than the native speakers ($M = 26.31$, $SD = 10.77$) and bilinguals ($M = 26.31$, $SD = 10.77$).

All the L2 learners spoke the Roman variety of Italian; the bilinguals' native varieties were central (34%, including Lazio, Marche, Tuscany, and Umbria) and southern (66%, including Abruzzo, Calabria, Campania, Molise, Apulia, and Sicily; five answers were missing). An analysis of a subset of participants showed that learners' and bilinguals' geminate-to-singleton ratio in L1 Italian did not differ (Mairano et al., 2018). The English native speakers mostly (66%) reported their native varieties as Standard British English or Received Pronunciation; the others originated from the Midlands, Jersey, Manchester, Northern Ireland, Wales, and Yorkshire (five answers were missing).

The learners were students at one of three high schools in Rome, who were studying English for 3 hours a week as a compulsory school subject, using British English textbooks. They had never been in an English-speaking country, or had only been in one briefly (*Mdn* = 0 months, range 0–4 months, except for one participant who reported having been in an English-speaking country for 36 months at age 4 years, nevertheless, this participant was included in the sample as his/her scores did not exceeed those of the other participants). The bilinguals were Italians who had studied English in Italian schools, had moved to the United Kingdom after the age of 18, and had been living there for 6.43 years on average (*SD* = 5.23). The bilinguals had a later age of onset of acquisition (*M* = 9.36 years, *SD* = 2.3) than the learners (*M* = 5.78 years, *SD* = 1.88), because most current high-school students started English in primary school, whereas previously it was generally taught after primary school. Bilinguals reported using English much more than learners (median values in hours per week: for reading, 15 for bilinguals vs. 2 for learners; for listening, 14 for bilinguals vs. 4 for learners; for speaking: 25 for bilinguals vs. 0.5 for learners).

Materials

For lists of stimuli, see the Supporting Information online: Appendix S1 for production tasks and Appendix S2 for awareness tasks. Stimuli are also available at the Open Science Framework (https://osf.io/p3q6d) and on IRIS https://www.iris-database.org.

Delayed Word Repetition

Materials were 20 word pairs in a consonant production task and 20 word pairs in a vowel production task. Each pair contained the same target sound, in the same phonological context, spelled with a single letter in one word of the pair and with a digraph in the other, such as the /p/ in *copy* and *floppy* (/ˈkɒpi/ and /ˈflɒpi/ in British English). We used two additional pairs for the pretesting training session of each task. Overall, based on word frequencies from the British National Corpus (accessed from http://corpora.lancs.ac.uk/BNCweb/), the words were more frequent in written than in spoken language (written:spoken frequency ratios: *Mdn* = 0.50 for words with single-letter spellings of the target sound; *Mdn* = 0.60 for words with digraph spellings), and words with single-letter spellings of the target sound were descriptively more frequent than words spelled with a digraph (*Mdn* = 18.71 vs. 12.16, respectively; $W = 954$, $p = .089$).

Consonant production task. The target consonants were the three English voiceless plosives: /p/, /t/, and /k/ (with seven, nine, and four word pairs for each plosive, respectively). Both words in a pair had:

- the target consonant in the same intervocalic context (for instance, the target /p/ is between /ɒ/ and /i/ in both *copy* and *floppy*);
- the same number of syllables (80% of pairs were disyllabic, the rest longer);
- the same primary stress position (post-tonic—after a stressed vowel—in 85% of pairs, because tonic position affects duration of Italian geminates; Payne, 2005); and
- the same morphological structure (the target consonant did not occur at the morpheme boundary in derived words, because fake gemination may arise in English when a derivational morpheme shares a phoneme with the base, for example, in the [n] of *unnamed*; Davis, 2011).

Six of the CC-words had Italian cognates. Five were orthographically congruent (i.e., the target sound is spelled with double letters in both English and Italian cognates, as with *attic*, cognate of Italian *attico*) and one was orthographically incongruent (i.e., the target sound is spelled with a single letter in the Italian cognate, as with *pepper*, cognate of *pepe*). Previous research found no effects of cognate status on the pronunciation of L2 English CC-sounds (Bassetti, 2017; Bassetti et al., 2018).

Vowel production task. Targets were the long vowels /iː/, /ɔː/, and /uː/, and the diphthong /əʊ/. Although /ɔː/ is less frequent than /iː/ and /uː/, it appears in very frequent words (e.g., *daughter*, *water*, *door*). The diphthong /əʊ/ was included because Italian speakers generally produce it as [oː] or [ɔː], and therefore its duration could be affected by orthography. To avoid potential confounds due to the phonetic context, we ensured that the target vowel within each word pair was followed by the same consonant (except for the pairs *movie–moody* and *chic–sheet*, where it was not possible to find suitable words; in these two pairs, consonants were matched in voice to avoid the confound of pre-fortis clipping; see Klatt, 1976).

Twenty digraph vowels were spelled with double letters (VV-pairs), such as /iː/ in *sees*, and 20 with a digraph containing different letters (V_1V_2-pairs), such as /ɔː/ in *boarder*. This combination of VV- and V_1V_2-pairs was chosen to be representative of the English language. We predicted that Italian speakers would produce more long vowels with double-letter spellings than with other digraphs (Bassetti et al., 2018). In eleven V-words, the length of the vowel was represented orthographically by the digraph <V_e> (e.g., *June*, /dʒuːn/). However, there is evidence that Italian teachers and students generally do not know that the so-called “silent e” marks the preceding vowel as long (Bassetti & Atkinson, 2015), and this was confirmed by participants in the present study.

We took into account variables that may affect vowel duration as follows. First, words in a pair had the same number of syllables, as this can affect vowel length (Klatt, 1973). Second, all but one of the target vowels were in a stressed closed syllable, because Italians may lengthen vowels in other positions (Bertinetto & Loporcaro, 2005). Third, because the voicing of the following consonant can affect English vowel duration (Klatt, 1976), 63% of vowels in closed syllables were followed by a voiced consonant and the rest by a voiceless one. Finally, one VV-word was a loanword in Italian (*zoom*).

Phrases were recorded by a female native speaker of Southern British English, in a sound-attenuated room using a Røde NT2-A microphone connected to an Alesis Multimix 12 Firewire mixer. Color pictures were selected from the Art Explosion library (Nova Development, 2004).

Rhyme Judgment

There were two tasks: a C–CC rhyme judgment task and a V–VV rhyme judgment task. Materials were 72 word pairs. The targets were 24 homophonic rhymes (rhymes having identical phonological forms but different orthographic forms), which contained the same target sound (12 target vowels and 12 target consonants) spelled with a single letter or a digraph, such as *scholar* and *dollar* (/ˈskɒlə/ and /ˈdɒlə/) or *machine* and *between* (/məˈʃiːn/ and /bɪˈtwiːn/). The control stimuli were 24 homophonic rhymes with orthographic differences other than number of letters (e.g., *lie* and *cry* (/laɪ/ and /kraɪ/). The filler items were 24 non-rhyming word pairs (e.g., *fox* and *blogs* (/fɒks/ and /blɒgz/). Two additional rhymes were created for the training session for each task.

