Does English Orthography Influence Bilingual Spanish Readers? The Effect of Grapheme Crosslinguistic Congruency and Complexity on Letter Detection

Carmen Hevia-Tuero[a], Sara Incera[b] and Paz Suárez-Coalla[a]

Author Note

[a] Laboratorio de Psicología del Lenguaje - Departamento de Psicología, Universidad de Oviedo, plaza Feijoo s/n, 33003, Oviedo, Asturias, Spain.

[b] Multilingual Laboratory - Eastern Kentucky University, 521 Lancaster Ave, Richmond, KY, 40475, USA

Carmen Hevia-Tuero https://orcid.org/0000-0003-1569-495X

Sara Incera https://orcid.org/0000-0001-9124-9204

Paz Suárez-Coalla https://orcid.org/0000-0001-9772-2680

We have no known conflict of interest to disclose. This study was supported by the Ministry of Science and Innovation of Spanish Government (predoctoral grant number FPU18/03368 and grant number PID2019-106868GB-I00). The funding sources did not have any involvement in the study.

Correspondence concerning this article should be addressed to Carmen Hevia-Tuero, Facultad de Psicología, University of Oviedo, plaza Feijoo s/n, 33003, Oviedo (Asturias), Spain. Email:

Abstract

Phonemic correspondences for a particular grapheme are not always congruent across languages. Also, some complex graphemes can be found in some languages but not in others. The purpose of this study is to determine if the congruency and complexity of English graphemes influence letter detection in L2 learners. We further investigated whether age group (7-, 9- and 11-year-old children, and university undergraduate) determines the size of these effects. Participants completed two different letter detection tasks using the mouse-tracking paradigm. Results from Experiment 1 indicate that only younger children are slightly affected by incongruent graphemes. Results from Experiment 2 show that all readers perform worse with complex graphemes. L2 learners interiorize English phonology at early stages, being barely affected by their native Spanish language. Importantly, L2 learners decode complex graphemes similarly to native English readers. Interpretations based on the BIA-d model are discussed.

Keywords: congruency, complexity, grapheme, bilingual, mouse tracking

Reading is a cognitive function which requires specific instruction to be acquired. Becoming an expert reader means being able to read both known and unknown words, which are not processed through the same strategy. In the dual-route cascaded (DRC) model (Coltheart et al., 2001) there are two different routes from print to speech: lexical, that allows known words processing through a sight-word recognition; and sublexical, which consists on applying grapheme-to-phoneme correspondences and allows reading nonwords and unknown words. While reading through the lexical route involves orthographic learning, the sublexical route also requires skilled phonological recoding. Moving from sublexical processing to lexical processing, which facilitates reading automation, depends on the implementation of a learning mechanism.

Share (1995) posited the *self-teaching hypothesis*, according to which children (as beginning learners) become expert readers by forming orthographic representations of each word through a self-teaching mechanism. Every successful decoding after learning the alphabetic code means an opportunity to acquire word-specific orthographic information. In the case of languages with a deep orthography (e.g., English), decoding implies processing the orthographic context and other elements (like syllables and rhymes). Despite potential differences between languages with deep and shallow orthographies, the self-teaching hypothesis has been tested in a variety of languages: Hebrew (Share, 1999, 2004), Dutch (de Jong & Share, 2007), English (Cunningham et al., 2002; Nation et al., 2007) and Spanish (Suárez-Coalla et al., 2016). Focusing on second languages, the hypothesis has also been evidenced in French (Chung et al., 2019) and English (Schwartz et al., 2014; van Daal & Wass, 2017). These last studies imply that many factors are involved in the orthographic learning of a second language (e.g., first and second language scripts proximity, transference of orthographic skills, instructional methods).

Research focused on second language literacy acquisition became prominent after the Common European Framework (Council of Europe, 2001), when second language (L2) learning at schools was enhanced by many European educational systems. English, in particular, gained a remarkable popularity in countries like Spain. As a result, Spanish children start their English instruction at very early ages. Furthermore, there is an increasing number of bilingual schools aimed at immersing Spanish children in English. In these environments, literacy acquisition occurs in both native and second language, thus, children face the challenge of learning to read simultaneously in two languages with dissimilar orthographic depths. The characteristics of the processing units vary across languages (Rau et al., 2015; Ziegler & Goswami, 2005), and may require different recoding strategies from those developed for the native language. Children have to learn two different orthographic codes, as well as discriminate between the two to avoid interference.

When starting second language instruction, the new language gets activated and an effort has to be made in order to inhibit the first language (Jared & Kroll, 2011). The L2 excitatory and L1 inhibitory connections are generated, and the connections gradually grow stronger with incremental exposure. This is essential, as the activation and inhibition of the appropriate languages is necessary to control the potential intrusion of the non-target language. The *language-nonselective access hypothesis* (Dijkstra & van Heuven, 2002) in bilinguals (or those who use two or more languages in their everyday life (Grosjean, 2010)) has been widely debated. Although the evidence coming from written words cannot be directly extended to other domains of bilingualism like spoken word recognition (Wang et al., 2020), authors have found empirical support for a nonselective access in visual word recognition (Dijkstra et al., 1999; van Heuven et al., 1998; Zhou et al., 2010). According to the *language nonselective access hypothesis*, when bilinguals read, lexical and sub-lexical information from both languages are subject to be activated. This coactivation produces cross-linguistic interference, the strength of which depends on variables like the specific orthography and phonology of each language. Interference can occur between different sets of languages, regardless of whether both languages have a common writing system or not (Bhide, 2015; Bialystok et al., 2005; Deacon et al., 2009; Duyck, 2005; Hamada & Koda, 2008; Howard et al., 2012; Jared et al., 2012; Jared & Szucs, 2002; Lallier & Carreiras, 2018; Lemhöfer et al., 2008; Ota et al., 2009; Sun-Alperin & Wang, 2008). In order to avoid this interference, bilinguals need an activation-inhibition mechanism described in the *bilingual interactive activation model* (BIA; Dijkstra et al., 1998) and its extensions the BIA + (Dijkstra & van Heuven, 2002) and the developmental BIA-d (Grainger et al., 2010). The BIA-d model discusses basic learning principles and addresses how these processes occur in early second language learners.

Proficient bilinguals are able to switch easily between their languages and their corresponding writing systems (Treutlein et al., 2017). However, young children starting their L2 instruction could be very sensitive to crosslinguistic interference. During early developmental stages, children are likely to transfer native language phonological rules and processing strategies, some of them at the level of spelling errors (Howard et al., 2012). These transferences might arise also at the grapheme unit level. When focusing on graphemes, it is important to determine the effects of cross-linguistic congruency and grapheme complexity. These characteristics influence native language reading so, given the additional drawback that cross-linguistic interference represents, they are likely to also impact L2 learners.

Congruency

In English (an opaque orthographic system) consonants have almost invariant letter-to-sound relationships, however vowels are the most irregular feature in the English orthography. Some of them can be pronounced in multiple ways ("a" can be pronounced /ɑː/ or /eɪ/), and some phonemes have multiple spellings (/ʊ/ can be spelled "ou" or "oo"). They have multiple correspondences, and their pronunciation is sensitive to orthographic context (Frith et al., 1998; Venezky, 1967). This is something that children have to deal with during their phonological recoding development (Share, 1995, 2008). Many researchers have reported that orthographic transparency, which varies across languages, has an effect on reading in monolinguals (Glushko, 1979; Jared, 1997; Seidenberg et al., 1984; Ziegler et al., 1997, 2003). This effect is stronger during early stages of literacy acquisition, before orthographic representations are built. It also depends on grapheme frequency as more frequent associations facilitate decoding. A processing conflict appears when a unit (e.g., "a") has multiple associations from orthography to phonology (/ɑː/, /eɪ/, /æ/...). This activation of multiple pronunciations results in longer reaction times, as evidenced through letter detection tasks (Lange, 2002).

In L2 learners multiple pronunciation activation is likely to occur even more often than in monolingual learners, as the associations might come from both L1 and L2 phonologies. Languages with the same writing system (both alphabetic) have many graphemes in common. But when graphemes are shared, they are not always congruent. The same grapheme can be associated to a different phoneme in each language. Language nonselective access induces an overlap of L1 and L2 grapheme-to-phoneme correspondences activation during reading. As a result, the association congruent with the native language will be more frequent and, therefore, stronger.

Many investigations have studied cross-linguistic phonology activation in bilinguals, using a variety of tasks like lexical decision (Duyck, 2005), reading aloud (Jared et al., 2012; Jared & Szucs, 2002; Mairano et al., 2018), spelling (Fashola et al., 1996; Howard et al., 2012) or picture-word interference task (Kaushanskaya & Marian, 2004). Commissaire and colleagues (2014) investigated specifically the effect of cross-linguistic phonological consistency in a letter detection task. They assessed it in a sample of French high school students learning English, a population for whom both languages have relatively deep orthographies. These researchers found faster reaction times in graphemes shared across languages. The results suggest that congruent correspondences connections (when a grapheme has the same phoneme in both languages) are stronger than incongruent ones (when a grapheme has a different phoneme in each language). Will congruency effects emerge in Spanish speakers learning English?

