



**HAL**  
open science

# Contribution of orthography to vocabulary acquisition in a second language: Evidence of an early word-learning advantage in elementary-school children

Florian Salomé, Eva Commissaire, Séverine Casalis

## ► To cite this version:

Florian Salomé, Eva Commissaire, Séverine Casalis. Contribution of orthography to vocabulary acquisition in a second language: Evidence of an early word-learning advantage in elementary-school children. *Journal of Experimental Child Psychology*, 2024, 246, pp.105978. 10.1016/j.jecp.2024.105978 . hal-04627090

**HAL Id: hal-04627090**

**<https://hal.science/hal-04627090>**

Submitted on 27 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Contents lists available at [ScienceDirect](#)

# Journal of Experimental Child Psychology

journal homepage: [www.elsevier.com/locate/jecp](http://www.elsevier.com/locate/jecp)



## Contribution of orthography to vocabulary acquisition in a second language: Evidence of an early word-learning advantage in elementary-school children



Florian Salomé<sup>a</sup>, Eva Commissaire<sup>b</sup>, Séverine Casalis<sup>a,\*</sup>

<sup>a</sup> Univ Lille, CNRS, UMR 9193, Sciences Cognitives et Sciences Affectives (SCALab), 59000 Lille, France

<sup>b</sup> Université de Strasbourg, Laboratoire de Psychologie Cognitive (LPC), 67000 Strasbourg, France

### ARTICLE INFO

#### Article history:

Received 17 November 2023

Revised 29 April 2024

Accepted 30 April 2024

#### Keywords:

Word learning

Second language

Orthographic facilitation

Paired-associate word-learning paradigm

Elementary school children

### ABSTRACT

Recent studies have shown that children benefit from orthography when learning new words. This orthographic facilitation can be explained by the fact that written language acts as an anchor device due to the transient nature of spoken language. There is also a close and reciprocal relationship between spoken and written language. Second-language word learning poses specific challenges in terms of orthography–phonology mappings that do not fully overlap with first-language mappings. The current study aimed to investigate whether orthographic information facilitates second-language word learning in developing readers, namely third and fifth graders. In a first experiment French children learned 16 German words, and in a second experiment they learned 24 German words. Word learning was assessed by picture designation, spoken word recognition, and orthographic choice. In both experiments, orthographic facilitation was found in both less and more advanced readers. The theoretical and practical implications of these findings are discussed.

© 2024 The Author(s). Published by Elsevier Inc.

\* Corresponding author.

E-mail address: [severine.casalis@univ-lille.fr](mailto:severine.casalis@univ-lille.fr) (S. Casalis).

## Introduction

Building vocabulary knowledge is fundamental in the early steps of learning a second language (L2) because it allows language understanding and production. Importantly, the early steps of foreign language learning start at school in many countries, several years after learning one's mother tongue. Exposure to the new language is rather limited; for example, in France pupils are exposed to 1 to 4 hr at most per week (Ministère de l'Éducation Nationale, 2023). In the vast majority of cases, English is taught. Otherwise, German or Spanish is taught, or regional languages such as Breton are taught. Although teaching focuses on spoken language, written language is introduced as soon as Grade 3 (Ministère de l'Éducation Nationale, 2019), and so orthography is increasingly involved in vocabulary learning. Therefore, phonological and orthographic representations of novel L2 words need to be learned and connected to semantic information. This raises the question of the extent to which the presence of orthography contributes to L2 word learning, in children who are still developing readers. This was the objective of the current work, for which we drew on recent literature in the field of vocabulary acquisition in the first language (L1).

### *Orthographic facilitation in an L1 in children*

Several studies have emphasized the contribution of orthography to L1 vocabulary acquisition, especially in children with typical development (Chambré et al., 2017, 2020; Ehri & Wilce, 1979; Jubenville et al., 2014; Li et al., 2016; Ricketts et al., 2009, 2021; Rosenthal & Ehri, 2008; Salins et al., 2023; Valentini et al., 2018; for a review, see Colenbrander et al., 2019). Whether examining isolated pseudoword learning (Jubenville et al., 2014; Li et al., 2016; Ricketts et al., 2009; Rosenthal & Ehri, 2008) or learning of words embedded within stories (Chambré et al., 2017, 2020; Ehri & Wilce, 1979; Valentini et al., 2018), an overall consistent benefit of the presence of orthography was evidenced, referred to as “orthographic facilitation.” These effects have been reported in several tasks focusing on various aspects of lexical processing. Frequently, children were asked to pronounce a novel word corresponding to a picture, that is, a production task assessing the semantic-to-phonological pathway as well as retrieval of the phonological form (Ricketts et al., 2009). Phonological processing without semantic processing has been assessed by spoken word recognition tasks, for example, by asking participants to choose between two phonological forms (Valentini et al., 2018). Orthographic processing has been assessed by production (spelling-to-dictation task; e.g., Jubenville et al., 2014) or recognition tasks (orthographic choice task; Valentini et al., 2018). Overall, orthographic facilitation was reported in these tasks as long as ceiling effects were avoided, for example, owing to the low number of items to be learned.

Orthographic facilitation may first result from the difference in properties between spoken and written modalities. Because phonology is more transient in time than orthography, it may explain why the presence of written information helps children to acquire novel phonological forms and to store them as strong phonological representations in their lexicon. According to this hypothesis, orthography acts as an anchoring device, promoting the memorization of novel spoken forms, in line with the orthographic mapping theory (Ehri, 2014, 2020). In addition to this visual advantage associated with multimodality, the tight and reciprocal relationships between spoken and written language may also account for the orthographic facilitation observed in L1 vocabulary learning (Ricketts et al., 2021; Valentini et al., 2018).

From the early stages of literacy acquisition, children automatically decode printed words (Share, 1995), thereby linking written language to the early and predominant oral modality. However, they also predict automatically and with increasing accuracy the written form of a new spoken word form (see the orthographic skeleton hypothesis by Wegener et al., 2018). These automatic bidirectional relationships may explain the benefit of a multimodal written/spoken learning condition by facilitating the mapping of print-to-sound correspondences of new word forms and the development of their corresponding phonological and orthographic representations. Perfetti and Hart (2002) proposed the *lexical quality hypothesis*, which states that the good quality of lexical representations depends on

the precision of phonological, orthographic, and semantic representations together with strong interconnections between them. More precisely, this means that “a word has a good quality representation if orthography efficiently activates its pronunciation and/or meaning” (Ricketts et al., 2009). Therefore, connections between orthography and phonology are a key aspect for establishing solid lexical representations.

Languages with transparent orthographies may be more prone to show orthographic facilitation than those with more opaque orthographies. In the former, decoding output and spoken forms closely match, which is not the case when orthography-to-phonology mappings are inconsistent. Jubenville et al. (2014) showed a larger benefit of orthography in French-speaking children who were exclusively exposed to consistent prints during learning compared with those who were not (but see Ricketts et al., 2009, for a lack of consistency effect on orthographic facilitation in English-speaking children). Reading expertise may also modulate the extent to which individuals benefit from orthography during novel word learning. Previous studies tested the influence of reading skills on orthographic facilitation among same-age children, but results were not consistent. Whereas Rosenthal and Ehri (2008) found that skilled readers benefit more from orthography when learning new words, this difference was not observed by Ricketts et al. (2009).

A recent study by Salins et al. (2023), conducted in Australia with English-speaking children examined the impact of age on orthographic facilitation. Three groups of children were considered: Group A (Grades 1 and 2), Group B (Grades 3 and 4), and Group C (Grades 5 and 6). The authors expected younger children (Group A) to show smaller benefits than older children (Groups B and C) due to their lower reading skills. The oldest children (Group C) were also expected to show less orthographic facilitation because of their increased ability to generate a mental image of spelling (i.e., an orthographic expectation; see Wegener et al., 2018). Thus, children from Group B (Grades 3 and 4) were expected to display the strongest orthographic facilitation effect. The authors examined orthographic facilitation through two tasks: a picture naming task and a picture–word matching task (looking at the correct picture among four pictures after hearing a word). Although there was no impact of grade level on the naming task, some effects appeared in the picture–word matching task. Here, younger children (Grades 1 and 2 and Grades 3 and 4) were faster in the orthographic condition, a pattern that was not observed in the older children (Grades 5 and 6). The authors concluded that all groups showed orthographic facilitation in learning spoken words and that the youngest groups benefitted most from the presence of spellings.

### *Orthographic facilitation in an L2 learning context*

The benefits of orthography consistently reported in L1 word learning cannot fully account for the challenges associated with the acquisition of vocabulary in an L2. Unlike pseudowords (or low-frequency words), learning a word in an L2 requires children to acquire L2 spoken and written forms and to store them in their lexicon through tight connections with preexisting semantic representations (Comesaña et al., 2009). These new representations also become progressively connected to formal L1 representations, leading to possible interactions at the lexical level (Grainger et al., 2010; Poarch et al., 2015). Moreover, these L2 word forms may have specific formal properties. Spoken forms are generally specific to the L2 owing to their phonetic features (see Marian et al., 2012). When it comes to written words, the situation can be completely different depending on whether or not the script is shared between languages. Here, we restricted ourselves to the case of sharing the same alphabet (the Latin alphabet). In this case, although some orthographic patterns are also specific or even exclusive to the L2 (e.g., “ght” occurs in English but not in French), most letters and graphemes are usually shared across Latin alphabetic languages. Critically, this leads to specific challenges in terms of orthography-to-phonology mappings in L2 words, which do not fully overlap with L1 orthography-to-phonology mappings (Botezatu, 2023; see Hayes-Harb & Barrios, 2021, for a review). Therefore, in addition to a problem of within-language print-to-sound consistency, which could be

alleviated when learning a language with transparent orthography-to-phonology mappings, L2 word learning may also confront children with a problem of interlinguistic *incongruency*<sup>1</sup> in orthography-to-phonology mappings. We use the term *congruency* instead of *inconsistency* to prevent any confusion with intralinguistic consistency. Owing to the tight connections between spoken and written language, relying on the preexisting orthography-to-phonology mappings that are specific to an L1 during the early steps of L2 learning results in initial mispronunciation of L2 forms. Therefore, the question is to know whether the benefit of orthography reported in L1 word-learning studies could be extended to an L2. If so, it suggests that children can rapidly overcome these initial incongruencies and adjust their preexisting orthography-to-phonological mappings to L2 ones.

So far, only two studies have explored whether the benefit of orthography reported in an L1 could be extended to L2 word learning in children (Hu, 2008; Krepel et al., 2021). Hu (2008) investigated the contribution of written information on pseudoword learning in Chinese fifth-graders learning English as an L2. Despite encouraging results showing an orthographic facilitation effect in a picture-naming task, the very low number of items used by the authors (three items per learning condition) and the heterogeneity of the sample in terms of experience with English as an L2 ruled out further generalizability. Recently, Krepel et al. (2021) had Dutch sixth-graders learn 12 English words (half of which were incongruent ones). Following an initial fixation cross, the L1 (Dutch) written form was displayed before or after the L2 (English) spoken and written forms for half the learning list. For the remaining half, the L2 written form was not presented during learning. The authors reported a learning benefit of L2 phonological forms when L2 written forms were displayed during learning. Critically, the L2 in this study was English, a language with which the children were already familiar.