C–CC rhyme judgment task. The target consonants were:/b/, /n/, /r/, /s/, and /l/. We considered three variables that may affect Italians’ rhyme judgments. First, we considered the legality of gemination in Italian: Gemination is not attested in word-final position in Italian native words (Bertinetto & Loporcaro, 2005), and Italian speakers of L2 English produce shorter consonants in English words containing double letters in word-final position than in those containing double letters in intervocalic position (Bassetti et al., 2018). To test for the effects of legality in L1 phonotactic rules, we included eight rhymes with intervocalic double consonants (e.g., *dollar–scholar*) and four with word-final ones (e.g., *hell–excel*). Second, we considered gemination position: No CC-sound occurred at the morpheme boundary in a compound word. Third, we considered cognate status: Of the three cognates used in this task, two were orthographically congruent with Italian and one incongruent.

V–VV rhyme judgment task. The targets were the long vowels /iː/, /ɔː/, and /uː/, and the diphthong /əʊ/. The three long vowels were spelled as V or VV (e.g., /iː/ in *me–free*); /əʊ/ was spelled as V or V_1V_2 (e.g., *alone–grown*). The target vowel was always lexically stressed, and in most cases it occurred in a closed syllable (i.e., it was followed by a consonant; three were in word-final position). There were no cognates or loanwords.

Open Questions About Rhyme Judgment

We used two additional English word pairs and two Italian word pairs, containing the same target double-letter graphemes <rr> and <oo>, to elicit L2 speakers' criteria for accepting or rejecting a rhyme in both languages: English consonants: *very–merry*; English vowels: *rule–fool*; Italian consonants: *caro–carro*; Italian vowels: *so–zoo*. Participants performed first the English and then the Italian task. They read each rhyme aloud, then decided whether it was a rhyme or not, and explained why.

Speaker Variables

Participants performed two standardized short-term memory tests—a digit recall task and a pseudoword recall task—from the Comprehensive Test of Phonological Processing, Second Edition (Wagner, Torgersen, Rashotte, & Pearson, 2013).

The questionnaire collected bio-measures (gender, age) and language background (native language[s], L2s). For L2 speakers, we also collected the following information: age of onset of acquisition; length of study; length of study with native-speaking teachers; length of residence in an English-speaking country; self-reported amount of English reading, listening, and interaction; rated importance of a nativelike pronunciation; and self-reported ability to imitate dialects and foreign accents. The learners also filled in a multiple-choice question about their motivations for learning English pronunciation; rated their desire to learn pronunciation; completed translated and adapted versions of the Pronunciation Attitude Inventory (Elliott, 1995) and the Pronunciation Learning Strategy Inventory (Berkil, 2009); and took the Oxford Placement Test (Allan, 1992), which tests listening comprehension, reading, and grammar. The bilinguals also filled in their self-reported amount of Italian reading, listening, and interaction; their self-assessed English proficiency in the four skills, using the levels of the Common European Framework of Reference for Languages or CEFR (Council of Europe, 2001); six 7-point scales measuring the emotionality of the English and Italian languages (Dewaele, 2010); and a measure of integrative orientation (Vancouver Index of Acculturation; Ryder, Alden, & Paulhus, 2000).

Individual Task Procedures

Delayed word repetition. The delayed word repetition task came from Bassetti (2017), except that the carrier sentence used in Step 4 (see below) was replaced with one suitable for the words included in the present study.

There were four steps in the task:

1. See a picture, and listen to a native speaker's production of an accompanying six-word phrase or sentence.
2. Count from seven to one in English to eliminate traces of the native speaker's production from phonological memory, and repeat the phrase (to ensure the participant understood and remembered the phrase).
3. Listen to a truncated version of the phrase from which the target word and all that followed was deleted.
4. Repeat the first (target) word missing from the truncated phrase in a carrier sentence ("The word ___ should follow") three times.

The carrier sentence ensured that the target word was in the nuclear position within the intonational unit. The three repetitions in step 4 produced a mean duration for each target sound.

Participants in the bilingual and English native speaker groups were presented with 80 trials, with the 20 consonant and 20 vowel word pairs. The total task duration was approximately 30 minutes. Pilot testing revealed that the task was too demanding for the L2 learners. They were therefore presented with 40 trials each (10 consonant pairs and 10 vowel pairs), from one of two counterbalanced lists.

Spelling task. This task tested participants' knowledge of the spelling of the target sounds from the delayed repetition task. Participants saw the pictures from the repetition task, presented in random order. On each trial, below the picture, participants saw the written form of the phrase they had heard in the production task, with a gap where the target word should have appeared. After clicking on a button on the screen, participants heard the target word, then typed the word using the computer keyboard, and pressed "enter" to start the next trial. If participants did not recognize a target word, they could listen to it up to three times by pressing three buttons on the computer screen. There were no time limits.

Rhyme judgment. Participants decided whether word pairs rhymed or not by pressing one of two buttons on a response box. The first half of the set of word pairs were presented within verses, and the second half of the set as word pairs in isolation. In the verse part, a two-line verse appeared in the centre of the computer screen (e.g., "Chocolate with a cherry, ask me if it's good and I'll say very"). In the word pair part, the two words were presented one above the other in the centre of the screen. Stimuli were presented in written format because some of the verses would have been difficult for the L2 learners to understand in the spoken modality.

Before the main rhyme judgment testing began, four verse trials were presented, with feedback, to clarify that they should evaluate sounds and not spellings, by, for instance, explaining that "I" and "lie" rhyme. Feedback was not provided during the main testing session.

After the task, participants were shown the English C-rhyme *very–merry*. They were asked to say whether this was a rhyme or not and to explain why. This was repeated with the English V-rhyme *rule–fool*. L2 speakers (i.e., the bilinguals and the learners) then also had to make a rhyme judgment and explain their judgment about the Italian C-rhyme *caro–carro* and then the V-rhyme *so–zoo*.

General Procedure

Participants were tested individually in the word repetition, spelling, and rhyme judgment tasks, in a sound-attenuated or quiet room in the presence of a researcher. Questionnaires were completed at home or at the end of the session according to the participant's preference. Participants within each group completed tasks in the same order; in all tasks, order of presentation of stimuli was randomized. Participants could take a rest whenever needed, both during and between tasks.

Native speakers and bilinguals were tested by a bilingual researcher with L1 Italian and L2 English on university premises in the United Kingdom. Participants first performed a perception task (to be reported elsewhere), followed by the word repetition task, the spelling task, rhyme judgment, the open questions, the memory tasks, and finally, a short reading-aloud task (the results of which were reported in Bassetti et al., 2018). The session lasted 1.5 hours for English native speakers, and 2 hours for the bilinguals, due to the additional tasks in Italian and slower overall performance.

The L2 learners were tested by an Italian researcher (the fourth author) in their school during normal school hours. In order to fit the sessions within normal schooling hours, the L2 learners were seen in two sessions that were between 1 and 3 weeks apart. The order of tasks was arranged so that each session would last approximately 50 minutes: The first session consisted of the perception and memory tasks; the second session consisted of word repetition, spelling, and reading-aloud tasks. The English proficiency test was administered during a normal English language class.

Participation in the study was voluntary. The learners received book tokens (as requested by parents or carers); bilinguals and native speakers were rewarded with money. The project received ethical approval from the Humanities and Social Sciences Research Ethics Committee of the University of Warwick.

Equipment

The word repetition, spelling, and rhyme judgment tasks were run on an Apple laptop running PsyScope X software (Cohen, MacWhinney, Flatt, & Provost, 1993), which controlled stimulus presentation and randomization, and recorded responses. Participants listened to recordings through headphones and interacted with the computer using a mouse or IoLab response box. Spoken production was recorded either with a Zoom H4N Pro digital recorder connected to a Shure SM10A headset microphone or to an AKG HSD171 headset with dynamic microphone, or with a Røde NT2-A microphone connected to an Alesis Multimix 12 Firewire mixer. Responses in the spelling task were entered using a keyboard.