Complexity

Not all graphemes are shared between languages. Some graphemes are specific to a particular language, as it is the case of some English complex graphemes formed exclusively by vowels and associated to a single sound (e.g., "ea" like in *beach* - /biʧ/). Graphemes, the written representations of phonemes, can be simple (if constituted by a single letter) or complex (if they are composed by two or more letters). Complex graphemes have their own phoneme-to-grapheme correspondence, as two or more letters are being processed as a whole unit in order to represent one phoneme (Joubert & Lecours, 2000). The existence of these graphemes, specifically when formed by vowels, is responsible for the apparent deep orthography of languages like English (Seidenberg et al., 1994). Moreover, the need to process other bigger-than-letter units impacts monolingual literacy acquisition: speakers of deep orthography languages like English reach reading accuracy later than speakers of shallow orthography languages like Spanish (Defior & Serrano, 2005; Seymour et al., 2003). Spanish speakers learning English have an additional difficulty with complex graphemes, as there are no graphemes with equivalent characteristics in the orthographic system of their native language. In Spanish, only 5 digraphs (complex graphemes composed of two letters) can be found, none of them formed exclusively by vowels (*ch*, *rr*, *ll*, *qu* and *gu*). Biliterate children must deal with two different phonological recoding strategies depending on each language's spelling-sound relationship: either pronouncing both vowels when they are found together (Spanish diphthongs and hiatuses; Aguilar, 1999; Face & Alvord, 2004), or identifying the complex grapheme and its correspondent phoneme (English /i/ for "ea"). Are Spanish children successful at recognizing complex graphemes in their second language? Or do they recode bigrams like they would do in their native language? A way to determine whether they are able to process these English specific graphemes is through investigating the *grapheme complexity effect*.

Detecting a letter forming part of a complex grapheme (e.g., detecting the letter "a" in *bean*) takes more time than detecting the same letter embedded in a simple grapheme (e.g., "a" in *park*) (Rey et al., 1998; Rey & Schiller, 2005). When processing the word *bean*, which includes a complex grapheme, two sub-lexical processes are activated: (1) Through letter detection the reader detects four letters, and (2) through grapheme detection the reader detects three graphemes. Both processes happen at the same time producing a conflict that delays reaction time and slows down identification. Complexity effects have been reported in monolingual adults (Rey et al., 1998, 2000) and children (Marinus & de Jong, 2011) (however see Chetail, 2020 for a contrary point of view). In bilinguals, Commissaire and colleagues (2014) evaluated how specific English sub-lexical units like complex graphemes are processed by L2 learners. In their study, French speakers attending high school performed a letter detection task in English and were affected by complexity (they showed a significant complexity effect). Moreover, in line with the congruency effects discussed above, letters embedded in complex graphemes shared between languages were recognized faster. These findings support a cross-linguistic complexity effect, pointing to a benefit when processing complex graphemes that are equivalent across languages.

It is important to highlight that (to our knowledge) no one has tested the complexity effect with L2 readers of a native language that does not have complex graphemes formed by vowels (like Spanish). This particularity makes it impossible to investigate “cross-linguistic complexity effects” with Spanish/English readers. However, it opens the door to a new and intriguing question: are speakers of Spanish affected by the complexity effect when reading in English? This would mean that they are able to process English orthography (in this case, specific complex graphemes) as native speakers of English typically do. From a reading development perspective, it is important to determine how young are Spanish readers when they start to be affected by complex English graphemes. At which age do Spanish children start processing complex graphemes as native speakers of English do?

The Present Study

In two experiments we investigate the effects of congruency and complexity across readers in age groups 7-, 9-, and 11-year-old children, as well as university undergraduates. The goal is to measure English grapheme processing by L2 learners of a shallow language (Spanish) during literacy acquisition. Participants, whom we will refer to as L2 learners, were native speakers of Spanish who were learning English as a second language. The age groups were selected in order to assess developmental differences across participants. Two different experiment were carried out with the same participants, one focusing on cross-linguistic grapheme congruency and the other focusing on grapheme complexity effects. These experiments are designed to investigate how Spanish students process English sub-lexical units, while keeping in mind the differences in orthography between these two specific languages. Previous studies (Commissaire et al., 2014; Marinus & de Jong, 2011; Rey et al., 2000) measured reaction times and errors. In this study we also measured mouse trajectories by using the MouseTracker software (Freeman & Ambady, 2010). Using the mouse-tracking paradigm (Spivey et al., 2005) it is possible to obtain a more detailed measure of the ongoing cognitive processes underlying word recognition. Instead of measuring overall performance, mouse-tracking captures the ongoing decision-making processes underlying how participants respond to written words. The procedure was approved by the Ethics Committee of Research of the Principality of Asturias, and it has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

EXPERIMENT 1

In this first experiment, we explored how L2 leaners process graphemes that are shared while being congruent (equivalent phoneme) or incongruent (different phonemes) across languages. Participants had to detect letters embedded in words, in which half of them were congruent graphemes, that is, pronounced like Spanish vowels (e.g., “a” in *park*); and half of them were incongruent graphemes, that is, pronounced completely different (e.g., “a” in *name*). Incongruent graphemes were associated to multiple phonemes, one of them shared with the native language phonology but many of them different from it (*name* /neɪm/, *talk* /tɔːk/). If participants are affected by congruency, and native language phonology is strongly activated, their performance will be worse with incongruent than with congruent graphemes.

Our hypotheses for the first experiment are:

- Main Effect of Congruency: All participants will answer more efficiently to graphemes congruent with their native language ("a" in *park*) than to incongruent graphemes ("a" in *name*), because associations shared across languages will be more strongly activated.
- Main Effect of Age: Younger students will perform worse across both type of graphemes (congruent, incongruent), as their expertise in reading English is expected to be lower.
- Congruency by Age Interaction: Younger students will show a larger congruency effect (have a larger difference in performance between congruent and incongruent graphemes), as they will be more sensitive to L1 interference.

Method

Participants

Participants were Spanish-English bilinguals. The sample included children who attended a bilingual public elementary school and undergraduate students. The sample was formed by 96 participants, including 24 undergraduate students from University of Oviedo (M_{age} = 20.2, *SD* = 21 months) and 72 elementary school students. Twenty-four students were around 7 years old (M_{age} = 7.7; *SD* = 3 months), 24 were around 9 years old (M_{age} = 9.7; *SD* = 3 months) and 24 were around 11 years old (M_{age} = 11.8; *SD* = 3 months). All the 7-, 9-, and 11-year-old children attended grades second, fourth, and sixth respectively. None of the participants (undergraduate or elementary school students) had cognitive, learning or behavioral impairments.

The school and the university were located in northern Spain. Undergraduate student's participation was compensated with extra points for their classes. Undergraduate students were exposed to English since Primary school, and they continue being exposed to English between 2 and 3 hours per week (M_{time} = 2.47 hours per week; SD_{time} = 1.55) during their university studies. The public school where data collection took place was chosen by the Spanish Government in 1996 to implement a bilingual learning program based on specific guidelines. The guidelines were developed and implemented by the Spanish Ministry of Education and the British Council as a result of a formal agreement signed in 1996. Elementary school students attend four hours of Literacy lessons per week. They also have teachers who are native English speakers. This instructional method emphasizes oral communication, as children start learning English before literacy acquisition. Furthermore, systematic teaching of phonics is contemplated during infant stage for both English and Spanish. Specific guidelines can be consulted in the Spanish/English integrated curriculum (Agudo et al., 2012). The recommendation of children not to take extra English lessons out of school is given to the families. The socioeconomic status of the students who attend this school is generally middle-income, but there are isolated cases of students that come from families with either lower or higher socioeconomic status.

Materials

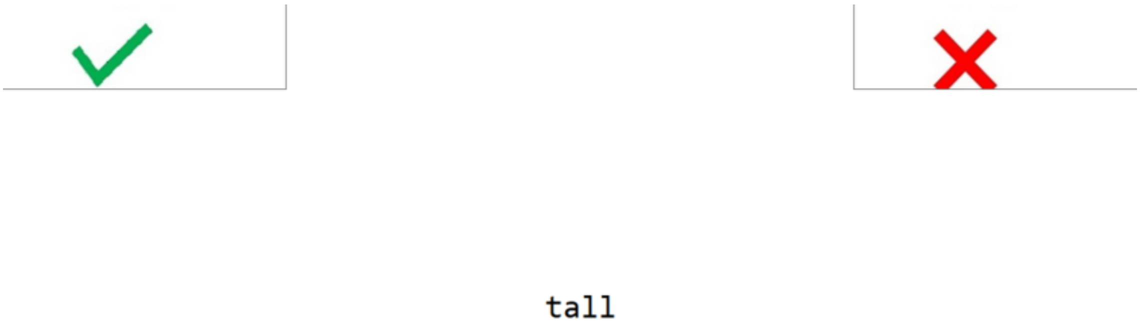
A total of 40 words were selected (see Appendix). The words contained one of the target letters (either the letter A or the letter I) and they were controlled for length, word frequency, and mean bigram frequency. The mean length was 4.00 characters (SD = 0.00) for the congruent condition, and 4.10 (SD = 0.31) for the incongruent condition, with no significant difference between the two (t = -1.45, p = .162). According to the MCWord database (Medler & Binder, 2005), the mean bigram frequency was 2,242.47 (SD = 1,049.48) for the congruent condition and 1,945.60 (SD = 1,028.93) for the incongruent condition with no significant difference between the two (t = 0.90, p = .372). The mean word frequency was 250.10 per million (SD = 185.78) for the congruent condition and 239.55 (*SD* = 184.55) for the incongruent condition with no significant difference between the two (*t* = 0.18, *p* = .858), according to the Children's Printed Word Database (Masterson et al., 2010).

Half of the selected words contained graphemes whose phonemic correspondence is shared across languages (e.g., "a" in *park*). The other half were words containing graphemes whose phonemic correspondence is different in English and Spanish (e.g., "a" in *name*). That is, English words that are read differently than how a native Spanish speaker would read them following Spanish grapheme-to-phoneme correspondences (GPC) conversion rules. In addition, 20 words in which the target letter was absent were added as fillers.

Two different versions of the experiment were created in order to counterbalance the response options. For half the participants the "present" correct response (green check mark image) was placed on the top left corner of the screen, while for the other half the "present" correct response was placed on the top right corner of the screen (see Figure 1). Each participant responded to 60 trials across three different conditions (20 words with the letter present containing a congruent grapheme, 20 words with the letter present containing an incongruent grapheme, and 20 letter-absent filler words). The order of presentation was random for all participants. Each participant (n = 96) responded to 60 trials for a total of 5760 observations.

Figure 1

Screenshot of the participants' view of the task (detecting the "a" in tall).