#### *Studies conducted in the adult population learning an L2*

Besides these two studies conducted in children (Grades 5 and 6), several studies have examined L2 word learning in adults in context (e.g., Elgort et al., 2015) or in isolation (e.g., Bartolotti & Marian, 2017; Bürki, Welby, Clément, & Spinelli, 2019; Escudero et al., 2023). Yet, most of the studies conducted in adults were actually designed to examine a specific aspect of spoken word learning, that is, whether phonetic processing is promoted by the exposure to the written form (see Hayes-Harb & Barrios, 2021, for a review). Interestingly, most of these studies showed a deleterious impact of orthography on learning L2 novel phonological contrasts, that is, which are not preexistent in the native language such as Italian consonant gemination (see Bassetti, 2017; Bassetti et al., 2018; Cerni et al., 2019), silent final consonants (Hayes-Harb et al., 2010), and minimal phonetic pairs (Escudero, 2015; Escudero et al., 2008, 2014; Escudero & Wanrooij, 2010) or with different scripts between the L1 and L2 (Showalter, 2018; Showalter & Hayes-Harb, 2015). In contrast, other studies concluded that spelling provided precise phonological information (Bürki et al., 2019; Welby et al., 2022). Beyond the phonetic deviation in vowel pronunciation, a benefit in spoken word recognition and production was observed (Bürki et al., 2019; Welby et al., 2022). These studies conducted in adults were generally focused on phonological processes (recognition and production) and used pseudowords as targets. Thus, they mirror studies of L1 word learning in children, where a new form is associated with a new meaning. Note that in several cases pseudowords were not attached to a meaning. This is not precisely what is at stake in L2 learning, especially in an academic context where a semantic representation already exists in an L1. In addition, a critical aspect is that children's reading experience and expertise change depending on their age. Whether their written word processing is automatized might affect the benefit of the presence of orthography in word learning. Therefore, L2 vocabulary learning in children needs to be further examined given that acquiring an L2 confronts children with different challenges than L1 word (or pseudoword) learning at different levels of reading experience.

<sup>1</sup> This problem of congruency can be illustrated as follows. The grapheme <u> is associated with the phoneme /u:/ in German and with the phoneme /y/ in French. Therefore, the German word <Schuh> will be pronounced /ʃy/ instead of /ʃu:/.

### The current study

Our study explored whether exposure to orthography facilitates the acquisition and memorization of L2 words in developing readers and the extent to which this orthographic advantage can be modulated by L1 reading experience. Using a paired-associated word-learning paradigm, we compared the learning of 16 German words (Experiment 1) and 24 German words (Experiment 2) in two groups of third- and fifth-graders who had French as their native language. We selected German for several reasons. First, this language is not familiar to children, unlike English (De Wilde et al., 2020), so we could test a completely unknown language. Second, German words differ from French words in terms of their orthography-to-phonology mappings. More precisely, German words can display a varying degree of orthography-to-phonology congruency with French grapheme-phoneme conversion (GPC) rules, thereby ensuring an ecological set of German features. For example, for the German word “Stern,” the grapheme <s> is associated with the phoneme /ʃ/ in German but with /s/ in French. On the contrary, the German word “Korb” can be read out rather correctly using French GPCs. Finally, German has a rather consistent orthography, meaning that French children would not need to deal with intralinguistic inconsistencies in addition to the problems of interlinguistic incongruencies mentioned above. During the learning phase, children were assigned to either an orthographic learning condition or a non-orthographic one. We chose a between-participants design rather than the more common within-participants design because we wanted to avoid any exposure to German spelling for participants in the non-orthographic condition. The orthographic condition was characterized by the simultaneous presentation of L2 written and spoken forms associated with a picture, reflecting their related concepts, whereas the non-orthographic condition substituted orthography by a series of hashes (to control for visual display). Although previous L1 word-learning studies used both production and recognition tasks, we used recognition measures to assess learning because L2 learners might have difficulties in pronouncing new phonemes, leading to potential difficulties in evaluating accuracy. The recognition task also allowed us to include many items to be learned, thereby ensuring robustness of the data. Finally, using recognition measures allowed us to focus on each constituent of the lexical representation—orthographic, phonological, and semantic (through the phonological-to-semantic connection) representations—and to determine how precisely they were integrated into the lexicon. To assess this, we designed three experimental tasks: (a) a picture designation task assessing the phonology-to-semantic pathway (choose the correct picture corresponding to the spoken form), (b) a spoken word recognition task (using a go/no-go task), and (c) an orthographic judgment task (recognition of the correct German written form among three written distractors). Learning performance was assessed immediately after the learning phase and after a 1-week delay to assess retention as previously examined in word-learning studies conducted in children (e.g., Chambré et al., 2017; Miles et al., 2016; Ricketts et al., 2021).

These tasks allowed us to establish whether any orthographic advantage occurred for different dimensions of vocabulary learning, including written, spoken, and conceptual forms. The study was mainly exploratory, given the scarcity of data in the domain of orthographic facilitation in a foreign language. Indeed, although there are several studies conducted in the L1 in monolinguals or in the L1/L2 in bilinguals (e.g., Jubenville et al., 2014), in most cases the children already knew the orthography-to-phonology mappings in the target language. Overall, we expected children assigned to the orthographic condition to outperform those in the non-orthographic one because they were the only ones to be exposed to the written form during learning. This was expected not only immediately after learning but also after a 1-week delay, in line with Ricketts et al.'s (2021) observations of a deeper knowledge of meaning after 1 week. In addition, we expected the delay to increase the benefit of orthography given that the simultaneous exposure to written and spoken forms may strengthen the lexical representation, which should be more robust to time decay (see Ricketts et al., 2021). With regard to the more specific tasks designed, we expected an orthographic advantage in phonological-to-semantic learning assessed through the spoken form-to-picture task given that the multimodal (spoken and written modalities) encoding of novel words may be particularly efficient to support semantic learning, in line with the lexical quality hypothesis (Perfetti & Hart, 2002). Regarding the spoken word recognition task, it remains unclear whether or not the presence of orthography facilitates the recognition of novel spoken forms. Indeed, children may be faced with great challenges when

decoding L2 words with L1 decoding rules while listening to L2 spoken forms. Despite these potential difficulties, we expected exposure to orthographic forms to still have a beneficial effect on spoken word learning given the existence of congruent parts of words on which to rely, even for words containing incongruencies (the word “Stern” contains one incongruent GPC between French and German but contains four congruent units). As for the orthographic task, it was obvious that children exposed to the written words would present higher scores in the orthographic judgment task compared with children in the non-orthographic condition. Importantly, the task was designed so that we could investigate the degree of precision of the orthographic representation by examining the type of errors produced by the participants.

In addition, two grades were tested to examine the effect of L1 reading experience on orthographic facilitation during L2 vocabulary learning. We tested third- and fifth-graders, because there is a consensus that reading automatization occurs around Grade 4 (Byrnes & Wasik, 2009). Although Salins et al. (2023) found an orthographic facilitation in an L1 for all groups of children from Grades 1 to 6, the effects were broader for the youngest groups. An explanation is that older children can generate orthographic expectancies even in the absence of orthographic exposure, making the presentation of orthography less critical. Although theoretically interesting, this mechanism might not apply to the L2 learning context of interest in the current study given that orthographic expectancies are potentially less strong due to specificities of L2 spoken (phonetic) forms, especially if children have not been exposed to foreign spellings during learning. Here, two views may be theoretically contrasted. On the one hand, older children might display more cross-language interference when exposed to orthographic forms of novel words to be learned because they have more automatized orthography-to-phonology mappings in the L1. This would suggest a weaker benefit of orthographic exposure for fifth graders than for third graders. On the other hand, older children may extract and rely on this new orthographic information more quickly, and therefore benefit more from orthographic exposure, because they are more familiar with written word processing. This would inversely suggest a stronger benefit of orthographic exposure for fifth-graders.

To recap, an orthographic facilitation was expected in all three tasks. We expected the presence of orthography to reinforce the new phonological representation (word recognition), to strengthen the connection between spoken words and semantics (picture designation), and to help memorize orthographic forms. As mentioned earlier, the impact of age is quite difficult to predict in an L2. Finally, vocabulary learning was tested twice, immediately and after a 1-week delay, allowing us to look at retention. Although it was hard to predict the impact of delay on retention overall, given that few studies have tested it in an L1 (Chambré et al., 2017; Ricketts et al., 2021), we did expect the orthographic facilitation to be stronger at the 1-week delay. Because it was difficult to anticipate how many foreign words children would be able to learn in a single session, we planned to conduct two experiments. In the first one children needed to learn 16 words, and in the second one children needed to learn 24 words.

## Experiment 1

### Method

#### Participants

An initial sample of 193 children was recruited from 11 elementary schools in the “Hauts de France” region of France. They were French native speakers or had been learning French for at least 6 years. Importantly, we selected participants from schools where exposure to foreign languages was limited to English for a maximum of 1 hr per week from Grade 3 onward. In addition, we ensured that our participants had no prior knowledge of German before the learning session by asking them to name selected images in German. None of them had specific educational needs. They had neither sensory nor diagnosed cognitive disorders. Unfortunately, several participants did not complete the whole study (i.e., learning, testing, and background cognitive and linguistic tasks) and therefore were removed from further analysis. Among the remaining 142 children, 69 were fifth-graders ( $M_{\text{age}} = 10.62$  years,  $SD = 0.49$ ; 44 girls) and 73 were third-graders ( $M_{\text{age}} = 9.02$  years,  $SD = 0.29$ ; 40 girls). Partici-



pants were randomly assigned to one of the two learning conditions; the orthographic condition involved the simultaneous presentation of written and spoken forms during learning ( $n = 72$ , including 34 third-graders and 38 fifth-graders), whereas the written form was substituted by a series of symbols in the non-orthographic condition ( $n = 60$ , including 29 third-graders and 31 fifth-graders). Their characteristics are summarized in Table 1. The experiment was approved by the research ethics committee of the University of Lille.

#### *Background cognitive and linguistic measures*

All the participants completed the background measures during the second testing session. Non-verbal reasoning was assessed through the colored progressive Raven matrices (Raven, 1981). Verbal intelligence was measured by the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) vocabulary subtest (Wechsler, 2005). Phonological short-term memory was assessed by the pseudoword repetition task subtest of the NEPSY II (Korkman et al., 2012), owing to the high correlation between these skills and vocabulary learning (Gupta et al., 2003). Reading age and reading skills (i.e., reading accuracy and fluency) were measured by the Alouette task (Lefavrais, 1967, 2005). In this test, children need to read a text of 265 words for a maximum of 3 min. If they read all the words in less than 3 min, their reading time is recorded. If not, the number of words read in 3 min is recorded. Errors are recorded in all cases. Reading accuracy was calculated by multiplying by 100 the ratio between the number of words accurately read compared with the total number of words read. Reading fluency refers to the number of words accurately read in a 3-min period: the higher the score, the more fluent the reading. The so-called reading age was obtained through reading using a table of correspondence between scores of accuracy and fluency and reading age (see Lefavrais, 1967).

Separate statistical descriptive analyses were conducted on third-graders and fifth-graders to ensure that participants were matched on their cognitive and linguistic skills in both learning conditions. They showed that third-graders were matched on all cognitive and linguistic skills (all  $p$  values  $> .05$ ) except for reading skills (for all three measures,  $p = .04$ ). Third-graders assigned to the orthographic condition exhibited higher scores in reading measures than those in the non-orthographic condition. Therefore, we used reading age as a covariable to control for differences in reading profiles across groups. The statistical descriptive analyses revealed homogeneous cognitive and linguistic skills in both groups of fifth-graders (all  $p$  values  $> .10$ ).

#### *Learning and testing phases*

The experiment was split into one learning session and two testing sessions. The immediate testing took place right after the learning session. The delayed testing session assessed learning performance after a 1-week delay. Learning and testing sessions took place in elementary schools, with the children being trained and tested individually. The experimental design is summarized in Fig. 1.

*Learning session.* The learning material included a set of 16 German words. This language was selected for its unfamiliarity to French children and its transparent orthography. The words were selected from the SUBTLEX-DE database (Brysbaert et al., 2011) and included four to seven letters ( $M = 5.00$ ,  $SD = 0.89$ ) and four to seven phonemes ( $M = 4.38$ ,  $SD = 0.62$ ). We ensured that their French equivalents were frequent enough so that their semantic representation was already integrated into the children's lexicon. Frequencies were extracted from the MANULEX database (Lété et al., 2004), which is a corpus base of children's reading books. They comprised from 20 occurrences per million for the French word *poire* (pear) to 774 occurrences per million for *tête* (head) ( $M_{\text{frequency}} = 169.90$ ,  $SD = 208.26$ ). To ensure the direct activation of the concept with exposure to the pictures, we also selected the words depending on their concreteness and imageability. Concreteness ratings of the French equivalents were estimated on a 5-point Likert scale ( $M = 4.61$ ,  $SD = 0.61$ ) (see Bonin et al., 2018). Imageability ratings were based on the Glasgow psycholinguistic norms (Scott et al., 2019) because, to our knowledge, no such database is standardized in French. Thus, imageability was estimated on a 7-point Likert scale ( $M = 6.69$ ,  $SD = 0.74$ ).