Data Analysis

Acoustic Analysis

The duration of each target sound in the repetition task was measured on the acoustic signal using the Praat software (Boersma & Weenink, 2016), following standard criteria (Turk, Nakai, & Sugahara, 2006). Consonant duration was measured as the duration of closure, which is the primary cue for gemination in Italian plosives (Esposito & Di Benedetto, 1999). Vowel duration was measured from the onset to the cessation of a clear formant pattern, especially relying on F2. For vowels followed by a fricative, nasal, or approximant, the vowel boundary was marked in correspondence of visible spectral discontinuity. In order to obtain a reliable measure, the average of the measurements from the three repetitions for each target sound was calculated.

Sound duration measurements were made by three expert phoneticians, of whom two were blind to the experimental hypotheses and one was the second author. Their intraclass correlation coefficients, calculated for each of the three pairs on 5% of the data, ranged from .97 to .98, all ps $< .001$.

Statistical Analysis

Statistical analyses were performed using R version 3.2.4 (R Core Team, 2018) with RStudio (RStudio Team, 2016) and the following additional packages: bear (Lee & Lee, 2014), car (Fox & Weisberg, 2019), doBy (Højsgaard & Halekoh, 2016), ggplot2 (Wickham, 2009), lme4 (Bates, Maechler, Bolker, & Walker, 2015), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017), lsmeans (Lenth, 2016), MuMIn (Bartoń, 2018), Publish (Gerds & Ozenne, 2018), and RePsychLing (Bates, Kliegl, Vasishth, & Baayen, 2015).

Delayed word repetition. Of the 16,800 word tokens, 5.7% (or 954) were missing, wrong, or not analysable acoustically. An additional 6.5% (1,036) were eliminated from the analysis because the participant had not spelled the target sound correctly in the spelling task. A long-sound-to-short-sound duration ratio was obtained for each of the remaining 6,634 word pairs, by dividing the duration of the target sound in the digraph word by its duration in the single-letter word (e.g., by dividing the duration of /t/ in *kitty* by the duration of /t/ in *city*).

Before modeling the data, we removed as outliers the top and bottom 1% of data for each group (native speakers, learners, and bilinguals), leaving 3,479 C–CC and 3,019 V–VV pairs, and then implemented a natural logarithmic transformation to address positive skew. We performed linear mixed effects modeling of the log-transformed ratios with the lmerTest package. The initial maximal mixed effects models (Barr, Levy, Scheepers, & Tily, 2013) included the fixed effect of language background (native speakers, bilinguals, learners), a random intercept for participant, a random slope for the effect of type of rhyme by participant, and a random intercept for word pairs. To simplify models, we used function rePCA in the RePsychLing package when random structure was overfitted (Bates et al., 2015), and we proceeded with model reduction using likelihood ratio tests. The Results section reports model summaries as tables, and results from Wald chi-square tests for significant fixed effects. We calculated geometric means and confidence intervals using function ci.mean in the Publish package; we calculated conditional and marginal R^2 using function r.squaredGLMM in the MuMIn package; finally, we performed group comparisons using function contrast in the lsmeans package with Tukey's method for adjusting p values. Visual inspection of fitted and residual values and Q–Q plots of residuals did not show deviations from linearity, homoscedasticity, or normality in any of the models.

Rhyme judgment. Data from 12 participants were lost or eliminated (five native speakers due to technical issues, and three bilinguals and four learners for incorrectly considering the Italian C–CC word pair *caro–carro* as a rhyme).

Because the outcome was binary, two logit mixed-effects models (Jaeger, 2008) were fitted to consonant and vowel data using the function glmer in the lmerTest package. The initial maximal models included the main effects and interaction between language background (native speakers, bilinguals, learners) and type of rhyme (control rhyme, C–CC/V–VV rhyme), as well as a random intercept for participant, a by-participant random slope for type of rhyme, and a random intercept for word pairs. Finally, to test whether higher metalinguistic awareness of sound length predicts lower duration ratio in speech production in L2 speakers, two models with mean log consonant or vowel duration ratios as the outcome variable were used. Two maximal mixed effects models were fitted to consonant and vowel data, with the interaction and fixed effects of language background (bilinguals, learners) and grand-mean centered awareness scores, and random intercepts for participants and word pairs, following the same procedure as in the word repetition analysis. Consonant and vowel reaction times were also analyzed using two maximal mixed-effect models with the same structure as used for the accuracy model, after eliminating the top and bottom 1% and carrying out log transformation of the data. For model reduction and results reporting, we followed the same procedure as used for the delayed word repetition task.

Speaker variables. Responses in the memory tasks and the proficiency test were scored following standard procedures. Scores in digit memory and pseudoword repetition memory were highly positively correlated; therefore, a composite short-term memory score was obtained by adding the z-scores together. The proportion of written input out of total input in English was obtained by dividing the participant's self-reported weekly number of reading hours by the sum of their reading and listening hours. The proportion of written L2 input was obtained by dividing the self-reported number of English reading hours by the sum of English and Italian reading hours, and the proportion of spoken L2 input was obtained similarly from self-reported listening hours. A composite mimicry ability measure was obtained from the self-rated dialect and foreign language mimicry scores, which were significantly positively correlated, Kendall's tau $= .46$, $p < .001$. A mean emotionality score for English was obtained by subtracting the Italian from the English emotionality rating for each item in the emotionality rating task, and a mean integrative orientation score was obtained by subtracting the Italian from the English rating for each item in the Vancouver Index. The participant's number of pronunciation learning strategies was obtained by counting the number of strategies each participant ticked in the Pronunciation Learning Strategy Inventory.

The results for the bilingual and learner groups were analyzed separately. This was necessary because they differed significantly on most variables: For instance, age of onset would not be a meaningful predictor across groups because bilinguals had later ages of onset (see Participants section). Visual inspection of fitted and residual values and Q–Q plots of residuals for the regression model of orthographic effects did not show obvious deviations from linearity, homoscedasticity, or normality in any regression model; in the only model with more than one predictor (orthographic effects in bilinguals' consonant production), variance inflation factors (all < 1.3) did not reveal collinearity.

Qualitative Analysis of Explanations Following Rhyme Judgment Test. Data from 25 L2 speakers were eliminated from the analysis of the responses to the open questions because their answers were not valid or else were lost due to technical issues. Responses were transcribed and coded using one of the five categories that emerged from the qualitative analysis. Categories describing the target sound were "double," "long," "strong/harsh," or "target sound (other)." Responses that described sound differences other than the target sound were categorized as "other" (e.g., word stress).

Results

Consonants

Effects of Language Background and Number of Consonant Letters on Consonant Duration in Speech Production

The dependent variable was the logged ratio of duration of the same consonant when spelled with double letters (CC) and when spelled with a single letter (C). The geometric mean of the CC:C ratio was 1.01 among English native speakers (95% CI [1.00, 1.02]), 1.33 among bilinguals (95% CI [1.31,1.35]), and 1.53 among L2 learners (95% CI [1.50, 1.56]).

The final model (see Table 1) included a fixed effect of language background (native speakers, bilinguals, learners), and random intercepts for participants and word pairs. There was a main effect of language background ($\chi^2 = 539.8, p > .001$). Post hoc tests revealed that the mean CC:C ratio was higher in the two L2 speaker groups than among native speakers (for bilinguals, $t = 15.480$; for learners, $t = 22.72$), and among learners than among bilinguals ($t = 7.436$; all ps $< .001$).