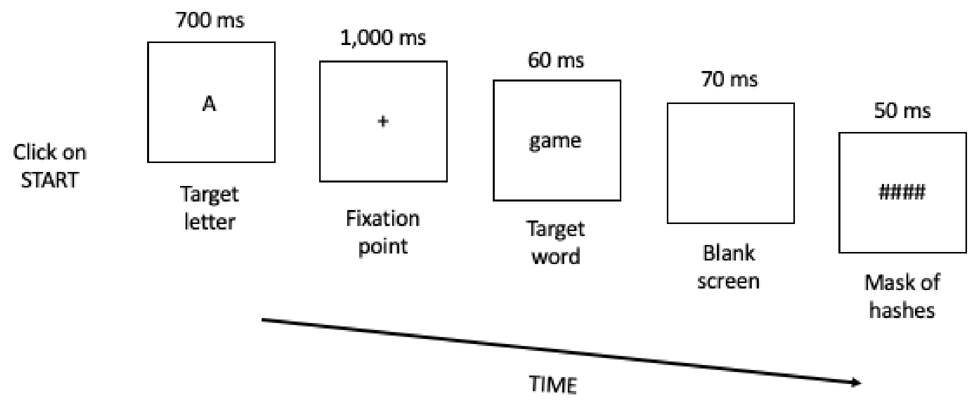
Procedure

A target letter detection task was created with the computer software MouseTracker (Freeman & Ambady, 2010). The mouse-tracking paradigm has been widely used in psycholinguistics research (Spivey et al., 2005), specifically in bilingualism (Barca & Pezzulo, 2015; Bartolotti & Marian, 2012; Incera, 2018; Incera et al., 2020; Incera & McLennan, 2016). A Medion Akoya S3409 laptop was used to present the stimuli to the participants, and participants were asked to answer using a wireless computer mouse and a large mouse pad (17.8 by 15.5 inches). Participants were tested individually (performance feedback was not provided). Testing took place in a room free of noise and distracting elements to ensure the accuracy of the results. Each participant was randomly assigned to one of the two versions of the experiment with the correct response ("green check mark") on the top right or left corners of the screen. Nonlinguistic trials (with the response options "Click Here") were included as a baseline preceding to the experiment. The purpose is to have a baseline measure of motor movement performance (independent of cognitive processes) before presenting the participants with stimuli they need to process, as well as a training phase in order to familiarize the students with the computer program (for a detailed discussion of the importance of including a baseline task when using the mouse tracking paradigm, see Incera & McLennan, 2018).

The procedure of the task followed previous works (Commissaire et al., 2014; Rey et al., 2000) except for a few modifications. The original task was designed for adults, so we increased the target word time to 66 ms in order to make the task accessible to children. Also, previous versions of the task used key press, while our version was adapted to use with the MouseTracker software (see the folder "Experiments" within the Open Science Framework). At the beginning of each trial, the START button and the response options (green tick for "yes", red cross for "no) appeared. As soon as participants clicked START, the target letter was displayed for 700 ms in uppercase in the center of the screen. After a fixation point of 1,000 ms, the target word appeared in lowercase for 60 ms. A blank screen presented for 70 ms replaced the word, and then a mask consisting of hashes appeared in the screen for 50 ms (see Figure 2). Participants started moving the mouse at word onset. They had to click "yes" if they detected the target letter in the word, or "no" if they did not. They were told to click on one of the two response options as quickly and accurately as possible. The cursor remained in the same position after the participants clicked on their response and while the START button appeared at the bottom of the screen. Participants had to move the cursor down to click on the START button that would initiate the next trial. Forcing participants to click START guaranteed that the starting position of the mouse was at the bottom center of the screen for all participants and items. If participants took more than 750 ms to initiate a mouse movement, a warning appeared instructing them to start moving the mouse earlier on in future trials. The task lasted about 10-15 minutes, depending on the age of the participant (i.e., younger children took longer than older children and undergraduate students).

Figure 2

Steps and timing of the online procedure (detecting the "a" in "game").



Analysis Plan

R (version 3.6.2) was used to run mixed model analyses using the lme4 package (version 1.1-21) (Bates et al., 2015). The independent variables included in the analyses were the between-participant variable age (7-, 9-, and 11-year-old children, undergraduates) and the within-participant variable congruency (congruent, incongruent). Trials of filler words that did not include the target letters were excluded from the analyses. The dependent variables included in the analyses were number of errors, reaction times, and mouse trajectory (the slope of the mouse position –*X*-coordinate– over time). In the case of reaction times and mouse trajectories, the clock started at the exact moment the target word appeared on the screen (see Figure 2 above). Furthermore, in the analysis of reaction times and mouse trajectories errors were removed. Outliers were excluded as well, deleting correct responses with reaction times over and under 2 standard deviations for each grade and condition. We started by including the crossed random effects of participants and items in all models. However, when the model was over fitted, we eliminated the random effect of items (even though different words were presented the task required participants to look for only two letters "a" and "i"). Models were compared using the Chi Square test; only factors that significantly contributed to model fit, as determined by a significant *p* value in the Chi square test, were included in the final model. The estimates and standard errors are reported for all factors that significantly improved model fit and are included in the final model.

The experiment, the data, and the scripts to reproduce the analyses are available at the Open Science Framework:

https://osf.io/w2buv/?view_only=66d5a48720c048bf89ee65aaa70c97cb

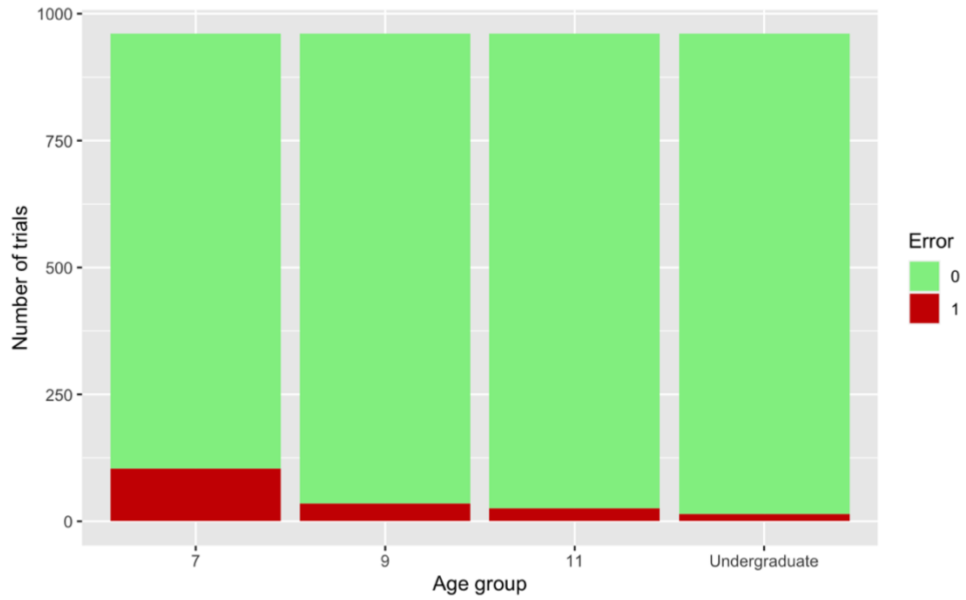
Results

Errors

When analyzing errors, model comparisons indicate that there is a main effect of Age ($\chi^2(3) = 37.12$, $p < .001$), no effect of Congruency ($\chi^2(1) = 1.82$, $p = .176$) and no Age by Congruency interaction ($\chi^2(3) = 2.72$, $p = .436$). The main effect of Age emerged because, as expected, age 7 group had more errors (104/960 - 10.8%), than age 9 group (35/960 - 3.6%; *Estimate* = -1.14, *SE* = 0.3), age 11 group (26/960, 2.7%; *Estimate* = -1.46, *SE* = 0.32) and undergraduate students (14/960, 1.4%; *Estimate* = -2.09, *SE* = 0.36). Not surprisingly, undergraduate students were the best performers (see Figure 3).

Figure 3

Number of correct (green) and incorrect (red) trials for students per age group (age 7, age 9, age 11 and undergraduate).



Reaction times

When analyzing reaction times, model comparisons indicate that there is a main effect of Age ($\chi^2(3)$ = 116.76, p < .001), no effect of Congruency ($\chi^2(1)$ = 0.43, *p* = .508), and no Age by Congruency interaction ($\chi^2(3)$ = 0.24, *p* = .970). Overall, students took about 2,000 ms to respond (*Estimate* = 2,151, *SE* = 55.5). Not surprisingly, the effect of Age emerged because age 7 group responded 598 ms (*SE* = 78.46) slower than age 9 group, 765 ms (*SE* = 78.46) slower than age 11 group, and 1,163 ms (*SE* = 78.43) slower than undergraduate students (see Figure 4 and Table 1).

Table 1

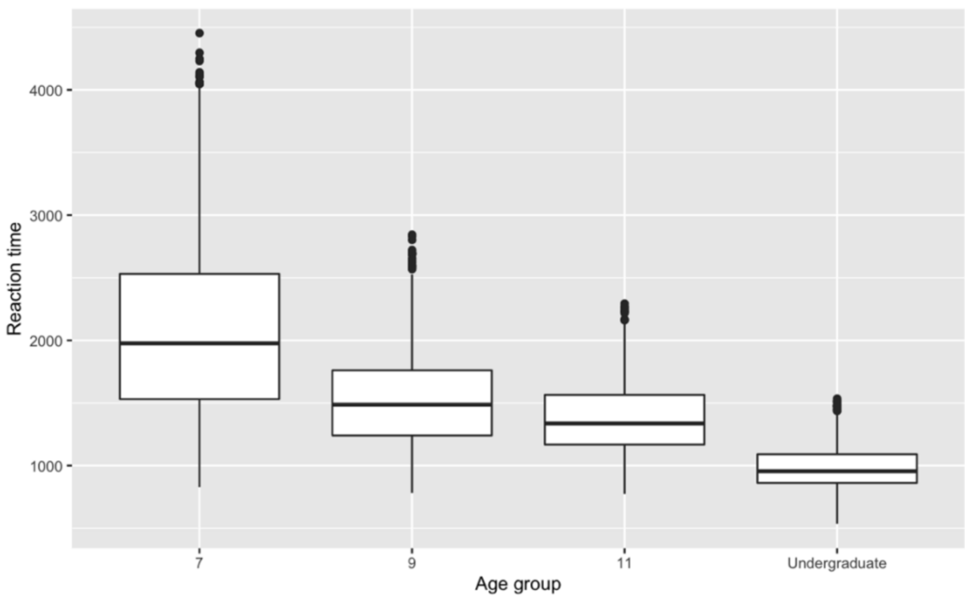
Descriptive statistics (means and standard deviations) of reaction for students per age group (age 7, age 9, age 11 and undergraduate) and condition. Differences across conditions were not significant.