Neither German–French cognate words nor German–English ones were included in the experiment. We also examined the formal properties of the stimuli. We measured the words' minimal bigram frequency across languages using the CLEARPOND database (Marian et al., 2012) to examine

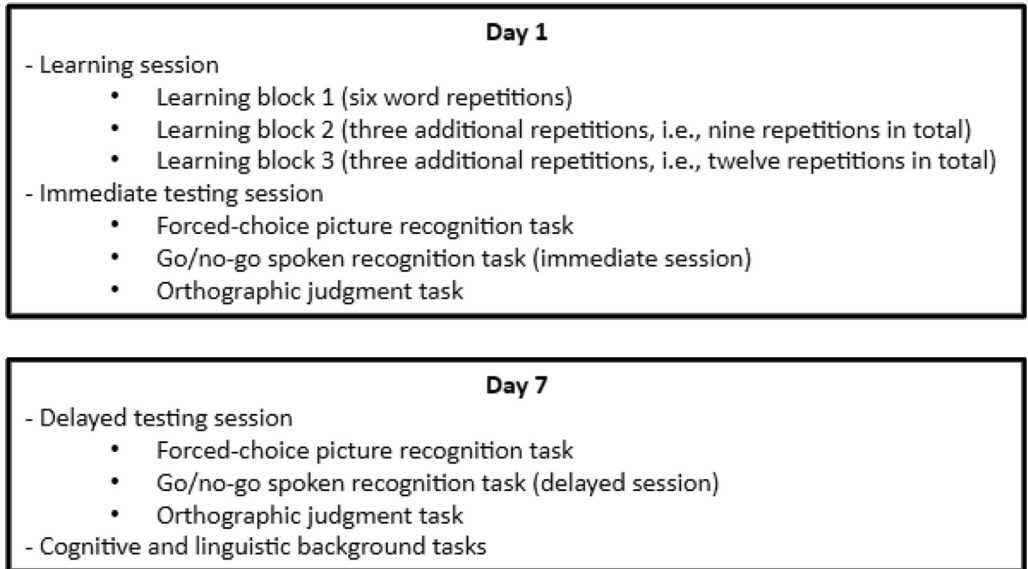


**Table 1**

Summary of participants' performances in cognitive and linguistic background tasks according to learning condition (orthographic vs. non-orthographic) and grade (third-graders vs. fifth-graders) (16 items)

	Orthographic condition		Non-orthographic condition		t test	p value	Cohen's d
	M	SD	M	SD			
<i>Third-graders</i>							
Chronological age	108.48	2.98	107.57	3.42	1.10	.28	
Reading age	113.06	17.89	104.54	15.95	2.12	.04	0.51
Reading fluency	251.17	86.27	209.32	87.19	2.02	.04	0.48
Reading accuracy	94.63	3.44	92.76	3.71	2.19	.04	0.52
NWRT (max = 40)	32.68	4.23	32.33	4.98	0.30	.76	
Raven matrices (max = 36)	23.59	3.79	22.07	3.77	-1.54	.13	
L1 vocabulary (max = 68)	24.09	5.47	23.76	5.36	0.26	.80	
<i>Fifth-graders</i>							
Chronological age	127.78	5.28	127.28	5.75	-0.37	.72	
Reading age	132.11	22.61	131.48	23.55	-0.11	.91	
Reading fluency	338.50	97.73	333.63	102.43	-0.20	.85	
Reading accuracy	96.39	2.47	96.90	1.62	0.96	.34	
NWRT (max = 40)	34.76	3.58	35.11	3.78	0.37	.71	
Raven matrices (max = 36)	26.78	4.32	26.21	5.00	0.37	.71	
L1 vocabulary (max = 68)	35.97	5.92	34.66	5.73	-0.90	.37	

Note. Reading measures were obtained from the Alouette test (Lefavrais, 1967, 2005). Reading age was inferred through reading accuracy and fluency by using standardized reading age tables (Lefavrais, 1967). Reading fluency scores reflect the number of words that would have been read in 3 min (reading fluency score > 265 means that the participant read the whole text in < 3 min). Reading accuracy was calculated by multiplying by 100 the ratio between the number of words read accurately and the number of words read. NWRT, nonword repetition task subtest of NEPSY-II (Korkman et al., 2012) and assessed short-term phonological memory skills; L1 vocabulary, first-language (French) vocabulary subtest of Wechsler Intelligence Scale for Children (Wechsler, 2005). Raven matrices were used as a measure of nonverbal reasoning skills (Raven, 1981).



**Fig. 1.** Organization of the learning and testing sessions.

their homogeneity in terms of orthographic markedness, a variable known to affect word learning (Bartolotti & Marian, 2017; Major, 1987; Pérez-Serrano et al., 2021). Finally, German words differed from French ones in terms of their orthography-to-phonology mappings. The selected German words were characterized by a variable degree of orthography-to-phonology congruency according to French GPC rules, thereby ensuring an ecological set of German features. For example, for the German word “Stern,” the grapheme <s> is associated with the phoneme /ʃ/ in German but with /s/ in French. Stimuli are presented in [Appendix 1 in the online supplementary material](#).

*Pictures and sounds.* The chosen German words were paired with black and white pictures selected from the MultiPic database (Duñabeitia et al., 2018). The pronunciation of each German word was recorded by three native German speakers using Audacity software (Version 2.1.1). We performed a normalization procedure as well as a noise attenuation procedure of the initial audio stream. The choice of offering three different voices for the learning phase was based on two factors. First, it ensured that participants would not rely on specific prosodic indices such as tone, pitch, and accentuation. Second, owing to the variability in speakers' pronunciations, participants needed to adjust their phonological representation of each German word, which may have helped them to improve their spoken recognition of these words (Tapia et al., 2021). Children heard the three voices in the learning phase with a rotation for each learning block. Similarly, they were tested with these three voices (one voice per test, with a rotation across participants).

*Procedure.* The learning session took place in each elementary school using a laptop computer. We first ensured that each picture was correctly recognized by asking children to name each picture in their native language. The learning session then started. In the orthographic condition, the German spoken and written forms were simultaneously presented, associated with their related picture. In the non-orthographic condition, the German written form was substituted by a series of hashes<sup>2</sup> (####) to match for the general visual display. Regardless of the learning condition, therefore, all participants were exposed to the same number of audiovisual sources of information during learning. [Fig. 2](#) illustrates an example of a trial according to the learning condition. Each picture (and written word in the orthographic condition) was displayed for 3 s. Children then needed to press the response key to access the next word. During the learning session, each German word was presented 12 times. This ensured that children had sufficient exposure to the learning material. This degree of exposure is consistent with the recommendations promoting vocabulary learning in a natural setting (see Beck et al., 2013). The learning session lasted 30 to 40 min.

*Testing sessions.* Learning performance was assessed by three computerized tasks: a picture designation task, a spoken recognition task, and an orthographic judgment task. They were carried out both immediately after learning and after a 1-week delay in the same order. After each experimental task, there was a shift in the speaker's voice.<sup>3</sup> Both testing sessions lasted 20 to 35 min. Background cognitive and linguistic skills were assessed at the end of the 1-week delayed testing. The testing material is available in the [appendices in the supplementary material](#) (see [Appendix 2](#) for the picture designation task, [Appendix 3](#) for the spoken recognition task, and [Appendix 4](#) for the orthographic judgment task).

*Picture designation task.* The picture recognition task was used to assess the phonology-to-semantic pathway. In each trial, an array of four pictures (i.e., the target and three trained pictures) was displayed on the computer screen in a 2 × 2 grid. The pronunciation of a German word was then provided through the children's headphones. Participants needed to select the picture to which the spoken word form referred. The positions of both the target and distractive pictures were counterbalanced between trials. Accuracy was recorded for each trial.

*Spoken word recognition task.* The spoken word recognition task was designed as a “purely” phonological task to assess the ability of children to recognize German spoken words and the degree of precision of their phonological representation.

Two lists of 16 pseudowords were designed using the Wuggy software (Keuleers & Brysbaert, 2010). They were close phonological distractors to the learning list, differing by one or two phonemes

<sup>2</sup> A series of hashes was preferred to a series of letter-like symbols, which could be confusing, as pointed by Ricketts et al. (2015).

<sup>3</sup> We used the same speakers for the learning and testing sessions.



Orthographic learning condition	Non-orthographic learning condition
<p style="text-align: center;">Birne</p>  <p style="text-align: center;">/bɪrnə/</p>	<p style="text-align: center;">#####</p>  <p style="text-align: center;">/bɪrnə/</p>

Fig. 2. Example of a learning trial according to the learning condition (orthographic vs. non-orthographic learning condition).

from the German words. These close phonological distractors were recorded at the same time as the German words by the three native speakers to ensure that their pronunciations were as word-like as possible.

For each trial, a spoken item (word vs. pseudoword) was played in the children’s headphones. No visual indices were provided during the experimental task except for a fixation cross presented in each trial. Participants needed to press the response key as fast as possible if they recognized the German spoken words, but they should refrain from making any response for pseudowords. This so-called go/no-go lexical decision task is recommended instead of the “standard” yes/no task in children (Moret-Tatay & Perea, 2011). The upcoming items were displayed immediately after pressing the response key or after a 3-s delay. Accuracy and response times were recorded.

*Orthographic judgment task.* The orthographic judgment task assessed whether participants were able to recognize the spelling of German words when exposed to their written form during learning. It also evaluated the degree of precision of the German orthographic representations by manipulating the nature of the distractors. Responses in the non-orthographic condition served as a baseline, providing information on the orthographic predictions generated by the children in the absence of orthographic exposure.

For each German word, three types of distractors were designed. Close orthographic distractors were created through a one-letter transposition (e.g., Pferd–Pfred, horse in English). Distant distractors shared only a small orthographic overlap with the target German word (one to three letters depending on word size, e.g., Pferd–Pfalt). The phonological distractors were homophonic with the target word when using the French orthography–phonology mappings (e.g., Pferd–Pferfe).

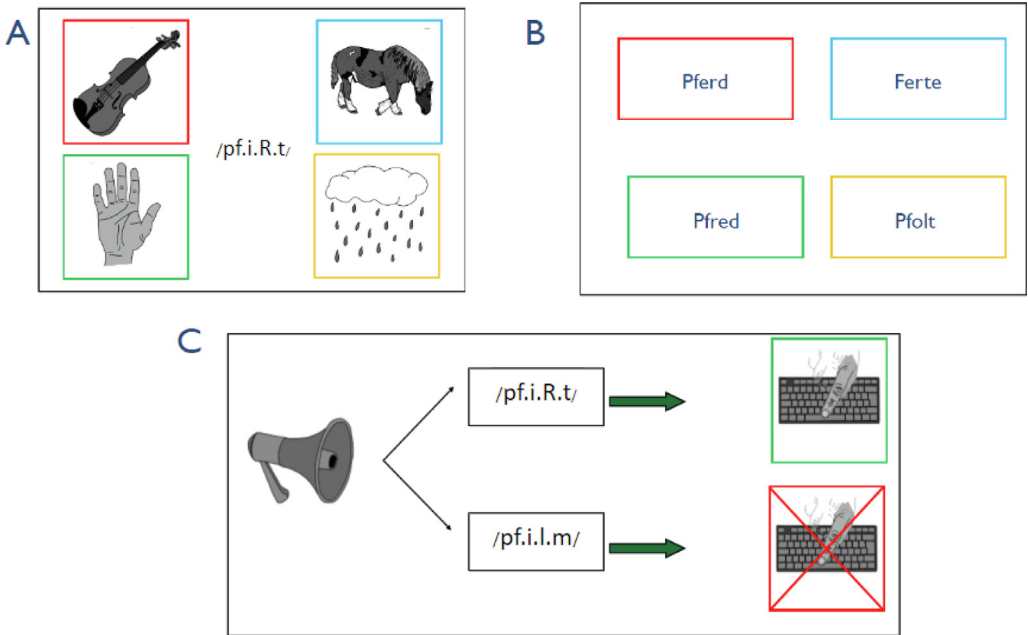
Four written versions of a German word (i.e., the target and three distractors) were displayed on the screen in a 2 × 2 grid. Participants needed to recognize the accurate written form of each German target word by clicking on one of the written transcriptions with the computer mouse. The positions of the target and the distractive written forms were counterbalanced across sessions and trials. Accuracy was recorded, as was the selected distractive spelling for each error.