Effects of Language Background and Number of Consonant Letters on Consonant Length Awareness

Figure 1 shows the mean percentage of correct responses in the C–CC rhyme judgment task.

Table 1 Results of mixed-model analysis of the effects of language background (native speakers, bilinguals, or L2 learners) on CC:C duration ratio

Random effects	Variance	*SD*
Participant Intercept	0.010	0.099
Word pair Intercept	0.004	0.064
Residual	0.047	0.217

Fixed effects	Estimate (*b*)	*SE*	95% CI	*t*	*p*
Intercept	0.012	0.019	[−0.251, 0.049]	0.641	.524
Language background					
Natives vs. bilinguals	0.276	0.018	[0.241, 0.311]	15.480	<.001
Natives vs. L2 learners	0.409	0.018	[0.374, 0.445]	22.723	<.001

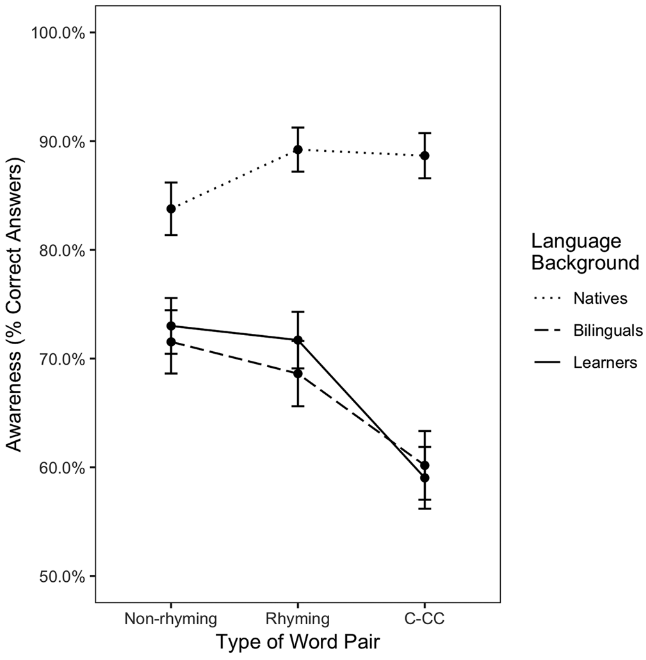

Figure 1 Mean percentage of correct rhyme judgment responses for native speakers, bilinguals, and learners for nonrhyming control, rhyming control, and C–CC stimulus pairs. Error bars represent 95% CI.

Table 2 Results of mixed-model analysis of the effects of language background (native speakers, bilinguals, or L2 learners) and type of rhyme (nonrhyming control, rhyming control, or C–CC) on accuracy in the C–CC rhyme judgment task

Random effects	Variance	*SD*
Participant		
Intercept (nonrhyming pair)	0.840	0.917
Nonrhyming vs. rhyming pair	1.911	1.382
Nonrhyming vs. C–CC pair	1.619	1.272
Word pair		
Intercept	0.912	0.955

Fixed effects	Estimate (*b*)	*SE*	95% CI	*z*	*p*
Natives					
Intercept (nonrhyming pairs)	2.176	0.316	[1.557, 2.795]	6.891	<.001
Nonrhyming vs. rhyming pairs	0.807	0.460	[−0.094, 1.708]	1.755	.079
Nonrhyming vs. C–CC pairs	0.423	0.449	[−0.457, 1.302]	0.941	.035
Nonrhyming pairs					
Natives vs. bilinguals	−0.971	0.201		−4.830	<.001
Natives vs. L2 learners	−0.855	0.192		−4.447	<.001
Nonrhyming vs. rhyming pairs					
Natives vs. bilinguals	−0.928	0.308		−3.011	.003
Natives vs. L2 learners	−0.855	0.296		−2.892	.004
Nonrhyming vs. C–CC pairs					
Natives vs. bilinguals	−1.088	0.286		−3.811	<.001
Natives vs. L2 learners	−1.276	0.274		−4.658	<.001

The model (see Table 2) that best explained the likelihood of a correct response included the interaction and fixed effects of language background (native speakers, bilinguals, learners) and type of word pair (nonrhyming control, rhyming control, C–CC), a random slope for the interaction between type of word pair and participant, and random intercepts for participant and for word pairs. There was an effect of language background ($\chi^2 = 178.33$, $p < .001$), and crucially there was an interaction between language background and type of rhyme ($\chi^2 = 24.72$, $p < .001$). In order to perform pairwise comparisons, results from rhyming and non-rhyming controls were merged into a single "control" category, and separate models were run for each language background. Although descriptively all L2 speakers were more likely to answer correctly a control than a C–CC word pair, the difference was significant only in the learner group ($b = -0.738$, 95% CI [−1.432, −0.045], $SE = 0.354, z = -2.087, p = .037$).

Reaction times were overall faster in native than in L2 speakers (natives vs. bilinguals: $b = 0.410, SE = 0.050$; natives vs. learners: $b = 0.541, SE = 0.047$; both ps $< .001$), but there was no effect of type of word pair or interaction.

Qualitative data confirmed and explained the quantitative findings for L2 speakers: 73% of bilinguals and 78% of learners incorrectly rejected the rhyme *very–merry*. Crucially, among those who rejected the rhyme and provided a valid explanation, 92% rejected the rhyme because the /r/ in *merry* is double, long, or stronger/harsher (bilinguals: 90%, $n = 31$; learners: 93%, $n = 60$). Examples of explanations include the following: "*Merry* has the double and *very* doesn't" (ESL36; all translations by first author); "they don't rhyme, because *merry* is longer than *very*" (BL67); "I pronounce *r-r-y* stronger, I pronounce the *r* in *very* as weaker" (ESL32). Some explicitly related English CC-consonants to Italian geminates: "[In *merry*] the *r* has a stronger sound, it's like the double in Italian" (BL34). The remaining 8% of valid responses mentioned sound differences other than the target consonant, such as word stress position. Furthermore, over a third of those who correctly considered *very–merry* a rhyme spontaneously clarified that double consonant letters do not represent long consonants in English, stating for instance that "[in English] double r is read only once" (ESL97). The reasons for rejecting C–CC rhymes were similar in English and Italian; *caro–carro* was mostly rejected because of the different pronunciation of the /r/ (93% of valid answers; bilinguals = 89%, learners = 95%).

The Effects of Awareness of Consonant Length on Consonant Duration in L2 Speakers

The final model (see Table 3) included fixed effects of L2 speakers' language background (bilinguals, learners) and of grand-mean centered accuracy in the C–CC rhyme judgment task (a measure of phonological awareness), and random intercepts for participants and word pairs. In both bilinguals and learners, higher awareness that consonant length is not contrastive in English was associated with lower CC:C duration ratio in speech production ($b = -0.181$, 95% CI [−0.335, −0.027], $SE = 0.079, t = -2.30, p = .023$).