Age group	Congruent	Incongruent
Age 7	2,107 (731)	2,125 (766)

Age 9	1,540 (417)	1,541 (380)
Age 11	1,373 (283)	1,390 (303)
Undergraduate	985 (175)	986 (169)

Figure 4

Reaction times for students per age group (age 7, age 9, age 11 and undergraduate).

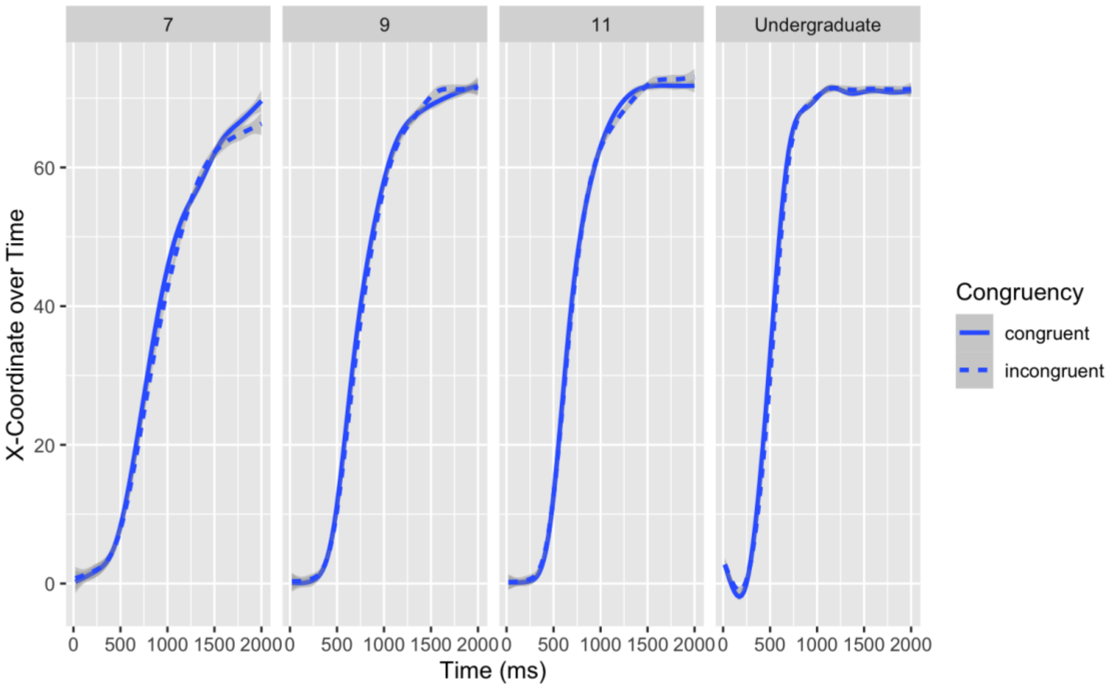


Mouse Trajectories

When analyzing the mouse position over time, model comparisons indicate that the slope of the mouse trajectory shows a main effect of Age ($\chi^2(3) = 18.49$, $p < .001$) and an Age by Congruency interaction ($\chi^2(7) = 47.68$, $p < .001$). However, the main effect of Congruency did not significantly improve model fit ($\chi^2(1) = 1.32$, $p = .250$). The significant Age by Congruency interaction that emerges is driven by the fact that the effect of Congruency (better performance for congruent than incongruent graphemes) only emerges in the age 7 group (see Figure 5). For the youngest children, the slope of the mouse trajectory is steeper (meaning that they move faster towards the correct response) when answering to congruent than incongruent graphemes (*Estimate* = - 0.39, *SE* = 0.17, *t*(350000) = -2.25, *p* = .024). In contrast, mouse trajectories went against the predicted pattern of responses in age 9 group (*Estimate* = 0.79, *SE* = 0.24, *t*(350000) = 3.21, *p* = .001), age 11 group (*Estimate* = 0.63, *SE* = 0.24, *t*(350000) = -2.58, *p* = .009) and undergraduate students (*Estimate* = 0.81, *SE* = 0.24, *t*(350000) = -3.31, *p* < .001).

Figure 5

The first two seconds of the mouse trajectories per age group (age 7, age 9, age 11 and undergraduate) responding to congruent and incongruent graphemes.



EXPERIMENT 2

In the second experiment we investigated if L2 learners were affected by complex English graphemes. In this case, the letters to be detected were embedded either in a simple grapheme formed by a single letter (e.g., "a" in *park*), or in complex graphemes specific to the English orthography (e.g., "a" in *beach*). If participants processed complex graphemes as a set (and not as two different letters), they would be sensitive to complexity effect. Therefore, they would be decoding English sub-lexical units following the English spelling rules, even though these complex graphemes do not exist in their native language (Spanish). Our hypotheses in the second experiment were:

- <u>Main effect of Complexity</u>: All participants will perform better when answering to simple than complex graphemes.
- <u>Main Effect of Age</u>: Younger children will perform worse than the rest of participants.
- <u>Complexity by Age Interaction:</u> Undergraduate students will have larger complexity effects (i.e., larger differences in performance between simple and complex graphemes). Undergraduates have been more exposed to written English and its orthography, therefore, they will be more likely to process complex graphemes as whole sets.

Method

Participants

The same children and undergraduate students that participated in Experiment 1 participated in Experiment 2. Again, participants were 7-, 9-, and 11-year-old children, while undergraduate students were recruited from the University of Oviedo. None of the participants had cognitive, learning or behavioral impairments.

Materials

48 words were selected for this task (see Appendix). They all contained one of the target letters (A, E, O), and they were controlled for word frequency, length and mean bigram frequency. The mean length was 4.29 characters (*SD* = 0.46) for simple condition, and 4.46 (*SD* = 0.51) for complex condition with no significant differences between both (t = 1.18, p = .242). The mean bigram frequency was 1,632.89 (*SD* = 929.88) for simple condition and 2083.86 (*SD* = 1,471.08) for complex condition with no differences between both (t = 1.26, p = .211) according to the MCWord database (Medler & Binder, 2005). The mean word frequency was 171.96 per million (*SD* = 115.14) for simple condition and 173.08 (*SD* = 174.14) for complex condition with no significant differences between both (t = 0.02, p = .979),

Selected words contain letters which may be part of a simple grapheme formed by a single letter (e.g., “a” in *park*), or embedded in a complex grapheme that is specific to the English orthography (e.g., “a” in *beach*). Filler words with absent target letters were also presented to the participants.

As in Experiment 1, response options were counterbalanced by creating two versions of the experiment. Each participant responded to 72 trials across three different conditions (24 letter-present words containing a simple grapheme, 24 letter-present words containing a complex grapheme, and 24 letter-absent filler words) for a total of 6912 observations.

Procedure

The procedure was the same as described in Experiment 1. The task also lasted about 10 to 15 minutes.

Analysis Plan

The analysis plan was described in Experiment 1. For Experiment 2, the independent variables included in the analyses were the between-participant variable age (7-, 9-, and 11-year-old children, undergraduates) and the within-participant variable complexity (simple, complex).

The experiment, the data, and the scripts to reproduce the analyses are available at the Open Science Framework: https://osf.io/w2buv/?view_only=66d5a48720c048bf89ee65aaa70c97cb

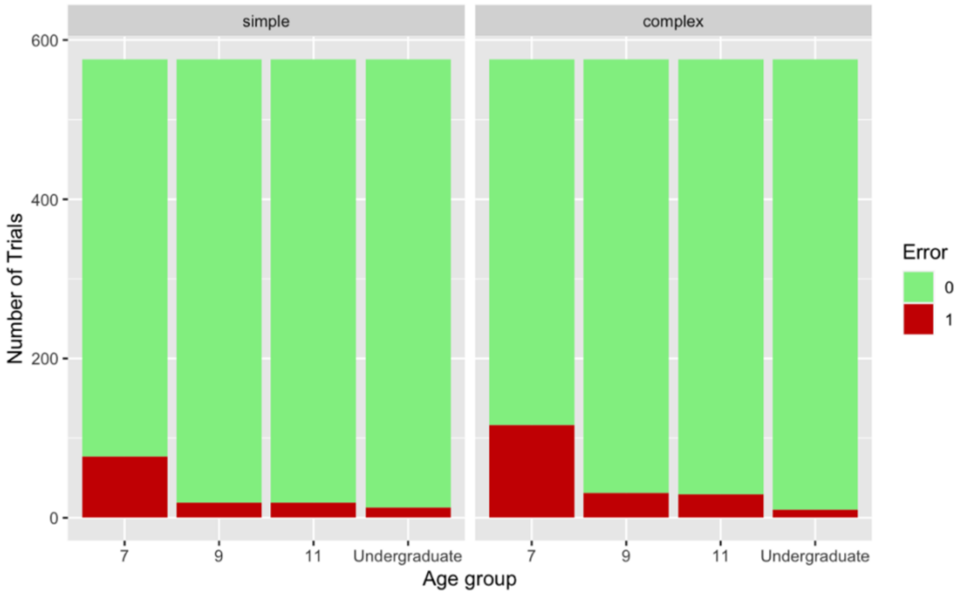
Results

Errors

When analyzing errors, model comparisons indicated that there is a main effect of Age ($\chi^2(3) = 48.7$, $p < .001$) and a main effect of Complexity ($\chi^2(1) = 6.06$, $p = .013$). However, the Age by Complexity interaction did not emerge ($\chi^2(3) = 3.33$, p = .342). The main effect of Age emerged because, not surprisingly, age 7 group had the worst performance (77/576 - 13.3% in simple graphemes; 116/576 - 20.1% in complex graphemes), followed by the age 9 group (19/576 - 3.2% in simple; 31/576 - 5.3% in complex; *Estimate* = -1.51, *SE* = 0.29), age 11 group (19/576 - 3.2% in simple; 29/576 - 5% in complex; *Estimate* = -1.57, *SE* = 0.29) and undergraduate group (13/576 - 2.2% in simple; 10/576 - 1.7% in complex; *Estimate* = -2.34, SE = 0.33) (see Figure 6). Furthermore, all the participants performed better with simple graphemes (128/2304 – 5.55%) than with complex graphemes (186/2304 – 8.07%), as we predicted in our first hypothesis (*Estimate* = 0.46, *SE* = 0.18).