Examples of trials for all tasks are provided in Fig. 3.

**Results**

*Data analysis*

Statistical analyses were conducted by using generalized mixed-effects models because this technique is the most suitable to take into consideration simultaneously the variability associated with participants, items, and schools. These statistical analyses were conducted on R software (R Core



**Fig. 3.** Example of a trial in each experimental task (A: forced-choice picture word recognition task; B: orthographic judgment task; C: go/no-go spoken word recognition task). No spoken information was provided during the orthographic judgment task (B). No written information was provided during the forced-choice picture word recognition task (A) or during the go/no-go spoken word recognition task (C). For the orthographic judgment task (B), there were a German word (Pferd) and three distractive written forms : a close orthographic distractor (Pfred), a distant orthographic distractor (Pfolt), and a phonological distractor (Ferte).

Team, 2017), using the *lmer* and *glmer* functions from the “lme4” package (Matuschek et al., 2017) as well as the “afex” package for estimating *p* values. Indeed, this approach was the most conservative one by producing lower Type I error rates compared with the *t-as-z* approach (Luke, 2017). We performed these analyses on the accuracy measures for each task. Statistical analyses on response times were conducted on the spoken word recognition task because it was the only experimental task that included a speed criterion. The random structure was initially set following a compromise between the maximal random structure (Barr et al., 2013) and the parsimonious one (Matuschek et al., 2017). This aimed to limit the models’ lack of convergence. Once the random structure was established, we then included the fixed effects of grade, learning condition, and session. This followed a model comparison approach for which each fixed and interaction effect was progressively included as long as it led to a better fitted model, that is, with the smallest Akaike information criterion (AIC). This statistical approach seemed to be highly relevant given that the literature did not provide a clear picture on how our variables might modulate the learning performance. For models with binary outcome variables, we computed mixed logistic models. For continuous outcomes, especially for the spoken word recognition task, we applied mixed linear models. Significant main and interaction effects were reported using a cutoff point of  $p < .05$ . Subset models were computed in the event of interaction to identify the contribution of each variable. As a reminder, third-graders were not matched on their reading skills, so we included reading age as a covariate in the statistical analyses. For all three experimental tasks, we present the best fitted model below. Descriptive statistics are presented for each experimental task in Table 2. We summarize the model and its parameters for each of the three experimental tasks in Table 3.

**Table 2**

Participants' performance on all three experimental tasks according to grade (third-graders vs. fifth-graders), orthographic learning condition (OLM vs. NOLM), and session (immediate vs. delayed) in both experiments.

	Experiment 1 (16 German words)								Experiment 2 (24 German words)							
	Third-graders				Fifth-graders				Third-graders				Fifth-graders			
	NOLM		OLM		NOLM		OLM		NOLM		OLM		NOLM		OLM	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Forced-choice recognition task (accuracy)																
Immediate	48.72	23.87	72.16	22.17	71.88	16.65	83.88	16.35	50.30	19.31	73.28	18.65	63.36	22.42	80.71	13.58
Delayed	51.76	21.01	68.94	20.14	72.60	15.88	80.72	14.58	54.46	16.59	74.53	15.07	61.80	18.61	80.86	13.14
Spoken recognition task (discrimination score)																
Immediate	1.28	0.58	1.55	0.64	1.73	0.66	1.80	0.61	0.88	0.47	1.43	0.59	1.29	0.66	1.60	0.64
Delayed	1.01	0.77	1.78	0.78	1.66	0.70	1.80	0.68	0.94	0.83	1.32	0.75	1.18	0.61	1.84	0.87
Spoken recognition task (response time)																
Immediate	1537	305	1572	315	1491	274	1576	225	1518	203	1573	211	1518	207	1502	285
Delayed	1562	289	1501	268	1434	242	1499	209	1574	214	1569	213	1523	181	1450	244
Orthographic judgment task (accuracy)																
Immediate	28.71	14.37	61.13	18.11	43.15	23.05	74.18	16.30	30.91	10.15	60.34	18.95	38.67	21.55	70.22	16.08
Delayed	29.81	12.36	63.48	18.97	44.76	24.28	73.85	19.08	31.85	14.00	56.90	20.74	38.41	18.59	68.06	16.79

Note. OLM, orthographic learning condition; NOLM, non-orthographic learning condition.

**Table 3**

Summary of the logistic mixed model analysis for variables predicting accuracy in the forced-choice recognition task and the orthographic judgment task

Model and predictors	Estimate	SE	95% CI		z value	p value
			LL	UL		
<i>Forced-choice recognition task</i>						
Intercept	0.216	0.091	0.095	0.494	-3.631	<.001
Reading age	1.015	0.003	1.008	1.021	4.296	<.001
Learning condition	2.491	0.355	1.885	3.293	6.411	<.001
Grade	2.123	0.364	1.516	2.971	4.387	<.001
Model = glmer (forced_choice\$Accuracy~Read_age + Learning_condition + Grade +(1 participant) + (1 item)+(1 school),data=forced_choice,family=binomial(link=logit),control=glmerControl(optimizer="bobyqa")						
<i>Orthographic judgment task</i>						
Intercept	0.061	0.025	0.028	0.136	-6.853	<.001
Reading age	1.016	0.003	1.009	1.023	4.791	<.001
Learning condition	4.247	0.553	3.290	5.483	11.100	<.001
Grade	1.629	0.256	1.197	2.217	3.103	.002
Model = glmer (forced_choice\$Accuracy~Read_age + Learning_condition + Grade +(1 participant) + (1 item)+(1 school),data=forced_choice,family=binomial(link=logit),control=glmerControl(optimizer="bobyqa")						

Note. CI, confidence interval; LL, lower limit; UL, upper limit.

*Picture designation task*

In the group of third-graders, 2 children assigned to the orthographic condition had recognition scores below 3 standard deviations (i.e., only one German word was accurately recognized); therefore, they were removed from the study. Thus, statistical analyses were conducted on 140 participants: 71 third-graders and 69 fifth-graders.

We computed mixed logistic models to analyze participants' performance in the forced-choice spoken word-picture recognition task. The best fitted model included reading age, grade (third-graders vs. fifth-graders), and learning condition (orthographic vs. non-orthographic) as fixed effects and by-participant, by-item, and by-school random intercepts, AIC = 4930,  $\chi^2(1) = 36.83, p < .001$ . The inclusion of the main effect of session did not fit the model, AIC = 4935,  $\chi^2(1) < 1, p = .70, ns$ . Despite the inclusion of reading age as a covariate, the main effect of grade remained significant, with an odds of accurate picture recognition 2.12 times higher in fifth-graders than in third-graders (78.38% vs. 58.46%,  $p < .001$ ). Children assigned to the orthographic learning condition had an odds of accurate picture recognition 2.49 times higher than that for children in the non-orthographic condition (74.44% vs. 60.26%,  $p < .001$ ). Interestingly, the inclusion of the interaction between grade and learning condition did not lead to a better fitted model,  $\chi^2(1) < 1, p = .81, ns$ .

*Spoken word recognition task*

Surprisingly, participants committed many false alarms for one of the pseudowords, namely Kitte. This pseudoword was phonologically too ambiguous to be correctly rejected by participants, so we decided to remove it from the analysis (i.e., 2.1% of the remaining data). We also removed response times that were shorter than 300 ms (i.e., 3.7% of the remaining data). This led to removing more than half of the responses for 3 third-grade children, who thus were excluded from the analysis. Given that this task may elicit certain response strategies such as pressing the space key for all items or for none, we fit our statistical approach into the framework of signal detection theory and calculated  $d'$  scores for sensitivity, which we termed *discrimination scores*. Discrimination scores were calculated as the difference between the z-transformed distribution of accurate word recognition (hits) and that of the acceptance of pseudowords (false alarms). Extreme recognition scores were adjusted by using [Macmillan and Kaplan's \(1985\)](#) transformation. A two-step cutoff was then applied to remove participants whose discrimination score was below 0 as well as those who committed more false alarms than the random level during the immediate testing session only. Unexpectedly, 12 participants (10 third-graders and 2 fifth-graders) met the cutoff and were removed from the study. Therefore, the statistical analyses were conducted on the remaining 117 participants, among whom 56 (including 30



third-graders) were assigned to the non-orthographic condition and 61 (including 25 third-graders) were in the orthographic learning condition.

*Discrimination scores.* We used generalized mixed models to analyze participants' discriminative scores for immediate and delayed testing. The random structure included only by-participant random intercepts given that discriminative scores were calculated on all items. The best fitted model included grade, reading age, and learning condition as fixed effects and by-participant random intercept,  $AIC = 476$ ,  $\chi^2(1) = 3.65$ ,  $p = .06$ . We found a main effect of the learning condition ( $p = .001$ , congruent), indicating that children assigned to the orthographic condition outperformed those in the non-orthographic condition (1.74 vs. 1.40). The main effect of grade was significant ( $p = .006$ ). Fifth-graders exhibited better discriminative performance than third-graders (1.75 vs. 1.38). The interaction between learning condition and grade did not reach significance ( $p = .06$ ). Again, the inclusion of the main effect of session did not lead to a better fitted model,  $AIC = 478$ ,  $\chi^2(1) < 1$ ,  $p = .71$ , *ns*.

*Response times.* Statistical analyses were conducted on raw response times using the generalized linear mixed model. The best fitted model included reading age and session as fixed effects and by-participant, by-item, and by-school random intercepts,  $AIC = 3769$ ,  $\chi^2(1) = 18.64$ ,  $p < .001$ . The main effect of session was significant, with faster response times in the delayed session than in the immediate one (1490 vs. 1570 ms,  $p < .001$ ). Reading age was a nonsignificant covariable in the analysis of response times ( $p = .40$ , *ns*).

#### *Orthographic judgment task*

Owing to technical issues, the performance in the orthographic judgment task was missing for 2 participants. Statistical analyses were conducted on the remaining 140 participants, with 70 children assigned to the non-orthographic learning condition (including 39 third-graders) and 70 in the orthographic condition (including 32 third-graders). Given that accuracy was a binary measure, we computed mixed logistic models. The best fitted model included reading age, learning condition, and grade as fixed effects and by-participant, by-item, and by-school random intercepts,  $AIC = 5185$ ,  $\chi^2(1) = 27.68$ ,  $p < .001$ . As expected, children who were exposed to orthography during learning exhibited an odds of accurate written form recognition 4.25 times higher than that of children who were not (68.66% vs. 35.78%,  $p < .001$ ). Despite the inclusion of reading age as a covariate variable, the effect of grade was still significant, with an odds 1.63 times higher among fifth-graders than among third-graders (60.50% vs. 35.78%,  $p < .001$ ). Neither the inclusion of the main effect of session ( $AIC = 5187$ ,  $p = .85$ , *ns*) nor the interaction between learning condition and grade ( $AIC = 5161$ ,  $p = .88$ , *ns*) better fit the model to the data.