Speaker variables. Multiple regression analysis was used to identify variables that predicted the magnitude of the effect of consonant spelling on L2 speakers' consonant duration (logged CC:C ratio). This revealed that weaker orthographic effects in bilinguals were predicted by proficiency (Oxford

Table 3 Results of mixed-model analysis of the effects of language background (native speakers, bilinguals, or L2 learners) and metalinguistic awareness of consonant length on CC:C duration ratio

Random effects	Variance	*SD*
Participant Intercept	0.015	0.124
Word pair Intercept	0.009	0.093
Residual	0.049	0.221

Fixed effects	Estimate (*b*)	*SE*	95% CI	*t*	*p*
Intercept	0.418	0.062	[0.296, 0.541]	6.699	<.001
C length awareness	−0.181	0.079	[−0.335, −0.027]	−2.298	.023
Bilinguals vs. learners	0.134	0.022	[0.091, 0.177]	6.140	<.001

Placement Test; $b = -2.731$, $p = .006$), short-term memory ($b = -2.173$, $p = .030$), and proportion of written input ($b = -3.623$, $p < .001$), adjusted $R^2 = .043$, $F(7, 774) = 18.02$, $p < .001$. Among learners, higher proficiency predicted weaker orthographic effects ($b = -0.001$, $p = .017$), adjusted $R^2 = .01$, $F(2, 644) = 5.86$, $p = .003$.

Vowels

Effects of Language Background and Number of Vowel Letters on Vowel Duration in Speech Production

The geometric mean of the VV:V ratio was 1.03 among English native speakers (95% CI [1.02, 1.04]), 1.10 among bilinguals (95% CI [1.08, 1.12]), and 1.07 among L2 learners (95% CI [1.05, 1.09]).

The final model (see Table 4) included a fixed effect of language background (native speakers, bilinguals, learners) and a random intercept for word pairs. There was a main effect of language background, $\chi^2 = 43.58$, $p < .001$). The mean VV:V ratio was higher in the two L2 speaker groups than among native speakers (post hoc contrasts for bilinguals, $t = -6.625$, $p < .001$; for learners, $t = -2.382$, $p = .046$), and among bilinguals than among learners ($t = 2.799$ $p = .014$).

Effects of Language Background and Number of Vowel Letters on Vowel Length Awareness

Figure 2 shows the mean percentage of correct responses in V–VV rhyme judgment.

Table 4 Results of mixed-model analysis of the effects of language background (native speakers, bilinguals, or L2 learners) on VV:V duration ratios

Random effects	Variance	*SD*
Word pair Intercept	0.009	0.096
Residual	0.040	0.199

Fixed effects	Estimate (*b*)	*SE*	95% CI	*t*	*p*
Intercept	0.038	0.022	[−0.005, 0.817]	1.725	.099
Language background					
Natives vs. bilinguals	0.054	0.008	[0.038, 0.070]	6.625	<.001
Natives vs. L2 learners	0.024	0.010	[0.004, 0.045]	2.382	.017

The model (see Table 5) that best explained the likelihood of a correct response included the fixed effects of language background (native speakers, bilinguals, learners) and type of word pair (nonrhyming control, rhyming control, V–VV), and random intercepts for participant and word pairs. There were effects of language background ($\chi^2 = 120.77$, $p < .001$) and type of word pair ($\chi^2 = 55.92$, $p < .001$), but no interaction, showing that native speakers were overall more accurate than either L2 group, and that control rhymes were judged more accurately than V–VV rhymes across groups.

Reaction times were overall faster in the English native speaker group than in the two L2 speaker groups ($\chi^2 = 249.07$, $p < .001$), but there was no effect of type of word pair or interaction.

Looking at the qualitative data, unlike with consonant results, only 30% of L2 speakers incorrectly rejected the *rule–fool* rhyme (bilinguals: 37%; learners: 26%), and vowel length was not their main reason (44% of rejections with valid explanations; bilinguals: 47%; learners: 42%): Just as many participants (42%) identified differences between the two vowels other than length (bilinguals: 37%; learners: 46%). Answers referring to vowel length were similar to those given for consonants, including, for instance, "*fool* is longer than *rule*" (BL67) and "I think this *u* is longer" (ESL068). However, participants identified differences other than length between the two vowels, such as that "the first [word] has a slightly more closed vowel" (ESL095). Regarding Italian *so–zoo*, 39% of Italians correctly rejected it as a rhyme (bilinguals: 47%; learners: 33%), in 50% of cases because of a double or longer vowel (bilinguals: 62%; learners: 39%).

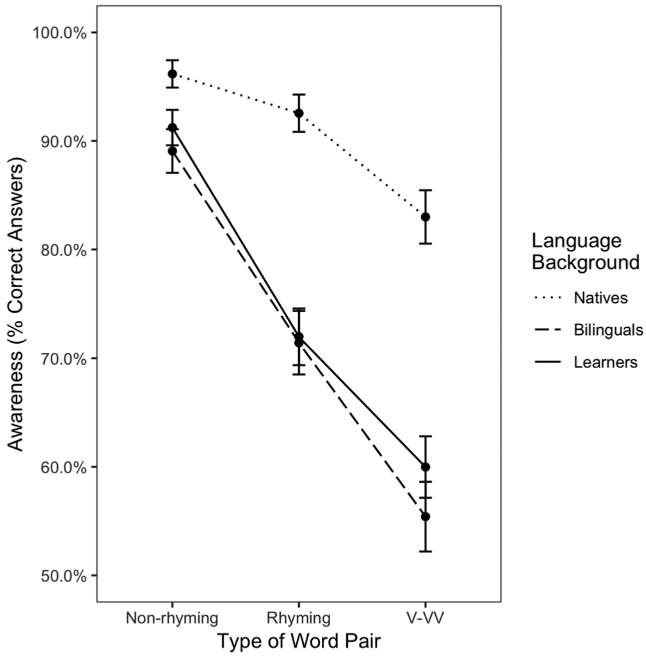

Figure 2 Mean percentage of correct rhyme judgment responses for native speakers, bilinguals, and learners for nonrhyming control, rhyming control, and V–VV stimulus pairs. Error bars represent 95% CI.

The Effects of Awareness of Vowel Length on Vowel Duration in L2 Speakers
The final model (see Table 6) included the fixed effect of L2 speakers' language background (bilinguals, learners) and a random intercept for word pairs. Unlike with consonants, accuracy in the V–VV rhyme judgment task was not associated with vowel duration ratios in speech production.

Speaker variables. Among bilinguals, no predictor was correlated with the orthographic effect (logged VV:V ratio). Among learners, Pronunciation Attitude Inventory score predicted weaker orthographic effects ($b = -0.048$, $p = .014$), adjusted $R^2 = .009$, $F(1, 575) = 6.10, p = .014$.

Discussion

This study aimed to link the effect on L2 speakers' speech production and metalinguistic awareness of an orthographic phenomenon that arises when a L2 speaker applies L1 rules for recoding print to sound. A further aim of the

Table 5 Results of mixed-model analysis of the effects of language background (native speakers, bilinguals, or L2 learners) and type of rhyme (nonrhyming control, rhyming control, or V–VV) on accuracy in the V–VV rhyme judgment task

Random effects			Variance		SD
Participant Intercept (nonrhyming pair)			0.809		0.900
Word pair Intercept			0.598		0.774
Fixed effects	**Estimate (*b*)**	***SE***	**95% CI**	***z***	***p***
Intercept (nonrhyming)	4.508	0.277	[3.965, 5.052]	16.261	<.001
Nonrhyming vs. rhyming	−1.697	0.460	[−2.350, −1.045]	−5.097	<.001
Nonrhyming vs. V–VV pairs	−2.442	0.333	[−3.094, −1.790]	−7.341	<.001
Natives vs. bilinguals	−1.761	0.175	[−2.104, −1.419]	−10.078	<.001
Natives vs. L2 learners	−1.599	0.167	[−1.927, −1.271]	−9.553	<.001

Table 6 Results of mixed-model analysis of the effects of language background (native speakers, bilinguals, or L2 learners) and mean scores in the V–VV rhyme judgment task on VV:V duration ratio

Random effects			Variance		SD
Word pair Intercept			0.016		0.126
Residual			0.050		0.223
Fixed effects	**Estimate (*b*)**	***SE***	**95% CI**	***t***	***p***
Intercept	0.094	0.029	[0.036, 0.151]	3.210	.004
Bilinguals vs. learners	−0.027	0.012	[−0.051, −0.004]	−2.302	.021

study was to investigate speaker-level predictors of the orthographic effect on L2 speech production. In brief, the results were as follows.