Figure 6

Number of correct (green) and incorrect (red) trials for participants per age group (age 7, age 9, age 11 and undergraduate) responding to simple and complex graphemes.



Reaction times

When analyzing reaction times, model comparisons indicated that there is a main effect of Age ($\chi^2(3) = 131.86$, $p < .001$), a main effect of Complexity ($\chi^2(1) = 30.97$, $p < .001$), and an Age by Complexity interaction ($\chi^2(3) = 26.45$, $p < .001$). The main effect of Age emerged because, not surprisingly, age 7 group responded 518 ms (*SE* = 87.47) slower than age 9 group, and 831 ms (*SE* = 87.51) slower than age 11 group. Undergraduate students were the fastest, with a difference of 1,324 ms (SE = 87.43) between them and the age 7 group. The main effect of Complexity emerged because detecting complex graphemes took 219 ms (SE = 28.61) longer than detecting simple graphemes. The Age by Complexity interaction emerged because the differences between simple and complex graphemes vary across age groups (see Figure 7 and Table 2). Specifically, age 7 participants were most affected by grapheme complexity. Across all age groups, there is a significant difference between the Simple and the Complex condition (*Estimate* = 219 ms, *SE* = 28.61, $t(625.36) = 7.68$, $p < .001$). This difference is larger for age 7 than for age 9 (*Estimate* = - 168, *SE* = 38.18, $t(3968.46) = -4.41$, $p <$

.001), age 11 (*Estimate* = - 138, *SE* = 38.22, t(3966.42) = -3.62, p < .001), and undergraduates (*Estimate* = - 173, *SE* = 37.92, t(3970.29) = -4.56, p < .001). Against our original prediction, the youngest (instead of the oldest) participants are the most affected by the complexity effect (show the biggest delay when processing complex graphemes).

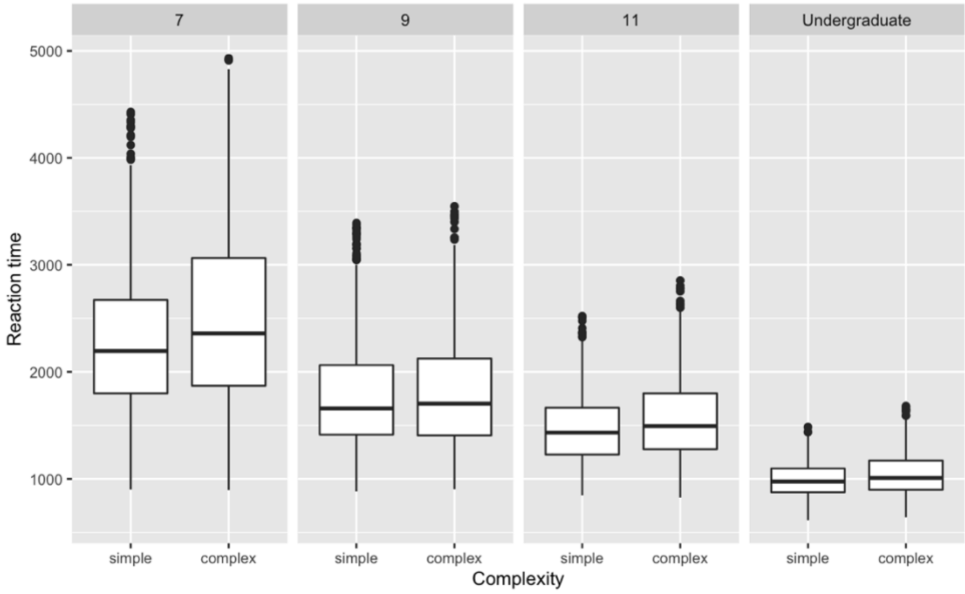
Table 2

Descriptive statistics (means and standard deviations) of reaction times for students per age group (age 7, age 9, age 11 and undergraduate) and condition.

Age group	Simple	Complex
Age 7	2,306 (687)	2,520 (846)
Age 9	1,782 (534)	1,827 (561)
Age 11	1,479 (333)	1,557 (389)
Undergraduate	998 (166)	1,043 (200)

Figure 7

Reaction times for participants per age group (age 7, age 9, age 11 and undergraduate) responding to simple and complex graphemes

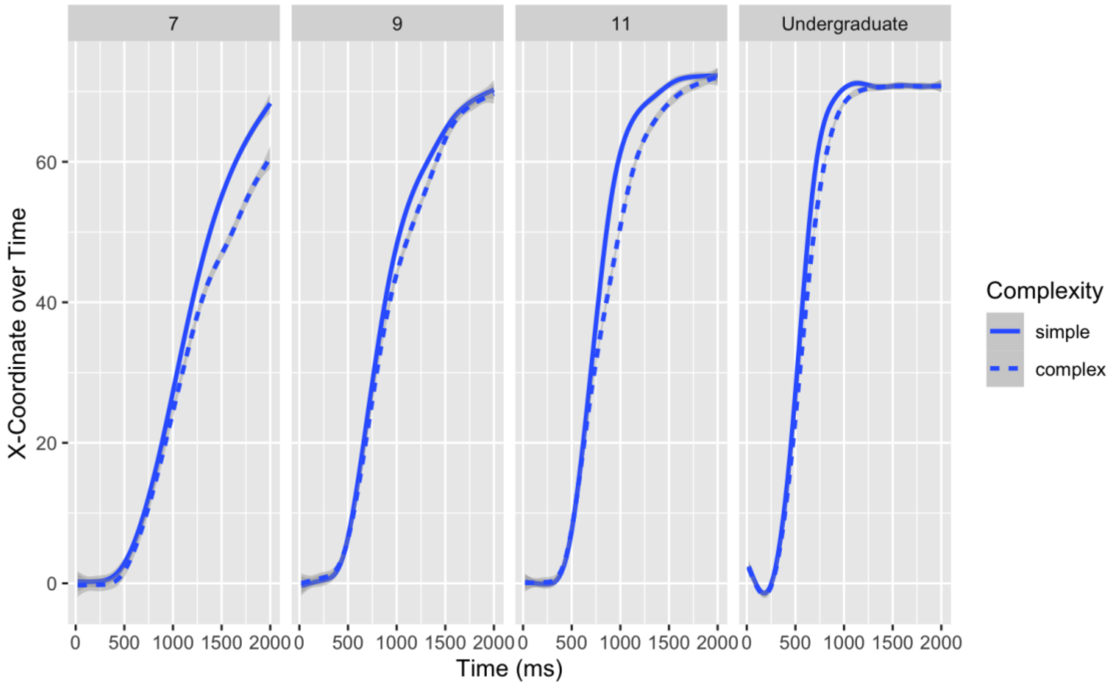


Mouse Trajectories

When considering the mouse position over times, model comparisons indicated that on the slope of the mouse trajectory there is a main effect of Age ($\chi^2(3) = 19.43$, $p < .001$), a main effect of Complexity ($\chi^2(1) = 16.11$, $p < .001$) and an Age by Complexity interaction ($\chi^2(7) = 460.51$, $p < .001$). The main effect of Age emerged because mouse trajectories were steeper (better performance) for undergraduates (*Estimate* = -0.35, *SE* = 1.12), followed by age 11 (*Estimate* = 3.71, *SE* = 1.12), age 9 (*Estimate* = 1.94, *SE* = 1.12) and age 7 (*Estimate* = 23.99, *SE* = 0.79) groups. The main effect of Complexity emerged because mouse trajectories were less steep (worse performance) with complex than simple graphemes across all the participants (*Estimate* = -2.98, *SE* = 0.18). The Age by Complexity interaction emerged because the effect of Complexity (better performance for simple graphemes) was larger for the age 7 group than the rest of the groups (see Figure 8). The differences between the simple and complex slopes of the mouse trajectories were more pronounced in age 7 group (*Estimate* = -2.98, *SE* = 0.18), than age 9 group (*Estimate* = 2.47, *SE* = 0.24), age 11 group (*Estimate* = 2.54, *SE* = 0.24) and undergraduate students (*Estimate* = 3.85, *SE* = 0.24). Across all age groups, there is a significant difference between the Simple and the Complex condition (*Estimate* = -2.98, *SE* = 0.18, *t*(409800) = -16.49, *p* < .001). This difference is larger for age 7 than for age 9 (*Estimate* = 2.48, *SE* = 0.24, *t*(409800) = 10.01, *p* < .001), age 11 (*Estimate* = 2.54, *SE* = 0.24, *t*(409800) = 10.24, *p* < .001), and undergraduates (*Estimate* = 3.85, *SE* = 0.24, *t*(409800) = 15.68, *p* < .001).

Figure 8

The first two seconds of the mouse trajectories per age group (age 7, age 9, age 11 and undergraduate) responding to simple and complex graphemes.