*Patterns of errors.* For exploratory purposes, we also conducted statistical analyses on the errors committed. As a reminder, three different types of distractors were presented in each trial: close orthographic, distant orthographic, and phonological. First, chi-square homogeneity tests were conducted to determine whether the errors committed were distributed evenly on the three types of distractors across learning conditions (orthographic vs. non-orthographic learning condition) and among third-graders and fifth-graders. Second, if significant and given that the chi-square homogeneity test is an omnibus test, we conducted Scheffé post hoc pairwise comparisons<sup>4</sup> between individual conditions in accordance with the Goodman procedure (Goodman, 1963) described by Franke et al. (2011). All pairwise comparisons are presented in Table 4.

A  $2 \times 3$  chi-square homogeneity test conducted between learning condition and type of error revealed that errors were not evenly distributed across learning conditions,  $\chi^2(2, 2138) = 90.32$ ,  $p < .001$ , Cramer's  $V = .21$ . Post hoc analyses showed that the proportion of errors committed was larger for close orthographic distractors in the orthographic learning condition than in the non-orthographic condition (54.6% vs. 33.1%),  $\chi^2(1, 858) = 90.32$ ,  $p < .001$ , but was smaller for distant

<sup>4</sup> The Scheffé post hoc test is usually used for analysis of variance but also, and to a lesser extent, for posttest comparisons following a chi-square homogeneity test. We used this specific post hoc test for its conservative approach, that is, minimizing Type I errors. The Scheffé critical value corresponds to the squared chi-square statistic.

**Table 4**

Summary of the mixed linear regression model for variables predicting discrimination between words and close phonological distractors in the spoken recognition task in the analysis including immediate and delayed sessions

Model and predictors	Estimate	SE	95% CI		t value	p value
			LL	UL		
<i>Immediate session vs. delayed session</i>						
Intercept	0.759	0.293	0.107	1.202	2.594	.01
Reading age	0.004	0.002	-0.000	0.009	1.380	.18
Learning condition	0.526	0.157	0.168	0.794	3.243	.001
Grade	0.467	0.167	0.236	0.838	2.860	.006
Learning Condition × Grade	-0.412	0.215	-0.857	-0.023	-1.91	.06
<i>Third-graders</i>						
Intercept	1.020	0.530	-0.037	2.016	1.96	.05
Reading age	0.001	0.005	-0.008	0.011	0.223	.82
Learning condition	0.547	0.175	0.218	0.902	3.120	.003
<i>Fifth-graders</i>						
Intercept	1.062	0.391	0.222	1.751	2.720	.009
Reading age	0.005	0.003	-0.000	0.011	1.667	.11
Learning condition	0.113	0.134	-0.134	0.394	0.840	.40

Model=lmer(go\_nogo\$Discrimination\_score ~ Read\_age+Learning\_condition + Grade + Grade\*Learning\_condition +(1|participant),data=go\_nogo).

Note. CI, confidence interval; LL, lower limit; UL, upper limit.

orthographic distractors (12.7% vs. 19.2%),  $\chi^2(1, 365) = 16.10, p < .001$ , and for phonological distractors (32.8% vs. 47.7%),  $\chi^2(1, 915) = 45.77, p < .001$ . Further statistical analyses were conducted following the same procedure as above to determine whether the distribution of errors was modulated by grade and by both grade and learning condition. They revealed that the proportion of committed errors was not equally distributed in fifth-graders and third-graders,  $\chi^2(2, 2138) = 13.42, p = .001$ , Cramer's  $V = .08$ . Post hoc analyses showed that fifth-graders committed a larger proportion of errors for close orthographic distractors compared with third graders (44.2% vs. 37.4%),  $\chi^2(1, 858) = 9.86, p = .01$ , but committed a smaller one for distant orthographic distractors (14.2% vs 19.0%),  $\chi^2(1, 365) = 8.87, p = .02$ . Chi-square homogeneity tests conducted between grade and learning condition did not reach significance ( $p > .05, ns$ ). These analyses are summarized in Table 5.

*Discussion*

Experiment 1 revealed orthographic facilitation in L2 vocabulary learning, including semantic, orthographic, and phonological dimensions, which was observed regardless of grade level. This suggests a generalized benefit from the presence of orthography in the early steps of L2 word learning, given that German was unknown to all participants and in readers as young as third-graders. Although this advantage was expected for orthographic and semantic learning, it was not as clear-cut for L2 phonological learning owing to the cross-language incongruencies of grapheme-to-phoneme mappings, making the contribution of orthography less predictive of L2 spoken representations. As mentioned in the Introduction, we conducted a second experiment in which the number of words to be learned was increased to 24. However, the construction of the list was similar, including both congruent and incongruent mappings.

**Experiment 2**

*Method*

*Participants*

Participants were 135 children recruited in the same elementary schools as in Experiment 1. Consistent with the previous experiment, we removed participants who did not complete the whole

**Table 5**

Summary of the pairwise comparisons conducted on the pattern of committed errors in the orthographic judgment task for Experiment 1

	$\psi$	SE	Scheffé test	p
<i>Close orthographic distractor</i>				
Non-orthographic condition vs. orthographic condition	-.215	.023	-9.54	<.001
Third-graders vs. fifth-graders	-.068	.022	-3.14	.01
Non-orthographic condition: Third-graders vs. fifth-graders	-.059	.026	-2.30	.07
<i>Distant orthographic distractor</i>				
Non-orthographic condition vs. orthographic condition	.065	.016	4.01	.0003
Third-graders vs. fifth-graders	.048	.016	2.98	.02
Non-orthographic condition: Third-graders vs. fifth-graders	.044	.021	2.09	.11
<i>Phonological distractor</i>				
Non-orthographic condition vs. orthographic condition	.149	.022	6.77	<.001
Third-graders vs. fifth-graders	.020	.022	0.91	.66
Non-orthographic condition: Third-graders vs. fifth-graders	.015	.027	0.58	.85

**Table 6**

Summary of the participants' performance on background tasks according to the learning condition (orthographic vs. non-orthographic) and grade (third-graders vs. fifth-graders) in Experiment 2 (24 items)

	Orthographic condition		Non-orthographic condition		t test	p value
	M	SD	M	SD		
<i>Third-graders</i>						
Chronological age	106.38	2.80	105.48	3.69	1.06	.30
Reading age	106.10	16.03	107.86	17.86	-0.39	.70
Reading fluency	211.46	84.27	221.71	91.25	-0.44	.66
Reading accuracy	94.49	3.31	93.81	3.93	0.71	.48
NWRT (max = 40)	34.48	2.31	34.00	3.56	0.59	.56
Raven matrices (max = 36)	22.43	4.53	23.04	6.09	-0.42	.68
L1 vocabulary (max = 68)	27.21	5.85	28.61	6.36	-0.87	.39
<i>Fifth-graders</i>						
Chronological age	128.30	5.67	127.16	5.39	0.79	.43
Reading age	126.04	20.81	118.91	26.60	1.15	.26
Reading fluency	311.22	84.76	288.25	115.15	0.44	.66
Reading accuracy	96.28	2.41	96.03	2.09	0.86	.39
NWRT (max = 40)	34.21	2.38	35.07	2.93	0.99	.33
Raven matrices (max = 36)	23.46	4.95	24.90	5.11	-1.08	.29
L1 vocabulary (max = 68)	33.27	7.89	31.29	7.80	0.95	.35

Note. Reading measures were obtained from the Alouette test (Lefavrais, 1967, 2005). Reading age was inferred through reading accuracy and fluency by using standardized reading age tables (Lefavrais, 1967). Reading fluency scores reflect the number of words that would have been read in 3 min (reading fluency score > 265 means that the participant read the whole text in < 3 min). Reading accuracy was calculated by multiplying by 100 the ratio between the number of words read accurately and the number of words read. NWRT, nonword repetition task subtest of NEPSY-II (Korkman et al., 2012) and assessed short-term phonological memory skills; L1 vocabulary, first-language (French) vocabulary subtest of Wechsler Intelligence Scale for Children (Wechsler, 2005). Raven matrices were used as a measure of nonverbal reasoning skills (Raven, 1981).

experiment, leaving us with 116 remaining participants, including 57 third-graders and 59 fifth-graders. Of these participants, 60 children, including 28 third-graders and 32 fifth-graders, were assigned to a non-orthographic condition. The 56 remaining children, including 29 third-graders and 27 fifth-graders, were assigned to an orthographic learning condition. Participants were matched on their cognitive and linguistic performance (see Table 5). When examining the profiles of participants in Experiments 1 and 2 (see Table 6), we found that children were not matched on some of the background tasks across our two experiments (see also Table 7 and General Discussion).

**Table 7**

Summary of the participants' performances on cognitive and linguistic background tasks between Experiment 1 (16 items) and Experiment 2 (24 items)

	Experiment 1 (16 items)		Experiment 2 (24 items)		t value	p value	Cohen's d
	M	SD	M	SD			
<i>Third graders</i>							
<i>Orthographic learning condition</i>							
Age (in months)	108.63	2.88	106.38	2.80	3.08	.003	0.79
Reading age	113.06	17.89	106.10	16.03	1.59	.12	
Reading fluency	251.17	86.27	211.46	84.27	1.82	.08	
Reading accuracy	94.63	3.44	94.49	3.31	0.16	.87	
NWRT	32.68	4.23	34.48	2.31	-1.97	.05	-0.51
Raven matrices	23.59	3.79	22.43	4.53	-0.14	.88	
L1 vocabulary	24.09	5.47	27.21	5.85	-2.15	.04	-0.55
<i>Non-orthographic learning condition</i>							
Age (in months)	108.00	3.81	105.18	3.69	3.03	.004	0.75
Reading age	104.54	15.95	107.86	17.86	-0.80	.43	
Reading fluency	209.32	87.19	221.71	91.25	-0.56	.58	
Reading accuracy	92.76	3.71	93.81	3.93	-1.12	.27	
NWRT	32.33	4.98	34.00	3.56	-1.48	.15	
Raven matrices	22.07	3.77	23.04	6.09	-1.85	.07	
L1 vocabulary	23.76	5.36	28.61	6.36	-3.35	.001	-0.84
	Experiment 1 (16 items)		Experiment 2 (24 items)		t value	p value	Cohen's d
	M	SD	M	SD			
<i>Fifth graders</i>							
<i>Orthographic learning condition</i>							
Age (in months)	127.73	5.92	128.30	5.67	-0.39	.70	
Reading age	131.43	23.10	126.04	20.81	0.96	.34	
Reading fluency	336.73	98.06	311.22	84.76	1.09	.28	
Reading accuracy	96.48	2.43	96.28	2.41	0.31	.76	
NWRT	34.50	3.61	34.21	2.38	0.31	.76	
Raven matrices	25.53	4.29	23.46	4.95	1.73	.09	
L1 vocabulary	35.25	5.94	33.27	7.89	1.13	.26	
<i>Non-orthographic learning condition</i>							
Age (in months)	127.34	5.58	127.16	5.39	0.14	.89	
Reading age	132.38	22.42	118.91	26.60	2.22	.03	0.55
Reading fluency	338.55	97.49	288.25	115.15	1.89	.06	
Reading accuracy	97.03	1.60	96.03	2.09	2.15	.04	0.54
NWRT	35.19	3.58	35.07	2.93	0.14	.89	
Raven matrices	26.25	4.79	24.90	5.11	1.08	.28	
L1 vocabulary	34.81	5.49	31.29	7.80	2.08	.04	0.52

Note. Reading measures were obtained from the Alouette test (Lefavrais, 1967, 2005). Reading age was inferred through reading accuracy and fluency by using standardized reading age tables (Lefavrais, 1967). Reading fluency scores reflect the number of words that would have been read in 3 min (reading fluency score > 265 means that the participant read the whole text in <3 min). Reading accuracy was calculated by multiplying by 100 the ratio between the number of words read accurately and the number of words read. NWRT, nonword repetition task subtest of NEPSY-II (Korkman et al., 2012) and assessed short-term phonological memory skills; L1 vocabulary, first-language (French) vocabulary subtest of Wechsler Intelligence Scale for Children (Wechsler, 2005). Raven matrices were used as a measure of nonverbal reasoning skills (Raven, 1981).