The results from a delayed repetition task revealed that L1 Italian speakers of L2 English produced consonants and vowels as long or short depending on their spelling in English words, whereas English native speakers did not. For consonants the effect was stronger in learners than in sequential bilinguals, whereas for vowels the impact was slightly stronger for bilinguals than for learners.

In a metalinguistic awareness (rhyme judgment) task, L1 Italian participants rejected English homophonic word pairs spelled with single and double consonant letters. As in the repetition task involving consonants, the effect was stronger among learners than sequential bilinguals. L2 speakers' performance on the consonant awareness tasks predicted the orthographic effect on speech production. Looking at speaker-level factors, English language proficiency predicted the magnitude of the orthographic effect for consonant production, with higher proficiency predicting a weaker effect. For vowels, learners with a higher desire to learn English pronunciation had a lower tendency to produce homophonic targets with different durations.

Effects of Orthographic Forms on Consonant Production and Awareness

Consonant Production

The results of the delayed repetition task revealed that the L1 Italian speakers of L2 English produced a consonant as long or short depending on its spelling, producing for instance a short [t] in *city* and a long [tː] in *kitty*, as shown by high consonant duration ratios. Native language (Italian versus English) explained 31% of the variance in the consonant duration ratio out of the total 47% explained by the model.

The magnitude of the orthographic effect was in line with previous findings, as follows. Among bilinguals, the average long-to-short consonant duration ratio was 1.33, slightly lower than the previously reported ratio of 1.38 (Bassetti et al., 2018; only results from English consonants in positions where gemination is legal in Italian are included, as all consonants in the present study were in legal positions). Among learners, the ratio was 1.53, which lies between the ratio of 1.39 reported by Bassetti et al. (2018) from a reading aloud task, and the 1.66 and 1.70 reported by Bassetti (2017) for reading aloud and word repetition, respectively. The ratio for native speakers in the current study was around 1, as in previous studies (Bassetti, 2017; Bassetti et al., 2018; Sokolović-Perović et al., 2019), confirming the lack of effect of number of letters on consonant duration in native speech production.

Naturalistic exposure seems to reduce the magnitude of orthographic effects for consonants, given that the effect was stronger in learners living in Italy than in sequential bilinguals living in the United Kingdom. As noted in the Introduction, Bassetti et al. (2018) and Mairano et al. (2018) had reported no differences in reading aloud tasks in subsamples of the two groups that took part in the present study. Different results may be due to the larger sample size in the current study. We note, however, that the bilinguals were also more proficient than the learners, so it is not possible, with the current results, to disentangle proficiency from the effect of naturalistic exposure.

Consonant Length Awareness

The next question was whether an effect of number of letters would be found in a metalinguistic awareness task. As predicted, in the C–CC rhyme judgment task the L1 Italian participants rejected English C–CC rhymes, such as *scholar* and *dollar*, because the same consonant was spelled with a single letter in one word and double letters in the other. The effect was again stronger among learners than among sequential bilinguals. English native speakers accepted these rhymes, because consonant length is not contrastive for them.

The qualitative data confirmed that L1 Italian speakers of L2 English make this phonological contrast: Their main reason for rejecting C–CC rhymes was consonant length. Of the approximately three quarters of L2 speakers who did not consider *very–merry* a rhyme, the vast majority (over 90%) reported that this was because *merry* contains a geminate. This was mostly described as a double or long consonant, and more rarely as a stronger or harsher consonant, reflecting the fact that duration is the main cue to gemination in Italian (although nondurational cues may also play a role; Payne, 2006). Length was also the main reason for rejecting the Italian *caro–carro* rhyme, showing that many Italians are applying the same phonemic contrast to English as to Italian. Furthermore, among those who correctly considered *very–merry* a rhyme, over a third spontaneously explained that in English double consonant letters do not represent geminates, showing that correctly answering this metalinguistic task is linked to an understanding that consonant length is not contrastive in English and that English and Italian grapheme–phoneme correspondences are different. The next question was then whether metalinguistic awareness would be reflected in speech production.

The Relationship Between Consonant Length Awareness and Consonant Production

Stronger consonant length awareness predicted weaker effects of orthographic forms on consonant duration (a smaller CC:C ratio) in learners' speech. This means that learners who thought that English had singleton and geminate consonants were more likely to produce long consonants in English words spelled with double consonant letters. This link between metalinguistic awareness and speech production supports the view that orthographic effects on L2 speech production are caused by orthography-induced phonological representations (Bassetti, 2006, 2008), because if orthographic effects on speech production were entirely due to the coactivation of orthography with phonology, then there should be no orthographic effects on language awareness. The next question, given the high level of interpersonal variability in orthographic effects on speech production, was whether orthographic effects may be linked to any L2 speaker characteristics.

Predictors of the Orthographic Effect on Consonant Production

The second main aim of this study was to provide a first exploration of speaker-level variables that may predict orthographic effects on L2 speech production. Results showed that English language proficiency predicts the magnitude of the orthographic effect on speech production in L2 speakers, as both a higher CEFR score in bilinguals and better performance in the Oxford Placement Test in learners predicted a weaker orthographic effect on consonant production. Although previous research had found no correlation between proficiency and the nativelikeness of segments (Iwashita et al., 2008), proficiency may in fact be more related to an understanding that English does not have long consonants than to the various complex abilities required to perceive and produce the segments of a nonnative language.

Predictors other than proficiency yielded more varied results. The regression analysis revealed a possible role of short-term memory in bilinguals. This extends previous findings that short-term memory is a strong correlate of pronunciation nativelikeness (Darcy et al., 2015), including in sound duration (Aliaga-Garcia et al., 2011), to the specific issue of orthographic effects. It appears that orthographic effect predictors differ between naturalistic and instructional settings, because short-term memory may be useful when one is surrounded by native speakers but less useful in a classroom setting with limited, often nonnativelike spoken input. The negative relationship between orthographic effect and proportion of written input out of total input among bilinguals was unexpected. Previous research had found a positive correlation between orthographic input and orthographic effects in beginner learners (Young-Scholten, 2002). Perhaps, in the context of naturalistic exposure, more exposure to writing shows better the contradiction between the double letters in spelling and the absence of consonant duration differences in speech; however, this is a counterintuitive result that warrants further research. Among learners, desire to learn pronunciation predicted lower levels of consonant gemination.

Effects of Orthographic Forms on Vowel Production and Awareness

Vowel spelling affected speech production but not metalinguistic awareness. In the delayed repetition task, the number of letters in the spelling of an English vowel affected vowel duration in Italians' L2 English speech production: The same vowel was produced with longer duration when spelled with a digraph than when spelled with a single vowel letter. The bilinguals' ratio of 1.10 was in line with previous reports (1.11 in Bassetti et al., 2018). The learners' ratio of 1.07 was smaller than in previous reports (1.14, Bassetti & Atkinson 2015; 1.12, Bassetti et al., 2018). It was also slightly smaller than the bilinguals' ratio, although this is unlikely to indicate large group differences, and previous studies found similar ratios in learners and bilinguals (Bassetti et al., 2018). The slightly stronger impact of vowel spelling on bilinguals than on learners suggests that some learners may not understand that English has different vowels, distinguished by length as well as qualitative differences.