DISCUSSION

In Experiment 1 congruency of grapheme-to-phoneme mappings across languages was manipulated. We compared participants' performance with congruent (e.g., detect "a" in *park*) and incongruent (e.g., detect "a" in *name*) words –not predicted by Spanish orthographic rules. In this first experiment we wanted to assess if Spanish children process grapheme units differently, depending on whether the pronunciation of the graphemes is congruent across languages or not. Our hypothesis was that (in line with results from deep orthographies like English-French) graphemes that are congruent across Spanish and English would be detected faster than incongruent ones, and that younger students would be more affected by cross-linguistic interference, that is, the lack of congruency. The results of errors and reaction times were similar, in that age 7 group had the worst performance but no effect of congruency emerged. The results of the mouse trajectories point to an age by congruency interaction, as congruency effects only emerged in age 7 group. These null results were surprising to us, as we had predicted that all students would be affected by their native language when processing graphemes. The youngest children are the only ones who perform better when responding to congruent than incongruent graphemes. The congruency effect does not emerge in older children or undergraduates. In sum, grapheme congruency does not have an influence on errors nor reaction times. As it is possible to observe in Figure 5, the effect of congruency is very small and only emerges relatively late in the mouse trajectories (after 1,500 ms have passed). Thus, researchers should be cautious when making claims based on this interaction.

In Experiment 2 grapheme complexity was manipulated. We compared simple (e.g., detect “a” in *park*) and complex and English-specific graphemes (e.g., detect “a” in *beach*). Congruency and complexity have an effect on orthography processing by monolingual speakers. Our aim with the second experiment was to assess if Spanish readers, speakers of a shallow language with nonequivalent complex graphemes, would be affected by complex English graphemes. Also, we investigated how L2 learners process sublexical units that do not exist in their native language. Participants across all age groups had less errors, lower reaction times, and steeper mouse trajectories when detecting a letter in a simple grapheme than when detecting a letter in a complex grapheme. Unsurprisingly, overall performance was better in the older participants; age 7 participants had more errors, as well as higher reaction times, and less efficient mouse trajectories than the rest of participants. Interestingly, the age by condition interaction emerged in the opposite direction as predicted. Instead of observing larger complexity effects in older participants (e.g., undergraduates), the differences between simple and complex graphemes in reaction times and mouse trajectories were largest for age 7 group.

In respect of our first experiment, our results do not support our initial hypothesis, as participants were barely affected by differences in congruency. The fact that the youngest students (age 7 group) were the only ones slightly affected by cross-linguistic interference point to the idea that as Spanish-English bilinguals grow older they are able to efficiently inhibit cross-linguistic interference. The lack of congruency effects in Spanish-English bilinguals emerges as a stark contrast to the congruency effects previously reported in French-English bilinguals (Commissaire et al., 2014). It is possible that cross-linguistic interference is more pronounced in bilinguals with two deep orthographies (French-English) than in bilinguals with one shallow and one deep orthography (Spanish-English). Nevertheless, we should also remark two main differences between their participants and our own sample that might account for the contradictory results. First, the instructional method followed by the French students is described as focused on written more than on oral skills. That is different from the instructional method of the school our participants attend, as this Spanish school focuses on oral communication in second language before the students learn to read in either language. Second, French students started their English instruction later on, when reading acquisition was completed. In contrast, our participants started their literacy acquisition in both languages at the same time. Furthermore, they were already exposed to English before reading learning; thus, our students are sequential bilinguals but simultaneous biliterates. Bearing in mind the described differences, we agree with the theoretical interpretation of Commissaire and colleagues. The influence of L1 sublexical phonology affects L2 learners in letter detection when graphemes are shared but not congruent. This is, when the orthographic mappings are not congruent across languages. Nevertheless, it only occurs during their first years of instruction. Our own findings point to a nonselective activation of both languages' phonology occurring at early stages of literacy acquisition. Younger children are more sensitive to cross-linguistic interference, as they have started their instruction recently and L1 inhibitory and L2 activatory connections are not well-established. The lack of congruency effects in our older participants might be interpreted as a stronger activation of L2 phonological codes, a result of the high exposure to the second language. Therefore, in line with the BIA-d model, once the students generate these connections (better inhibition of native language and stronger activation of second language) the congruency effect is no longer detectable.

Regarding complexity, we determined whether early knowledge of English is accompanied by the interiorization of English specific graphemes. If Spanish students were to process English-specific complex graphemes during reading, it would mean that they are able to adapt their processing strategies to the target language. Indeed, results showed that all participants performed better with simple than complex graphemes. In complex graphemes, performance was worsened due to the conflict produced by the letter and the grapheme units coactivation. These results supporting complexity effects are in line with those reported by Rey and colleagues (2000) and Marinus & De Jong (2011) in monolinguals, and Commissaire and colleagues (2014) in bilinguals.

A surprising result that warrants further consideration is the fact that the complexity interaction in Experiment 2 emerged in the opposite direction as predicted. While (in line with our predictions) complexity effects emerged, the effects were larger for younger instead of older participants. If complexity effects were due solely to proficiency levels (e.g., more proficient individuals having larger complexity effects because of their automatic processing of the English orthography) we should have seen a larger effect in the older participants. Instead, we observed larger complexity effects in the youngest group of participants (7-year-old children). Chetail (2020) found no evidence for a complex grapheme effect in undergraduate students and attributed the effects observed to a phonological confusion. Since our older participants were barely affected by congruency in Experiment 1, it seems unlikely that the origin of differences between simple and complex graphemes in older participants is phonology. One possibility, though, is that our younger participants were indeed affected by phonological confusion. The origin of this confusion is that participants would detect the “a” slower in beach than in park, because they activate the phoneme which corresponds with that grapheme (/biʧ/ compared to /pɑrk/). This type of phonological confusion could result in increased differences between simple and complex graphemes for the youngest children, as they were affected by congruency in Experiment 1. Nonetheless, another possible explanation is the issue of response time scaling. It is a known phenomenon that effects tend to increase when the scale increases. In this case, reaction times for younger participants were much larger and had a greater variability than reaction times for older participants. As age 7 participants have longer reaction times, the effect of complexity might appear larger. Last, but not least, the number of letters (more vowels to process) in the complex graphemes might have affected the younger (less expert) participants in a way that older participants were not affected. Although Marinus and de Jong (2011) discarded serial processing as the reason of the complexity effect found in their Dutch participants (children with and without dyslexia), we do not reject this possibility in Spanish-English bilinguals. Given the Spanish orthographic characteristics, sublexical serial processing is a more plausible explanation of our results with younger participants. Older students could be using a global lexical reading strategy joined to a strong knowledge of English complex graphemes, which contributed to reduce differences between simple and complex graphemes. In the case of younger participants, the appliance of a serial phonological recoding in word processing could be the origin of the increase in the complexity effect. One or several of these explanations could be behind our results in Experiment 2. In order to determine the root cause of these effects, further studies about this issue would be necessary.

Regarding bilingual processing, the results of both experiments are consistent with the BIA-d model and the non-selective hypothesis. Cross-linguistic interference is more likely to happen on children when they start biliteracy acquisition (Bialystok et al., 2005; Jared et al., 2012; Sun-Alperin & Wang, 2008). In the case of Experiment 1, our younger participants could be facing the challenge of dealing with overlapping activation of L1 and L2 phonology codes for shared graphemes. After getting experience with the second language, L2 excitatory and L1 inhibitory connections strengthen, hence the lack of congruency effect in older participants. In their case, L2 grapheme-to-phoneme correspondences were strongly activated and the inhibition of L1 phonology guaranteed a correct and impartial processing of the cross-linguistic incongruent grapheme.

Regarding Experiment 2, the fact that complex graphemes are found in English but not in Spanish leads to another consideration. Complex graphemes might be processed as orthographic markers, which activate the language node described in BIA-d model (Grainger et al., 2010), thus facilitating processing. Orthographic markers are letters or bigrams that can be found only in one language. This language specific orthography facilitates language membership recognition, as it has been demonstrated by several authors (Casaponsa et al., 2014; Oganian et al., 2016; van Kesteren et al., 2012). In addition, the written introduction of the task in English indicated that English words would be appearing. This fact, joined to complex graphemes working as orthographic markers, could have made the participants activate their "English mode" (Grosjean, 1989, 2001). That is, an activation of the English connections while inhibiting the Spanish connections.

However, it seems plausible that sensitiveness to crosslinguistic interference among young children is not only a product of an immature language control mechanism, but the interaction of different reading strategies as well, as reading mechanisms are influenced by both the native and the second language (Bhide, 2015; Goswami et al., 1998; Lallier et al., 2016). Indeed, transferences during reading are highly dependent on the linguistic features shared across both languages (Lallier & Carreiras, 2018), and bilinguals can use different strategies depending on the characteristics of each language orthography. Differences in processing between shallow and deep orthography languages are found in French-German (de León Rodríguez et al., 2016) or English-Welsh (Egan et al., 2019) bilinguals. These investigations reported an adjustment in reading strategies according to the language context. However, the mentioned studies samples were formed by adults. Indeed, after years of instruction and prolonged exposure, bilinguals process each orthography in a similar way to native speakers of the language in particular (Goodwin et al., 2015; Treutlein et al., 2017). The versatility observed in our undergraduate participants suggests that English-Spanish bilinguals also adapt to English orthography strategies after their first years of instruction. But what happens during foundation literacy acquisition? The transition from a sublexical to a lexical route, described in reading developmental models (Ehri, 1992; Share, 1995), was demonstrated in monolingual readers of shallow orthographies like Italian (Orsolini et al., 2006; Zoccolotti et al., 2009) and Spanish (Cuetos & Suárez-Coalla, 2009). Bhide (2015) suggested that early literacy experience with a shallow orthography leads to a higher reliance on phonological recoding and sublexical processing in a second language. Our results from younger participants evidenced this, being consistent with other studies in French-Spanish (Lallier et al., 2014) and French-Basque (Lallier et al., 2016) bilinguals. Further studies should assess the potential advantages or disadvantages of this, but van Daal & Wass (2017) have demonstrated a positive effect of knowing a shallow orthography when learning a second deeper orthography (English).