*Learning and testing sessions*

*Words.* An additional 8 supplementary German words were selected from the SUBTLEX database (see Brysbaert et al., 2011) and were added to the 16 used in Experiment 1 (see Appendix 5 in supplementary material for the list of stimuli). The two learning lists (i.e., 16 items and 24 items) were comparable regarding linguistic characteristics, especially their letter and phoneme length, and the level of phonological and orthographic overlapping of the French equivalents and their frequency, concreteness, and imageability (all p values > .10).

**Pictures and sounds.** An additional 8 black and white pictures were selected from the MultiPic database (Duñabeitia et al., 2018). All words were pronounced by the same three native German speakers as in Experiment 1, and the same recording and standardization of the audio stream were performed.

**Learning session and testing tasks.** The procedures for both learning and testing sessions were identical to those in Experiment 1. The only difference was that learning lasted about 40 to 45 min and testing lasted about 30 min.

The testing material is available in the appendices in the [supplementary material](#) (see [Appendix 6](#) for the picture designation task, [Appendix 7](#) for the spoken recognition task, and [Appendix 8](#) for the orthographic judgment task).

## Results

Performance in all experimental tasks is detailed in [Table 2](#). The model and its parameters for each experimental task are presented in [Tables 8–10](#). Consistent with Experiment 1, the inclusion of the main effect of session did not better fit to the data for each experimental task,  $\chi^2 < 1$ , all  $p$  values  $> .10$ . Therefore, we did not include the effect of session in the following analyses.

### Picture designation task

The best fitted model included learning condition and grade as fixed effects and by-participant, by-item, and by-school random intercepts,  $AIC = 6035$ ,  $\chi^2(1) = 10.77$ ,  $p = .001$ , as in Experiment 1. We found a significant main effect of the learning condition ( $p < .001$ ), with an odds 2.29 times higher in children in the orthographic condition than in children in the non-orthographic condition. There was also a main effect of grade ( $p = .001$ ), with an odds of accurate recognition 1.79 times higher in fifth-graders than in the third-graders. The interaction between grade and learning condition did not lead to a better adjustment of the model,  $AIC = 6037$ ,  $\chi^2 < 1$ ,  $p = .41$ , *ns*. A summary of the model is presented in [Table 8](#).

### Spoken word recognition task

Response times shorter than 300 ms were removed prior to further statistical analysis (i.e., 6.07% of the remaining data). This led to removing more than 50% of the response times for 6 participants (i.e., 4 third-graders and 2 fifth-graders), who therefore were excluded from the analysis. Once again, we calculated discrimination scores for each participant, excluding from the analysis the 7 participants whose false alarm rate was above the random level. Statistical analyses were conducted on the 103 remaining participants, among whom 49 (including 23 third-graders) were assigned to the non-orthographic condition and 54 (including 28 third-graders) were in the orthographic condition.

**Discrimination scores.** Following the same procedure as in Experiment 1, we found that the best fitted model included learning condition and grade as fixed effects and by-participant random intercepts,  $AIC = 408$ ,  $\chi^2(1) = 8.93$ ,  $p = .003$ . There was a significant main effect of learning condition, suggesting higher discrimination scores in the orthographic condition than in the non-orthographic condition (1.54 vs. 1.09,  $p < .001$ ). There was also a main effect of grade, with fifth-graders outperforming third-graders (1.47 vs. 1.16,  $p = .004$ ). Again, the inclusion of the main effect of session did not lead to a better adjusted model,  $AIC = 409$ ,  $\chi^2(1) < 1$ , *ns*. A summary of the model is presented in [Table 9](#).

**Response times.** The best fitted model included grade as a fixed effect and by-participant, by-item, and by-school random intercepts,  $AIC = 5007$ ,  $\chi^2(1) = 2.70$ ,  $p = .09$ , with a marginal main effect of grade ( $p = .07$ ). Thus, third-graders tended to recognize German words more slowly than fifth-graders (1559 vs. 1471 ms).

### Orthographic judgment task

Statistical models were computed on accuracy measures using logistical mixed models. The best fitted model here included learning condition and grade as fixed effects and by-participant, by-item, and by-school random intercepts,  $AIC = 6864$ ,  $\chi^2(1) = 5.74$ ,  $p = .02$ . As in Experiment 1, we found

**Table 8**

Summary of the logistic mixed model analysis for variables predicting accuracy in the picture designation task and in the orthographic judgment task in Experiment 2 (24 items)

Model and predictors	Estimate	SE	95% CI		z value	p value
			LL	UL		
<i>picture designation task</i>						
Intercept	1.116	0.309	0.649	1.919	0.396	.692
Learning condition	2.254	0.331	1.691	3.004	5.547	<.001
Grade	2.033	0.441	1.329	3.111	3.269	.001
Model = glmer (forced_choice\$Accuracy~ Learning_condition + Grade +(1 participant) + (1 item)+(1 school),data=forced_choice,family=binomial(link=logit),control=glmerControl(optimizer="bobyqa")						
<i>Orthographic judgment task</i>						
Intercept	0.427	0.070	0.310	0.588	-5.225	<.001
Learning condition	3.855	0.461	3.051	4.872	11.296	<.001
Grade	1.424	0.203	1.077	1.883	2.480	.01
Model = glmer (forced_choice\$Accuracy~ Learning_condition + Grade +(1 participant) + (1 item)+(1 school),data=forced_choice,family=binomial(link=logit),control=glmerControl(optimizer="bobyqa")						

Note. CI, confidence interval; LL, lower limit; UL, upper limit.

**Table 9**

Summary of the mixed linear regression model for variables predicting discrimination between words and close phonological distractors in the spoken recognition task in the analysis including immediate and delayed sessions in Experiment 2 (24 items)

Model and predictors	Estimate	SE	t value	p value
<i>Immediate session vs. delayed session</i>				
Intercept	0.941	0.104	9.028	<.001
Learning condition	0.445	0.116	3.835	<.001
Grade	0.338	0.116	2.917	.004
Model=lmer(go_nogo\$Discrimination_score ~ Learning_condition + Grade + (1 participant),data=go_nogo).				

**Table 10**

Summary of the pairwise comparisons conducted on the pattern of committed errors in the orthographic judgment task for Experiment 2 (24 items)

	$\psi$	SE	Scheffé test	p
<i>Close orthographic distractor</i>				
Non-orthographic condition vs. orthographic condition	-.197	.019	-10.26	<.001
Third-graders vs. fifth-graders	-.069	.018	-3.80	<.001
Orthographic condition: Third-graders vs. fifth-graders	-.087	.032	-2.69	.03
Non-orthographic condition: Third-graders vs. fifth-graders	-.079	.021	-3.73	.001
<i>Distant orthographic distractor</i>				
Non-orthographic condition vs. orthographic condition	.074	.014	5.17	<.001
Third-graders vs. Fifth-graders	.010	.014	0.68	.79
Orthographic condition: Third-graders vs. fifth-graders	.017	.022	0.79	.73
Non-orthographic condition: Third-graders vs. fifth-graders	.013	.018	0.72	.77
<i>Phonological distractor</i>				
Non-orthographic condition vs. orthographic condition	.123	.019	6.55	<.001
Third-graders vs. Fifth-graders	.059	.018	3.26	.005
Orthographic condition: Third-graders vs. fifth-graders	.069	.030	2.29	.07
Non-orthographic condition: Third-graders vs. fifth-graders	.066	.022	2.96	.01

a significant main effect of learning condition, with an odds of accurate written word recognition 3.86 times higher in the orthographic learning condition than in the non-orthographic one (65.70% vs. 35.02%,  $p < .001$ ). The main effect of grade was also significant, with an odds of accurate spelling 1.42 times higher in fifth-graders than in third-graders (52.54% vs 44.55%,  $p = .01$ ). See Table 10 for a summary of the model.



*Pattern of errors.* Statistical analyses were conducted following the same procedure as in Experiment 1. The statistical analyses showed that the pattern of errors was comparable to that of Experiment 1. Indeed,  $2 \times 3$  chi-square homogeneity tests revealed that errors were not evenly distributed across learning conditions,  $\chi^2(2, 2138) = 90.32, p < .001$ , Cramer's  $V = .21$ , and across grade,  $\chi^2(2, 2138) = 15.21, p = .001$ , Cramer's  $V = .07$ ). Post hoc pairwise comparisons in the event of heterogeneous distribution of committed errors showed two results. First, it yielded a larger proportion of close orthographic distractor errors in the orthographic learning condition compared with the non-orthographic one (53.4% vs. 33.7%),  $\chi^2(1, 1183) = 105.18, p < .001$ , but it yielded a lower proportion for distant errors (13.5% vs. 20.9%),  $\chi^2(1, 543) = 26.78, p < .001$ , and for phonological errors (33.1% vs. 45.4%),  $\chi^2(1, 1215) = 42.86, p < .001$ . Second, we found a larger proportion of close orthographic distractors (44.0% vs. 37.1%),  $\chi^2(1, 1183) = 14.47, p = .001$ , and a lower proportion of phonological distractors (38.1% vs. 44.0%),  $\chi^2(1, 1215) = 10.64, p = .004$ , in fifth-graders compared with third-graders, but we found no differences for distant orthographic errors,  $\chi^2(1, 543) < 1, p = .79, ns$ . The details of the distribution of distractors are presented in [Appendix 9](#) in the [supplementary material](#).

## General discussion

Identifying the nature of orthographic facilitation in vocabulary learning is a hotly discussed issue. In studies conducted in children and in an L1, several researchers have argued that the benefit of spelling exposure may reflect the strong interplay between orthography and phonology (see orthographic mapping theory in [Ehri, 2014](#)) owing to the extensive experience in linking the spoken to the written modality with increasing reading skills. In an L2, adult studies show that providing orthographic information helps in memorizing more words, even though learning *without* orthography helps L2 learners to pay more attention to phonemes and to pronounce new words more accurately ([Welby et al., 2022](#)). Information regarding L2 learners who are developing readers is lacking, especially about how reading experience influences the contribution of orthography.

In two experiments involving 16 and 24 items to be learned respectively, we explored whether orthography contributes to the encoding and memorization of L2 vocabulary, as reported in L1 word and pseudoword learning in adults and children (see [Colenbrander et al., 2019](#)). We tested the impact of being exposed or not to orthography in L2 word learning in three tasks reflecting three aspects of lexical learning: a picture designation task to assess phonology-to-semantic pathway learning, a spoken word recognition task to assess phonological learning, and an orthographic choice task for orthographic learning. We also evaluated whether this orthographic facilitation was modulated by the degree of reading experience in the L1 by comparing third-graders and fifth-graders. Overall, orthographic facilitation was observed in all three lexical measures and in both experiments, with no interaction with grade.

Our first task was a picture designation task designed to reflect the phonology-to-semantic pathway. Our results were clear: The presence of orthography was beneficial at each time point (immediately and 1 week later), each grade level (Grade 3 and Grade 5), and whatever the number of items to be learned (16 vs. 24). Therefore, we posit that orthography is beneficial in learning the link between L2 phonological words and their meanings. Given that the picture designation task did not involve any orthographic information, the benefit of exposure to a written form during vocabulary learning may be explained by a connection between orthographic and phonological information, leading to the reinforcement of the lexical phonological form itself and/or the phonology-to-semantics link. This interpretation is supported by the results of the spoken word recognition task, designed to assess the quality of phonological representations per se by asking participants to discriminate learned words from phonemically close distractors. Orthographic facilitation was found again in both experiments, with higher discrimination scores when children were exposed to orthography during learning than when they were not. Such a pattern of results is consistent with previous studies conducted in L1 (pseudo)word learning ([Jubenville et al., 2014](#); [Miles et al., 2016](#); [Ricketts et al., 2009, 2021](#); [Salins et al., 2023](#); see [Colenbrander et al., 2019](#), for a review). However, there are major differences between word learning in an L1 and an L2 even when L1 studies use pseudowords instead of words (e.g. [Ricketts et al., 2009](#)). Indeed, children must learn new meanings in L1 studies, which was not the case

in the current study. Importantly, orthography-to-phonology mappings are less familiar in an L2 and sometimes are even unknown, as in the current experiments. In adults, previous studies already found an advantage of orthography in spoken word recognition, although it can also have an interfering effect on pronunciation and L2 phoneme discrimination (Hayes-Harb & Barrios, 2021; Welby et al., 2022). In some L2 word-learning experiments conducted in adults, participants needed to learn pseudowords (involving learning new meanings) and were familiar with the orthography of the L2 (e.g., Welby et al., 2022). The current word-learning study examined developing readers exposed to a completely unknown L2. Thus, the question of the existence of orthographic facilitation in vocabulary learning was essential given that orthographic processing is not fully automatized in children, especially before Grade 4. We could have expected a stronger effect in older children given that fifth-graders have more automatic reading processes than third-graders (Byrnes & Wasik, 2009). Critically, these children showed the benefit of orthography without a significant interaction with age in both experiments.