Looking at the results for vowel length awareness in the rhyme judgment task, contrary to predictions, only a third of L2 speakers thought that *fool* and *rule* did not rhyme, and less than half of those speakers attributed this to vowel length. Furthermore, L2 speakers' accuracy in the task was not related to vowel duration ratios in speech production. The most likely explanation is that the consonant rhyme judgment task measured participants' understanding that different consonant spellings do not correspond to consonants of different length, whereas vowel rhyme judgment assessed participants' knowledge of the pronunciation of the specific words being tested. This is because English vowels can be long or short (as a correlate of the distinction between lax and tense vowels), and therefore performance in the production and awareness tasks for vowels reflects the participants' knowledge of the vowel in the specific word being tested.

Finally, looking at predictors of orthographic effects on vowel production, learners with a higher desire to learn English pronunciation (as shown by their score on the Pronunciation Attitude Inventory) had a lower tendency to produce homophonic vowels with different durations, in line with the results for consonants. It appears that among instructed learners, those with a stronger desire to learn to pronounce English are also better at learning the pronunciation of specific words. This also further confirms that predictors of orthographic effects on L2 speech production vary across naturalistic and instructed environments.

Limitations and Future Directions

One limitation of the study was that we were not able to include all target consonants and vowels in the experimental tasks (we included three consonants and four vowels in delayed repetition, and five consonants and four vowels in rhyme judgment). This was because it was necessary to control for a range of confounding variables that could have had an effect on duration, and also to include words that were imageable and accessible to learners. This meant that the range of words available as stimuli was restricted. Second, the bilingual participants had a higher level of English proficiency than the learners, so it was not possible to disentangle proficiency from the effect of naturalistic exposure in this study. The generalizability of our findings would be increased if the effects could be shown for an increased range of consonants and vowels, and also if proficiency and naturalistic exposure could be disentangled.

Finally, we included a number of speaker-level variables in investigating potential predictors of the magnitude of orthographic effects. However, the results were nonsignificant or were weak in terms of statistical significance in several cases, and counterintuitive in one case (namely, the finding for bilinguals of a negative relationship between the magnitude of the orthographic effect with consonants and proportion of written input out of total input). It would be informative in future research to use alternative measures of the speaker-level variables that were used in the present study to see whether different results are obtained. It could also be informative to investigate the role of speaker-level variables that were not included in the current study, such as measures of executive function.

Conclusion

In conclusion, our results indicated orthographic effects on L2 phonology such that Italian native speakers established a long–short contrast and used consonant and vowel length contrastively in their L2 English, not only in production as previously shown (Bassetti, 2017; Bassetti & Atkinson, 2015; Bassetti et al., 2018), but also, in the case of consonants, in awareness. We also found that proficiency was a predictor of orthographic effects on phonology (i.e., the higher the proficiency, the smaller the orthographic effect). In terms of limitations, the two groups were not directly comparable on a number of variables. For instance, age of onset of acquisition—arguably a crucial variable in L2 phonology—could not be compared due to generational differences. The orthographic effect might have been more pronounced with a larger set of stimuli, but choice of materials was limited, as noted above. Furthermore, the present analysis used various self-reported variables, which are suitable for an exploratory study but may yield different results from objective measures, at least for those variables where there is no evidence that self-ratings and objective measures are related.

This study has contributed to the field in three ways. It is likely the first research to show a direct link between the effects of the same orthographic phenomenon on speech production and on metalinguistic awareness. Second, by using larger samples than is usual in this line of research, it allowed for a meaningful exploration of speaker-level predictors of orthographic effects on speech production. Finally, it was arguably the first study to combine experimental and qualitative data to provide a fuller explanation of orthographic effects on L2 phonology.

Final revised version accepted 20 March 2020

Open Research Badges

This article has earned an Open Materials badge for making publicly available the components of the research methods needed to reproduce the reported procedure. All materials that the authors have used and have the right to share are available at https://www.iris-database.org and https://osf.io/p3q6d. All proprietary materials have been precisely identified in the manuscript.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Aliaga-Garcia, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, *47*, 1–14. https://doi.org/10.2478/psicl-2011-0002

Allan, D. (1992). *Oxford placement test*. Oxford: Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2018). MuMIn: Multi-Model Inference (R package version 1.40.4) [Computer software]. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a Foreign Language. *Written Language and Literacy*, *9*, 95–114. https://doi.org/10.1075/wll.9.1.07bas

Bassetti, B. (2007). Effects of hanyu pinyin on pronunciation in learners of Chinese as a foreign language. In A. Guder, X. Jiang, & Y. Wan (Eds.), *The cognition, learning and teaching of Chinese characters* (pp. 155–179). Beijing, China: Beijing Language and Culture University Press.

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191–206). Clevedon, UK: 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*, 1835–1842. https://doi.org/10.1037/xlm0000417

Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics*, *36*, 67–91. https://doi.org/10.1017/s0142716414000435

Bassetti, B., Escudero, P., & Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, *36*, 1–6. https://doi.org/10.1017/s0142716414000393

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*, 577–597. https://doi.org/10.1177/0023830918780141

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. Retrieved from arXiv preprint arXiv:1506.04967.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berkil, G. (2009). *A closer look at pronunciation learning strategies, L2 pronunciation proficiency and secondary variables influencing pronunciation ability*. Ankara, Turkey: Bilkent University. Retrieved from http://repository.bilkent.edu.tr/handle/11693/14676

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*, 131–151. https://doi.org/10.1017/s0025100305002148

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.19) [Computer software]. Retrieved from http://www.praat.org/

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271. https://doi.org/10.3758/bf03204507

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139087636

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*. Strasbourg, France: Council of Europe.

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

Davis, S. (2011). Geminates. In M. van Oostendrop, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology: Vol. 2. Suprasegmental and prosodic phonology* (pp. 873–897). Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0037

Derwing, B. L. (1992). Orthographic aspects of linguistic competence. In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 193–210). Amsterdam: John Benjamins.

Derwing, B. L., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*, 163–185. https://doi.org/10.1075/tsl.21.16der

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/lllt.42

Detey, S., & Nespoulous, J.-L. (2008). Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua*, *118*, 66–81. https://doi.org/10.1016/j.lingua.2007.04.003

Dewaele, J.-M. (2010). *Emotions in multiple languages*. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230289505

Ehri, L. C., & Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Applied Psycholinguistics*, 1, 371–385. https://doi.org/10.1017/s0142716400009802

Elliott, R. A. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal*, *79*, 530–542. https://doi.org/10.1111/j.1540-4781.1995.tb05456.x

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, *17*, 384–395. https://doi.org/10.1017/s1366728913000436

Esposito, 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

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, *91*, 370–389. https://doi.org/10.1121/1.402780

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125–3134. https://doi.org/10.1121/1.413041

Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (3rd ed.). Thousand Oaks, CA: Sage.