Considering our results for both experiments, just before middle-childhood seems to be a key period in literacy acquisition. This stage of clear improvement was already evidenced in monolinguals (Seymour et al., 2003), but also in bilingual children, as Howard and colleagues (2012) found in their study about errors in spelling. A specific comparison of first stages of reading acquisition (6-10 years old) would shed light on L2 orthography processing in this crucial period. Specifically, it would confirm the degree in which children learning to read in a shallow orthography rely on the same strategies when learning a deeper orthography. It would also be interesting to compare children with different level of exposure to their second language, as it influences cross-linguistic interference (Brysbaert et al., 2017). Language exposure is likely to be involved in how bilinguals learn to process language specific features. These results will inform about current instructional methods, collecting evidence about the most indicated approach to teach bilingual children to read and write in both languages. As suggested by Murray and colleagues (Murray et al., 2019), reassessing our current instructional methods is necessary because of their importance for literacy acquisition. Furthermore, different techniques and tasks should be compared in order to compile more information about the variables that affect second language orthographic learning, as their impact on language processing depend on the task demands (Fischer-Baum et al., 2014). In this sense, our results stand for the advantages of using the mouse-tracking paradigm (Freeman & Ambady, 2010; Spivey et al., 2005) to record the unfolding of participants' responses. There were effects that emerged in the sensitive mouse trajectories despite not emerging in overall measures like errors or reaction times. Reaction times are useful and give information about how difficult to be processed an item is. Mouse trajectories give information about how this item is processed (and therefore, why would it be difficult to be processed). The mouse-tracking paradigm broadens the information that can be obtained from the same task (Marinus & de Jong, 2011; Rey et al., 2000), as it gives information about what happens during the processing and not only the outcomes.

One of the limitations of the study is that we did not collect individual socioeconomic status data from our participants. Future studies should do so, as this aspect can have a potential impact on dual language development. Regarding the generalizability of these results, our participants were selected from a particular school, therefore, our results might be only suitable to children attending bilingual schools with similar instructional methods. Also, different reading strategies applied by bilingual children should be confirmed through other tasks to determine to what extent these results emerge across cognitive processes.

In summary, this investigation provides information about how Spanish children learning English process sublexical units. Furthermore, it contributes to the discussion of how speakers of a shallow-orthography language learn to read in a deep-orthography language. The weak congruency effect, only evidenced in the mouse trajectories of the age 7 group, points to a strong activation of L2 and a strong inhibition of L1 phonology since very early on. Cross-linguistic interference might still affect more proficient children, but the language node is likely to be activated enough to avoid processing differences. Moreover, students learn to process language specific graphemes, as it was reflected by the complexity effect during letter detection. Interestingly, younger participants performance showed more differences between simple and complex graphemes, which seen from a developmental perspective could indicate a greater reliance in a serial processing strategy at early stages. The reading strategies that children apply during literacy acquisition in languages with a shallow orthography are likely to influence how they process second languages with deep orthographies. The complexity of written language processing, the differences in orthography between languages, the variety of instructional methods and the diversity of the bilingual experience make this area of research an exciting topic of investigation full of potential discoveries. Broaden knowledge about reading development (specifically focused on second language literacy acquisition) will lead to a better understanding of the challenges that children must face in bilingual education. Certainly, helping children become proficient readers across languages will have a positive impact on their lives and on society at large. As Stephen Krashen put it: "*We acquire language when we* *understand messages, when we understand what people tell us and when we understand what we read*."

Acknowledgements

We want to thank the students and staff from the school that collaborated with the project, as well as all the families that provided consent.

This study was supported by the Ministry of Science and Innovation of Spanish Government (predoctoral grant number FPU18/03368 and grant number PID2019-106868GB-I00). The funding sources did not have any involvement in the study.

REFERENCES

Agudo, M., González, R., Hill, E., Justo, M. A., Kelly, R., Locke, M., Medrano, P., & Reilly, T. (2012). *Spanish/English infants integrated curriculum* (S. G. T. S. G. de D. y Publicaciones (ed.)). Centro Nacional de Innovación e Investigación Educativa (CNIIE). https://www.mecd.gob.es/dam/jcr:e67e2e92-6ad8-44aa-be7c-ecb2e09ae59c/integrated-curriculum-infants.pdf

Aguilar, L. (1999). Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication*, *28*(1), 57–74. https://doi.org/10.1016/S0167-6393(99)00003-5

Barca, L., & Pezzulo, G. (2015). Tracking second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS ONE*, *10*(2), 1–14. https://doi.org/10.1371/journal.pone.0116193

Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, *36*(6), 1129–1147. https://doi.org/10.1111/j.1551-6709.2012.01243.x

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

Bhide, A. (2015). Early literacy experiences constrain L1 and L2 reading procedures. *Frontiers in Psychology*, *6*(1446), 1–16. https://doi.org/10.3389/fpsyg.2015.01446

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*(1), 43–61. https://doi.org/10.1207/s1532799xssr0901_4

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530–548. https://doi.org/10.1017/S1366728916000353

Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts: The impact of orthographic markedness. *Frontiers in Psychology*, *5*(MAY), 1–30. https://doi.org/10.3389/fpsyg.2014.00424

Chetail, F. (2020). Are graphemic effects real in skilled visual word recognition? *Journal of Memory and Language*, *111*(December). https://doi.org/10.1016/j.jml.2019.104085

Chung, S. C., Chen, X., Commissaire, E., Krenca, K., & Deacon, S. H. (2019). Testing the self-teaching hypothesis in second language reading. *Writing Systems Research*, *0*(June), 1–11. https://doi.org/10.1080/17586801.2019.1625478

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Commissaire, E., Duncan, L. G., & Casalis, S. (2014). Grapheme coding in L2: How do L2 learners process new graphemes? *Journal of Cognitive Psychology*, *26*(7), 725–739. https://doi.org/10.1080/20445911.2014.951363

Council of Europe. (2001). *Common European Framework of Reference for*

Languages: Learning, teaching, assessment. Cambridge University Press.

Cuetos, F., & Suárez-Coalla, P. (2009). From grapheme to word in reading acquisition in Spanish. *Applied Psycholinguistics*, *30*(4), 583–601. https://doi.org/10.1017/S0142716409990038

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, *82,* 185–199. https://doi.org/10.1016/S0022-0965(02)00008-5

de Jong, P. F., & Share, D. L. (2007). Orthographic learning during oral and silent reading. *Scientific Studies of Reading*, *11*(1), 55–71. https://doi.org/10.1080/10888430709336634

de León Rodríguez, D., Buetler, K. A., Eggenberger, N., Preisig, B. C., Schumacher, R., Laganaro, M., Nyffeler, T., Annoni, J.-M., & Müri, R. M. (2016). The modulation of reading strategies by language opacity in early bilinguals: An eye movement study. *Bilingualism: Language and Cognition*, *19*(3), 567–577.

Deacon, S. H., Wade-Woolley, L., & Kirby, J. R. (2009). Flexibility in young second-language learners: Examining the language specificity of orthographic processing. *Journal of Research in Reading*, *32*(2), 215–229. https://doi.org/10.1111/j.1467-9817.2009.01392.x

Defior, S., & Serrano, F. (2005). The initial development of spelling in Spanish: From global to analytical. *Reading and Writing*, *18*(1), 81–98. https://doi.org/10.1007/s11145-004-5893-1

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496–518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(03), 175–197. https://doi.org/10.1017/s1366728902003012

Dijkstra, T., van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, *38*(3–4), 177–196.

Duyck, W. (2005). Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1340–1359. https://doi.org/10.1037/0278-7393.31.6.1340

Egan, C., Oppenheim, G. M., Saville, C., Moll, K., & Jones, M. W. (2019). Bilinguals apply language-specific grain sizes during sentence reading. *Cognition*, *193*(July 2018), 104018. https://doi.org/10.1016/j.cognition.2019.104018

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In *Reading acquisition.* (pp. 107–143). Lawrence Erlbaum Associates, Inc.

Face, T. L., & Alvord, S. M. (2004). Lexical and acoustic factors in the perception of the Spanish diphthong vs. Hiatus contrast. *Hispania*, *87*(3), 553–564. http://www.elsevier.com/locate/scp

Fashola, O. S., Drum, P. A., Mayer, R. E., & Kang, S.-J. (1996). A cognitive theory of orthographic transitioning: Predictable errors in how spanish-speaking children spell english words. *American Educational Research Journal*, *33*(4), 825–843. https://doi.org/10.3102/00028312033004825

Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: Evidence from ERPs. *Language,* *cognition and neuroscience*, *29*(10), 1342–1355. https://doi.org/10.1080/23273798.2014.927067.Frequency

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German and English-speaking children. *Scientific Studies of Reading*, *2*(1), 31–54. https://doi.org/10.1207/s1532799xssr0201

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674–691. https://doi.org/10.1037/0096-1523.5.4.674

Goodwin, A. P., August, D., & Calderon, M. (2015). Reading in multiple orthographies: Differences and similarities in reading in Spanish and English for English learners. *Language Learning*, *65*(3), 596–630. https://doi.org/10.1111/lang.12127

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, *19*(1), 19–52. https://doi.org/10.1017/s0142716400010560

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmannn (Eds.), *Language acquisition across linguistic and cognitive systems (vol. 52)* (Issue October 2014, pp. 267–283). John Benjamins Publishing. https://doi.org/10.1075/lald.52.18gra

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*, 3–15. https://doi.org/10.1016/0093-

934X(89)90048-5

Grosjean, F. (2001). The Bilingual's language modes. In J. L. Nicol (Ed.), *Explaining linguistics. One mind, two languages: Bilingual language processing* (pp. 1–22). Blackwell Publishing.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Harvard University Press.