The orthographic task was designed to evaluate how children memorized L2 word spellings through a recognition task to avoid the floor effects potentially observed in production. This task also made it possible to identify the type of errors made and to understand how orthographic information was encoded. As expected, children who were exposed to orthography performed a more accurate recognition of the novel written words than those who were not. We designed three categories of distractors (e.g., for the German word “Berg” [mountain]: (a) close orthographic distractors characterized by a large form overlap with the target word [Breg], (b) distant distractors with a restricted form overlap [Bimp], and (c) phonological distractors that were homophonic with the target form when stemming on L1 orthography-to-phonology mappings [Bergue]). Thus, the latter reflected the prediction of the written form a child may have when exposed to the spoken form, based on L1 French orthography-to-phonology mappings, in line with the orthographic skeleton hypothesis (Wegener et al., 2018). Interestingly, irrespective of the degree of reading experience, participants tended to commit more close orthographic errors (Breg for Berg) than the other two types of errors when they were exposed to orthography during learning. However, they preferred the homophonic distractors when they were not exposed to orthography during learning (Bergue for Berg). In the absence of written information, thus, children seemed to rely on L1 orthography-to-phonology mappings to build an orthographic representation. Thus, they were able to learn novel written forms after a rather limited exposure to orthography during learning. Although previous studies conducted in an L1 strongly evidenced how fast L1 orthography is learned while reading (e.g., for orthographic choice: Lucas & Norbury, 2014; for production: Miles et al., 2016), our study extends this observation for L2 word learning.

Two grades were tested to examine, for the first time, the effect of L1 reading experience in orthographic facilitation during L2 vocabulary learning. We tested third-graders and fifth-graders because there is a consensus that reading automatization occurs around Grade 4 (Byrnes & Wasik, 2009). Two views may be theoretically contrasted. On the one hand, older children might display less orthographic facilitation due to greater cross-language interference when exposed to the orthographic forms of the novel words to be learned given that they have more automatized orthography-to-phonology mappings in the L1. On the other hand, they may extract and rely on this new orthographic information more quickly and show greater benefit from orthographic exposure given that they are more familiar with written word processing, and that there is still some partial overlap across languages beyond incongruencies. Note that French third-graders and fifth-graders were exposed to some English words at school. Because older children are more familiar with English orthography than younger children, it could help them to deal with navigation among incongruencies even when faced with a new foreign language like German. Our data cannot help to disentangle these aspects given that no interaction with grade was significant. Although older children systematically outperformed younger children in all three tasks, the benefit of orthography was comparable across grades in both experiments and in all tasks. The similar contribution of orthography in both the picture designation and spoken word recognition tasks, where foils were close phonological distractors, contrasts with the strong grade effect observed in the spelling recognition task, especially in children exposed to orthography. Fifth-graders recognized about 60% of correct spellings, whereas third-graders recognized only 35%. This means that older children have developed stronger skills in orthographic memorization. Thus, although fifth-graders processed written words more accurately, they showed no greater ortho-

graphic facilitation effect in the phonological and semantic tasks than third-graders. Both groups preferred close orthographic foils. Overall, this indicates that younger children relied on partial orthographic encoding to reinforce their phonological representations of L2 words. The more experienced the children, the more letters contributed to reinforcing their phonological representations, leading to higher scores. However, the processes cannot be considered as qualitatively different. Interestingly, the [Salins et al \(2023\)](#) study, which was designed to test for age effects, also evidenced a generalized orthographic facilitation in the L1, not differing across grade levels in the picture naming task. Note, however, that an age effect was found in the picture-word matching task, where the orthographic facilitation was limited to the youngest groups (from Grade 1 to Grade 4). Additional studies, probably using more fine-grained measures—such as the graded measures used by [Ricketts et al. \(2021\)](#)—will be needed to further explore the impact of reading experience on orthographic facilitation in a foreign language.

We tested children both immediately and 1 week later in the two experiments, allowing to test retention. The pattern of results never differed between both testing sessions; the three lexical dimensions involved in vocabulary learning did not undergo any decay after 1 week. Several studies conducted in an L1 also tested retention after 1 week (e.g., [Chambre et al., 2017, 2020](#); [Ricketts et al., 2021](#); [Valentini et al., 2018](#)). All observed that word knowledge was retained over time, even after 8 months ([Ricketts et al., 2021](#)). Of course, one cannot rule out that the children were trained during the week, although this is rather unlikely given that German words are not familiar to the general French population. In addition, children did not mention any training when asked. In addition, we found no interaction between learning condition and testing session, suggesting a comparable orthographic effect in both sessions. This result runs counter to our hypothesis of an enhanced orthographic facilitation effect after 1 week due to the strengthening of connections among orthography, phonology, and semantic information over time. Future studies using additional tasks capturing lexical engagement processes may be needed to examine the effect of time on lexical consolidation in more detail. Note that we found faster response times in Experiment 1 in all groups, suggesting a kind of consolidation, which was not observed, however, in Experiment 2.

We ran two experiments in the current study because it was difficult to anticipate the number of German words that could be retained, providing us with the opportunity to replicate and guarantee robustness. The results were remarkably consistent whether children needed to learn 16 words (Experiment 1) or 24 words (Experiment 2), thereby allowing us to confidently discuss the nature of orthographic facilitation in L2 word learning. This also shows that the increasing cognitive load during memorization did not modulate the orthographic facilitation effect. However, the percentage of correct responses (in the three tasks) was similar in the two experiments regardless of the number of items to be learned. This suggests that the *quantity* of words learned was not constant, which is quite hard to explain. Increasing the number of words should increase the memory load, leading to lower recognition scores. On the other hand, it may provide more opportunities to become familiar with the phonemes of this new L2 and, for the children exposed to orthographic information, to reinforce orthography-to-phonology mappings in the L2. Again, this pioneering study in an L2 learning context should be followed by others to answer questions raised by the results, including how learners might adapt their strategies to optimize learning.

The congruency issue should also be examined in greater depth, that is, the possible impact of cross-language (in)congruency of print-to-sound mappings. It is hard to adopt an “all-or-none” approach to this question given that phonemes do differ across languages; no word can be considered as purely congruent. However, orthography-to-phonology mappings between L1 and L2 can be considered very close (e.g., *hat* in English and German) or distant (e.g., *brave* could be read out as /BraFe/ in German). Interestingly, two very recent studies reported contradictory results in English as L2 learning. [Krepel et al. \(2021\)](#) found that the benefit of orthographic exposure in L2 learning, assessed through a translation task, was significant for congruent words but not for incongruent words. A different pattern was observed by [Welby et al. \(2022\)](#) with English-like pseudowords. They reported an orthographic facilitation effect in picture naming and matching tasks (production and recognition tasks, respectively) for words whose vowels had incongruent grapheme-to-phoneme mappings across languages. However, orthographic exposure did pull the vowels toward more L1-like pronunciations. In our experiments, the stimuli included both congruent and incongruent L2 words, matched on

length and frequency of L1 translation equivalents. Although it was not possible to make reliable comparisons owing to the low number of items per condition, examination of the descriptive statistics revealed that mean accuracy was very close for both types of items.

Finally, the languages at stake should be considered with regard to the facilitation effect, especially the impact of the L1 in L2 word learning. Certainly, it should be different to learn L2 words in a consistent or inconsistent orthography. In addition, children's L1 should also have an impact on how they can rely on orthographic information. This should be tested in subsequent studies. Here, children whose L1 orthography (French) is relatively inconsistent learned a new language whose orthography is rather consistent (German), thereby limiting orthographic learning difficulties to the problem of cross-language *incongruency* without any within-language inconsistency issue.

Our study has some limitations. First, we used a between-participants design, leading to the possibility that interindividual differences between groups also account for differences in the orthographic facilitation effect. Although studies conducted in an L1 generally have used a within-participants design (but see [Jubenville et al., 2014](#), for a between-participants design), we took the opportunity of the children's unfamiliarity with the German orthography to keep the non-orthographic group completely outside of this knowledge. Therefore, orthographic expectancies that could have interfered with orthographic facilitation ([Salins et al., 2023](#)) were very limited. However, having used a large battery of cognitive and language tasks allowing us to compare the groups (and to statistically control variables if necessary), we can be reassured that the orthographic facilitation effect observed is not due to more general differences between the children. Another limitation is that our study did not include a power analysis, which was not possible to compute before conducting the current experiments due to the lack of studies on this topic in children learning an L2. However, it would have been beneficial to compute power analysis after the first experiment if the two studies had been carried out successively, which unfortunately was not the case.

An important caveat, suggested by a reviewer, is that our measures of word learning cannot address one key dimension of vocabulary learning, namely the ability to use words. Because children learned only a few words (16 or 24) from an unknown language, it was inherently not possible to test this. Assessments in more ecological conditions should be conducted to answer this point. Along the same line, in our word recognition task distractors were phonologically close to the target words, meaning that participants needed to have built a very precise spoken representation of novel L2 words to succeed in the task. Therefore, our tasks were not designed to examine partial word learning. In a recent study, [Ricketts et al \(2021\)](#) used graded measures of word learning within a dynamic assessment. Children were first asked to provide a definition; if they were not able to do so, a semantic cue was provided. Finally, if they failed at this step, they were asked to select the correct definition among four definitions. In our study, a more fine-grained approach—including both close and distant distractors, for example—could have revealed a different pattern of results (e.g., effects of grade and delay). Finally, during the learning phase, written words were presented for only 3 s. This duration may have been too short for the youngest children and may have contributed to their worse performance. Future studies may explore the duration of written word presentation in more detail. Note that the number of times a word needs to be repeated for successful learning should also be considered.

In conclusion, the use of a paired-associate word-learning paradigm promoted the acquisition and memorization of a large set of L2 German words among French native children from two grade levels. Learning performance was high and robust to time decay, with a systematic advantage for children exposed to orthographic forms of L2 novel words over children without such information. Therefore, our study concurs with previous studies conducted in an L1 in spite of the differences between L1 word learning and L2 word learning. Indeed, semantic information is already present in L2 word learning. Conversely, new phonemes must be learned in the L2 as well as new orthography-to-phonology mappings. Our results can be explained by the strong interplay between orthographic and phonological processing: phonological representations are quickly and automatically activated upon the presentation of a written word. Conversely, spoken language activates their written correspondents, in line with the lexical quality hypothesis ([Perfetti & Hart, 2002](#)). Given these promising results, it is essential to provide evidence of their possible transfer to a classroom setting. It seems essential to document when and how much written information should be provided when children learn a foreign language at school. Future experimental studies should probe how L2 orthography is processed during

vocabulary learning, including its impact on phonological and semantic learning, taking into account the psycholinguistic properties of novel words such as orthographic features and print-to-sound congruency.

### CRedit authorship contribution statement

**Florian Salomé:** Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Eva Commissaire:** Writing – review & editing, Validation, Formal analysis, Conceptualization. **Séverine Casalis:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

### Data availability

Data will be made available on request.

### Acknowledgment

This study was funded by grants from the Agence Nationale de la Recherche (ANR) APPREL2 project awarded to Séverine Casalis (16-CE28-0009).

### Appendix A. Supplementary material

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.jecp.2024.105978>.