Gerds, T. A., & Ozenne, B. (2018). Publish: Format output of various routines in a suitable way for reports and publication [Computer software]. Retrieved from https://CRAN.R-project.org/package=Publish

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, *35*, 623–647. https://doi.org/10.1006/jmla.1996.0033

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*, 367–381. https://doi.org/10.1177/0023830910371460

Højsgaard, S., & Halekoh, U. (2016). doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities (R package version 4.5-15) [Computer software]. Retrieved from https://CRAN.R-project.org/package=doBy

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*, 24–49. https://doi.org/10.1093/applin/amm017

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformations or not) and towards Logit Mixed Models. *Journal of Memory and Language*, *59*, 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Klatt, D. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America*, *54*, 1102–1104. https://doi.org/10.1121/1.1914322

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of Acoustic Society of America*, *59*, 1208–1221. https://doi.org/10.1121/1.380986

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest: Tests in linear mixed effects models [Computer software]. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Larson-Hall, J. (2006). What does more time buy you? Another look at effects of long-term residence on production accuracy of English/ɹ/and/l/by Japanese speakers. *Language and Speech*, *49*, 521–548. https://doi.org/10.1177/00238309060490040401

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139166621

Lee, H.-y., & Lee, Y.-j. (2014). bear: Data analysis tool for average bioequivalence and bioavailability [Computer software]. Retrieved from https://CRAN.R-project.org/package=bear

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

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

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, *21*, 81–108. https://doi.org/10.1017/S0272263199001035

Muneaux, M., & Ziegler, J. C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, *19*, 641–660. https://doi.org/10.1080/01690960444000052

Nova Development. (2004). *Art Explosion library*. Calabasas, CA: Nova Development Co.

Payne, E. M. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association*, *35*, 153–181. https://doi.org/10.1017/s0025100305002240

Payne, E. M. (2006). Non-durational indices of gemination in Italian. *Journal of the International Phonetic Association*, *36*, 83–95. https://doi.org/10.1017/s0025100306002398

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215. https://doi.org/10.1006/jpho.2001.0134

R Core Team. (2018). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

RStudio Team. (2016). *RStudio: Integrated development for R* [Computer software]. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com

Ryder, A. G., Alden, L. E., & Paulhus, D. L. (2000). Is acculturation unidimensional or multidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of Personality and Social Psychology*, *79*, 49–65. https://doi.org/10.1037/0022-3514.79.1.49

Shively, R. L. (2008). L2 acquisition of [β], [ð̞] and [ɣ] in Spanish: Impact of experience, linguistic environment, and learner variables. *Southwest Journal of Linguistics*, *27*(2), 79.

Sokolović-Perović, M., Bassetti, B., & Dillon, S. (2019). (Online first). English orthographic forms affect L2 English speech production in native users of a non-alphabetic writing system. *Bilingualism: Language and Cognition*, https://doi.org/10.1017/s136672891900035x

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, *26*, 233–253. https://doi.org/10.1111/j.1467-1770.1976.tb00275.x

Taft, M. (2006). Orthographically influenced abstract phonological representation: Evidence from non-rhotic speakers. *Journal of Psycholinguistic Research*, *35*, 67–78. https://doi.org/10.1007/s10936-005-9004-5

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). Berlin, Germany: De Gruyter. https://doi.org/10.1515/9783110914641.1

Vokic, G. (2011). When alphabets collide: Alphabetic first-language speakers' approach to speech production in an alphabetic second language. *Second Language Research*, *27*, 391–417. https://doi.org/10.1177/0267658310396627

Wagner, R. K., Torgersen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing – Second Edition (CTOPP-2)*. PsycTESTS. https://doi.org/10.1037/t52630-000

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis* [Computer software]. New York: Springer-Verlag. Retrieved from http://ggplot2.org
Young-Scholten, M. (2002). Orthographic input in L2 phonological development. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honour of Henning Wode* (pp. 263–279). Trier: Wissenschaftlicher Verlag Trier.
Young-Scholten, M., Akita, M., & Cross, N. (1999). Focus on form in phonology: Orthographic exposure as a promoter of epenthesis. In P. Robinson & N. O. Jungheim (Eds.), *Pragmatics and pedagogy: Proceedings of the third PacSLRF* (Vol. *2*, pp. 227–233). VolTokyo: Aoyama Gakuin University.
Young-Scholten, M., & Langer, M. (2015). The role of orthographic input in second language German: Evidence from naturalistic adult learners' production. *Applied Psycholinguistics*, *36*, 93–114. https://doi.org/10.1017/s0142716414000447
Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania*, *77*, 470–481. Retrieved from http://cervantesvirtual.com/servlet/SirveObras/13540511434137617422202/p0000009.htm#i_14 https://doi.org/10.2307/344974
Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review*, *5*, 683–689. https://doi.org/10.3758/bf03208845

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Stimuli List for the Production Tasks.
Appendix S2. Stimuli List for the Awareness Tasks.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Sound Spelling Affects Both Pronunciation and Phonological Awareness in a Second Language

What This Research Was About And Why It Is Important

Does the spelling of our second (L2) language affect how we think about and pronounce the sounds of our L2? Previous research showed that Italians pronounce an English sound (consonant or vowel) as longer when it is spelled with two letters than one (e.g., they pronounce a longer “t” in “kitty” than in “city”). This is because in Italian long and short consonants distinguish words from each other, and long consonants are spelled with two letters. This study examined whether Italians actually think that English has long and short sounds.

We asked Italians to judge whether two English words rhyme, using word pairs where the same sound is spelled with one or two letters (e.g., *scholar-dollar*). We found that many Italians believed that such words do not rhyme. Crucially, those who made more errors in judging English rhymes were more likely to pronounce long sounds in words spelled with two letters, showing the link between sound categorisation and pronunciation. We also asked whether spelling affects pronunciation in some L2 speakers more than others. We found that the effects were more evident in classroom students than in immigrants living in the UK, and less evident in those with higher proficiency and in students with a strong desire to learn pronunciation.

What the Researchers Did

- Participants were 100 Italian high-school students of English, 80 Italians living in the United Kingdom, and 80 English native speakers.
- We compared the same sound (consonant or vowel) in one word where the sound is spelled with one letter and one word where it is spelled with two letters, such as the “t” in “city” and “kitty," and the “i” in “skis” and “cheese."
- To test whether participants pronounced long and short sounds (consonants and vowels) in English, they repeated aloud a series of words after hearing a native speaker’s pronunciation.
- To test whether participants thought that English has long and short sounds, they saw a series of word pairs and had to decide whether the words rhymed or not, then explain why.
- To find out who is more likely to be influenced by spelling, we collected data about factors such as language proficiency and how much participants wanted native-like pronunciation.

What the Researchers Found

- The Italians often thought that English has long and short sounds (consonants and vowels).
- The Italians often produced the same consonant or vowel as longer when it was spelled with two letters.
- Those who thought that English has long and short sounds were also more likely to pronounce long and short sounds.
- Spelling (number of letters) affected pronunciation more strongly: for consonants than vowels; and in high-school students than in immigrants living in the United Kingdom.
- Spelling affected pronunciation less strongly: in participants with higher proficiency; and in learners with a stronger desire to learn pronunciation.

Things to Consider

- We still do not know who is more likely to be influenced by spelling, for instance, those who start learning the language at an older age or those who are less exposed to spoken language.
- Language teachers and learners should be aware that spelling may affect how second language speakers categorize and pronounce the sounds of their second languages.

Materials and data: Materials are publicly available at https://www.iris-database.org and https://osf.io/p3q6d

How to cite this summary: Bassetti, B., Mairano, P., Masterson, J., & Cerni, T. (2020). Sound spelling affects both pronunciation and phonological awareness in a second language. *OASIS Summary* of Bassetti, Mairano, Masterson, & Cerni (2020) Effects of orthographic forms on L2 speech production and phonological awareness, with consideration of speaker-level predictors. In *Language Learning*. https://oasis-database.org