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning*, *58*(1), 1–31. https://doi.org/10.1111/j.1467-9922.2007.00433.x

Howard, E. R., Green, J. D., & Arteagoitia, I. (2012). Can yu rid guat ay rot? A developmental investigation of cross- linguistic spelling errors among Spanish-English bilingual students. *Bilingual research journal*, *35*(2), 164–178. https://doi.org/10.1080/15235882.2012.703637.Can

Incera, S. (2018). Measuring the timing of the bilingual advantage. *Frontiers in Psychology*, *9*, 1983. https://doi.org/10.3389/fpsyg.2018.01983

Incera, S., & McLennan, C. T. (2016). Mouse tracking reveals that bilinguals behave like experts. *Bilingualism: Language and Cognition*, *19*(3), 610–620. https://doi.org/10.1017/S1366728915000218

Incera, S., & McLennan, C. T. (2018). Bilingualism and age are continuous variables that influence executive function. *Aging, Neuropsychology, and Cognition*, *25*(3), 443–463. https://doi.org/10.1080/13825585.2017.1319902

Incera, S., Tuft, S. E., Fernandes, R. B., & McLennan, C. T. (2020). Using mouse tracking to investigate auditory taboo effects in first and second language speakers of American English. *Cognition and Emotion*, *0*(0), 1–9. https://doi.org/10.1080/02699931.2020.1741339

Jared, D. (1997). Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*(4), 505–529.

Jared, D., Cormier, P., Levy, B. A., & Wade-Woolley, L. (2012). Cross-language activation of phonology in young bilingual readers. *Reading and Writing*, *25*(6), 1327–1343. https://doi.org/10.1007/s11145-011-9320-0

Jared, D., & Kroll, J. F. (2011). Cognitive processes in bilingual reading. *Dyslexia across Languages: Orthography and the Brain-Gene-Behavior Link.*, *July*, 262–280. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc8&NEWS=N&AN=2011-27130-015

Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition*, *5*(3), 225–239. https://doi.org/10.1017/s1366728902003024

Joubert, S. A., & Lecours, A. R. (2000). The role of sublexical graphemic processing in reading. *Brain and Language*, *72*(1), 1–13.

Kaushanskaya, M., & Marian, V. (2004). Activation of non-target language phonology during bilingual visual word recognition: Evidence from eye-tracking. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 654–659. papers2://publication/uuid/92FC1313-E458-4F4A-BC2D-A5B777B1C853

Lallier, M., Acha, J., & Carreiras, M. (2016). Cross-linguistic interactions influence reading development in bilinguals: A comparison between early balanced French-Basque and Spanish-Basque bilingual children. *Developmental Science*, *19*(1), 76–89. https://doi.org/10.1111/desc.12290

Lallier, M., & Carreiras, M. (2018). Cross-linguistic transfer in bilinguals reading in two alphabetic orthographies: The grain size accommodation hypothesis. *Psychonomic Bulletin and Review*, *25*(1), 386–401.

https://doi.org/10.3758/s13423-017-1273-0

Lallier, M., Valdois, S., Lassus-Sangosse, D., Prado, C., & Kandel, S. (2014). Impact of orthographic transparency on typical and atypical reading development: Evidence in French-Spanish bilingual children. *Research in Developmental Disabilities*, *35*(5), 1177–1190.

Lange, M. (2002). Activation of multiple phoneme associates of graphemes in visual word recognition. *Brain and Language*, *81*(1–3), 610–620. https://doi.org/10.1006/brln.2001.2551

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Mairano, P., Bassetti, B., Cerni, T., & Sokolović-Perović, M. (2018). Effects of L1 orthography and L1 phonology on L2 English pronunciation. *Revue Française de Linguistique Appliquée*, *23*(1), 45–57.

Marinus, E., & de Jong, P. F. (2011). Dyslexic and typical-reading children use vowel digraphs as perceptual units in reading. *Quarterly Journal of Experimental Psychology*, *64*(3), 504–516. https://doi.org/10.1080/17470218.2010.509803

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology*, *101*(2), 221–242. https://doi.org/10.1348/000712608X371744

Medler, D. A., & Binder, J. R. (2005). *MCWord: An On-Line Orthographic Database of the English Language*. http://www.neuro.mcw.edu/mcword/

Murray, B. A., McIlwain, M. J., Wang, C. hsuan, Murray, G., & Finley, S. (2019). How do beginners learn to read irregular words as sight words? *Journal of Research in Reading*, *42*(1), 123–136. https://doi.org/10.1111/1467-9817.12250

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*, 71–84. https://doi.org/10.1016/j.jecp.2006.06.004

Oganian, Y., Conrad, M., Aryani, A., Heekeren, H. R., & Spalek, K. (2016). Interplay of bigram frequency and orthographic neighborhood statistics in language membership decision. *Bilingualism*, *19*(3), 578–596. https://doi.org/10.1017/S1366728915000292

Orsolini, M., Fanari, R., Tosi, V., De Nigris, B., & Carrieri, R. (2006). From phonological recoding to lexical reading: A longitudinal study on reading development in Italian. *Language and Cognitive Processes*, *21*(5), 576–607. https://doi.org/10.1080/01690960500139355

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, *111*(2), 263–269. https://doi.org/10.1016/j.cognition.2008.12.007

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, *130*, 92–105. https://doi.org/10.1016/j.jecp.2014.09.012

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, *68*, 71–80. https://doi.org/10.1016/S0010-0277(98)00051-1

Rey, A., & Schiller, N. O. (2005). Graphemic complexity and multiple print-to-sound associations in visual word recognition. *Memory and Cognition*, *33*(1), 76–85. https://doi.org/10.3758/BF03195298

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, *75*(1), 1–12. https://doi.org/10.1016/S0010-0277(99)00078-5

Schwartz, M., Kahn-Horwitz, J., & Share, D. L. (2014). Orthographic learning and self-teaching in a bilingual and biliterate context. *Journal of Experimental Child Psychology*, *117*(1), 45–58. https://doi.org/10.1016/j.jecp.2013.08.008

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1177–1196. https://doi.org/10.1037/0096-1523.20.6.1177

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*(3), 383–404. https://doi.org/10.1016/S0022-5371(84)90270-6

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*(4), 267–298. https://doi.org/10.1016/j.jecp.2004.01.001

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Suárez-Coalla, P., Álvarez-Cañizo, M., & Cuetos, F. (2016). Orthographic learning in Spanish children. *Journal of Research in Reading*, *39*(3), 292–311. https://doi.org/10.1111/1467-9817.12043

Sun-Alperin, M. K., & Wang, M. (2008). Spanish-speaking children's spelling errors with English vowel sounds that are represented by different graphemes in English and Spanish words. *Contemporary Educational Psychology*, *33*, 932–948. https://doi.org/10.1016/j.cedpsych.2007.12.005

Treutlein, A., Schöler, H., & Landerl, K. (2017). Recoding strategies of German learners of English as a foreign language. *Reading and Writing*, *30*(6), 1215–1230. https://doi.org/10.1007/s11145-016-9719-8

van Daal, V. H. P., & Wass, M. (2017). First- and second-language learnability explained by orthographic depth and orthographic learning: A "natural" Scandinavian experiment. *Scientific Studies of Reading*, *21*(1), 46–59. https://doi.org/10.1080/10888438.2016.1251437

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*(3), 458–483. https://doi.org/10.1006/jmla.1998.2584

van Kesteren, R., Dijkstra, T., & de Smedt, K. (2012). Markedness effects in Norwegian-English bilinguals: Task-dependent use of language-specific letters and bigrams. *Quarterly Journal of Experimental Psychology*, *65*(11), 2129–2154. https://doi.org/10.1080/17470218.2012.679946

Venezky, R. L. (1967). English orthography: Its graphical structure and its relation to sound. *Reading Research Quarterly*, *2*(3), 75–105. https://doi.org/10.2307/747031

Wang, X., Hui, B., & Chen, S. (2020). Language selective or non-selective in bilingual lexical access? It depends on lexical tones! *PLoS ONE*, *15*(3), 1–23. https://doi.org/10.1371/journal.pone.0230412

Zhou, H., Chen, B., Yang, M., & Dunlap, S. (2010). Language nonselective access to phonological representations: Evidence from Chinese-English bilinguals. *Quarterly Journal of Experimental Psychology*, *63*(10), 2051–2066. https://doi.org/10.1080/17470211003718705

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language*, *37*(4), 533–554. https://doi.org/10.1006/jmla.1997.2525

Ziegler, J. C., Perry, C., & Coltheart, M. (2003). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychonomic Bulletin and Review*, *10*(4), 947–953. https://doi.org/10.3758/BF03196556

Zoccolotti, P., de Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing*, *22*(9), 1053–1079. https://doi.org/10.1007/s11145-008-9144-8

Appendix: List of stimuli for Experiment 1 and Experiment 2

Table A

Graphemes by congruent and incongruent condition used in Experiment 1

Congruent		Incongruent	
Target letter "a"	**Target letter "i"**	**Target letter "a"**	**Target letter "i"**
Card	Pink	Game	Five
Farm	Give	Name	Rice
Half	Hill	Late	Nine
Hard	Kiss	Wake	Bike
Past	Milk	Cake	Bird
Path	Sick	Tall	Dirt
Bark	Wish	Salt	Girl
Dark	Gift	Talk	Wife
Fast	Kill	Walk	First
Bath	Miss	Wall	Birth

Table B

Graphemes by simple and complex condition used in Experiment 2

Simple			Complex		
Target letter "a"	**Target letter "e"**	**Target letter "o"**	**Target letter "a"**	**Target letter "e"**	**Target letter "o"**
Bath	Best	Corn	Bread	Beach	Board
Black	Desk	Frog	Clean	Head	Boat
Glad	Dress	Shop	Coat	Learn	Coal
Lamp	Guess	Soft	Dead	Meal	Coast
Plant	Left	Song	Dream	Leaf	Float
Sand	Neck	Storm	Read	Speak	Goat
Star	Send	Torch	Toad	Teach	Load
Yard	Spell	Word	Toast	Team	Road