### References

- 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(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bartolotti, J., & Marian, V. (2017). Bilinguals' existing languages benefit vocabulary learning in a third language. *Language Learning*, 67(1), 110–140. <https://doi.org/10.1111/lang.12200>.
- Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1835–1842. <https://doi.org/10.1037/xlm0000417>.
- Bassetti, B., Sokolovi-Perovi, M., Mairano, P., & Cerni, T. (2018). Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs. *Language and Speech*, 61(4), 577–597. <https://doi.org/10.1177/0023830918780141>.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.
- Bonin, P., Méot, A., & Bugaiska, A. (2018). Concreteness norms for 1,659 French words: Relationships with other psycholinguistic variables and word recognition times. *Behavior Research Methods*, 50(6), 2366–2387. <https://doi.org/10.3758/s13428-018-1014-y>.
- Botezatu, M. R. (2023). The impact of L1 orthographic depth and L2 proficiency on mapping orthography to phonology in L2-English: An ERP investigation. *Applied Psycholinguistics*, 44(2), 237–263. <https://doi.org/10.1017/S0142716423000176>.
- Brysaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424. <https://doi.org/10.1027/1618-3169/a000123>.
- Bürki, A., Welby, P., Clément, M., & Spinelli, E. (2019). Orthography and second language word learning: Moving beyond “friend or foe?”. *Journal of the Acoustical Society of America*, 145(4), EL265–EL271. <https://doi.org/10.1121/1.5094923>.
- Byrnes, J., & Wasik, B. A. (2009). Picture this: Using photography as a learning tool in early childhood classrooms. *Childhood Education*, 85(4), 243–248. <https://doi.org/10.1080/00094056.2009.10523090>.
- Cerni, T., Bassetti, B., & Masterson, J. (2019). Effects of orthographic forms on the acquisition of novel spoken words in a second language. *Frontiers in Communication*, 4, 31. <https://doi.org/10.3389/fcomm.2019.00031>.
- Chambré, S. J., Ehri, L. C., & Ness, M. (2017). Orthographic facilitation of first graders' vocabulary learning: Does directing attention to print enhance the effect? *Reading and Writing*, 30(5), 1137–1156. <https://doi.org/10.1007/s11145-016-9715-z>.
- Chambré, S. J., Ehri, L. C., & Ness, M. (2020). Phonological decoding enhances orthographic facilitation of vocabulary learning in first graders. *Reading and Writing*, 33(5), 1133–1162. <https://doi.org/10.1007/s11145-019-09997-w>.
- Colenbrander, D., Miles, K. P., & Ricketts, J. (2019). To see or not to see: How does seeing spellings support vocabulary learning? *Language, Speech, and Hearing Services in Schools*, 50(4), 609–628. [https://doi.org/10.1044/2019\\_LSHSS-VOIA-18-0135](https://doi.org/10.1044/2019_LSHSS-VOIA-18-0135).
- Comesaña, M., Perea, M., Piñero, A., & Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Child Psychology*, 104(1), 22–33. <https://doi.org/10.1016/j.jecp.2008.10.004>.



- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure: How do word-related variables and proficiency influence receptive vocabulary learning? *Language Learning*, 70(2), 349–381. <https://doi.org/10.1111/lang.12380>.
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 71(4), 808–816. <https://doi.org/10.1080/17470218.2017.1310261>.
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, 18(1), 5–21. <https://doi.org/10.1080/1088438.2013.819356>.
- Ehri, L. C. (2020). The science of learning to read words: A case for systematic phonics instruction. *Reading Research Quarterly*, 55, S45–S60. <https://doi.org/10.1002/rrq.334>.
- Ehri, L. C., & Wilce, L. S. (1979). The mnemonic value of orthography among beginning readers. *Journal of Educational Psychology*, 71(1), 26–40. <https://doi.org/10.1037/0022-0663.71.1.26>.
- Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, 30(5), 506–528.
- Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, 36(1), 7–22. <https://doi.org/10.1017/S014271641400040X>.
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36(2), 345–360. <https://doi.org/10.1016/j.wocn.2007.11.002>.
- Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, 17(2), 384–395. <https://doi.org/10.1017/S1366728913000436>.
- Escudero, P., Smit, E. A., & Angwin, A. J. (2023). Investigating Orthographic Versus Auditory Cross-Situational Word Learning With Online and Laboratory-Based Testing. *Language Learning*, 73(2), 543–577.
- Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, 53(3), 343–365. <https://doi.org/10.1177/0023830910371447>.
- Franke, T. M., Ho, T., & Christie, C. A. (2011). The chi-square test: Often used and more often misinterpreted. *American Journal of Evaluation*, 33(3), 448–458. <https://doi.org/10.1177/1098214011426594>.
- Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Annals of Mathematical Statistics*, 34(1), 152–177.
- 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. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (Vol. 52, pp. 267–283). John Benjamins.
- Gupta, P., MacWhinney, B., Feldman, H. M., & Sacco, K. (2003). Phonological memory and vocabulary learning in children with focal lesions. *Brain and Language*, 87(2), 241–252. [https://doi.org/10.1016/S0093-934X\(03\)00094-4](https://doi.org/10.1016/S0093-934X(03)00094-4).
- Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, 54(3), 297–326. <https://doi.org/10.1017/S0261444820000658>.
- 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(3), 367–381. <https://doi.org/10.1177/0023830910371460>.
- Hu, G. (2008). The misleading academic discourse on Chinese-English bilingual education in China. *Review of Educational Research*, 78(2), 195–231. <https://doi.org/10.3102/0034654307313406>.
- Jubenville, K., Sénéchal, M., & Malette, M. (2014). The moderating effect of orthographic consistency on oral vocabulary learning in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 126, 245–263. <https://doi.org/10.1016/j.jecp.2014.05.002>.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633. <https://doi.org/10.3758/BRM.42.3.627>.
- Korkman, M., Kirk, U., & Kemp, S. (2012). *NEPSY II—Bilan Neuropsychologique de l'Enfant* (2nd ed.). Éditions du Centre de Psychologie Appliquée (ECPA).
- Krepel, A., de Bree, E. H., & de Jong, P. F. (2021). Does the availability of orthography support L2 word learning? *Reading and Writing*, 34(2), 467–496. <https://doi.org/10.1007/s11145-020-10078-6>.
- Lefavrais, P. (1967). *Test de L'Alouette, Manuel et tests*. Éditions du Centre de Psychologie Appliquée (ECPA).
- Lefavrais, P. (2005). *Alouette-R: Test d'analyse de la vitesse en lecture à partir d'un texte*. Éditions du Centre de Psychologie Appliquée (ECPA).
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36(1), 156–166. <https://doi.org/10.3758/BF03195560>.
- Li, H., Zhang, J., Ehri, L., Chen, Y., Ruan, X., & Dong, Q. (2016). The role of orthography in oral vocabulary learning in Chinese children. *Reading and Writing*, 29(7), 1363–1381. <https://doi.org/10.1007/s11145-016-9641-0>.
- Lucas, R., & Norbury, C. F. (2014). Orthography facilitates vocabulary learning for children with autism spectrum disorders (ASD). *Quarterly Journal of Experimental Psychology*, 67(7), 1317–1334. <https://doi.org/10.1080/17470218.2013.859714>.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49, 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98(1), 185–199. <https://doi.org/10.1037/0033-2909.98.1.185>.
- Major, R. C. (1987). Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition*, 9(1), 63–82.
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One*, 7(8)e43230. <https://doi.org/10.1371/journal.pone.0043230>.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>.



- Miles, K. P., Ehri, L. C., & Lauterbach, M. D. (2016). Mnemonic value of orthography for vocabulary learning in monolinguals and language minority English-speaking college students. *Journal of College Reading and Learning*, 46(2), 99–112. <https://doi.org/10.1080/10790195.2015.1125818>.
- Ministère de l'Éducation Nationale. (2019). *Guide pour l'enseignement des langues vivantes Oser les langues vivantes étrangères à l'école*. <https://eduscol.education.fr/document/347/download>.
- Ministère de l'Éducation Nationale (2023). *L'organisation des enseignements au collège*. <https://eduscol.education.fr/619/l-organisation-des-enseignements-au-college>.
- Moret-Tatay, C., & Perea, M. (2011). Is the go/no-go lexical decision task preferable to the yes/no task with developing readers? *Journal of Experimental Child Psychology*, 110(1), 125–132. <https://doi.org/10.1016/j.jecp.2011.04.005>.
- Pérez-Serrano, M., Nogueroles-López, M., & Duñabeitia, J. A. (2021). Incidental vocabulary learning with subtitles in a new language: Orthographic markedness and number of exposures. *PLoS One*, 16(2)e246933. <https://doi.org/10.1371/journal.pone.0246933>.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11, pp. 67–86). John Benjamins. <https://doi.org/10.1075/swll.11.14per>.
- Poarch, G. J., Van Hell, J. G., & Kroll, J. F. (2015). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism: Language and Cognition*, 18, 357–371. <https://doi.org/10.1017/S1366728914000558>.
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Raven, J. C. (1981). *Progressive matrices*. Etablissement d'Applications Psychotechniques (EAP).
- Ricketts, J., Bishop, D. V., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *Quarterly Journal of Experimental Psychology*, 62(10), 1948–1966. <https://doi.org/10.1080/17470210802696104>.
- Ricketts, J., Dawson, N., & Davies, R. (2021). The hidden depths of new word knowledge: Using graded measures of orthographic and semantic learning to measure vocabulary acquisition. *Learning and Instruction*, 74, 101468. <https://doi.org/10.1016/j.learninstruc.2021.101468>.
- Ricketts, J., Dockrell, J. E., Patel, N., Charman, T., & Lindsay, G. (2015). Do children with specific language impairment and autism spectrum disorders benefit from the presence of orthography when learning new spoken words? *Journal of Experimental Child Psychology*, 134, 43–61.
- Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology*, 100(1), 175–191. <https://doi.org/10.1037/0022-0663.100.1.175>.
- Salins, A., Cupples, L., Leigh, G., & Castles, A. (2023). Orthographic facilitation of oral vocabulary acquisition in primary school children. *Quarterly Journal of Experimental Psychology*, 76(5), 1045–1056. <https://doi.org/10.1177/17470218221102916>.
- Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, 51(3), 1258–1270. <https://doi.org/10.3758/s13428-018-1099-3>.
- 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](https://doi.org/10.1016/0010-0277(94)00645-2).
- Showalter, C. E. (2018). Impact of Cyrillic on native English speakers' phono-lexical acquisition of Russian. *Language and Speech*, 61(4), 565–576. <https://doi.org/10.1177/0023830918761489>.
- Showalter, C. E., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, 36(1), 23–42. <https://doi.org/10.1017/S0142716414000411>.
- Tapia, J. L., Rocabado, F., Martínez, E. M. R., Martínez, M. V., & Perea, M. (2021). The facilitating effect of narrator variability in incidental word learning: Where did I hear it? In *5th International Neurobiology of Speech and Language Conference* (pp. 40–41).
- Valentini, A., Ricketts, J., Pye, R. E., & Houston-Price, C. (2018). Listening while reading promotes word learning from stories. *Journal of Experimental Child Psychology*, 167, 10–31. <https://doi.org/10.1016/j.jecp.2017.09.022>.
- Wechsler, D. (2005). *Échelle d'intelligence de Wechsler pour enfants: WISC-IV*. Éditions du Centre de Psychologie Appliquée (ECPA).
- Wegener, S., Wang, H. C., de Lissa, P., Robidoux, S., Nation, K., & Castles, A. (2018). Children reading spoken words: Interactions between vocabulary and orthographic expectancy. *Developmental Science*, 21(3)e12577. <https://doi.org/10.1111/desc.12577>.
- Welby, P., Spinelli, E., & Bürki, A. (2022). Spelling provides a precise (but sometimes misplaced) phonological target. Orthography and acoustic variability in second language word learning. *Journal of Phonetics*, 94, 101172. <https://doi.org/10.1016/j.wocn.2022.101172>.