RESEARCH ARTICLE

Learning words with unfamiliar orthography: The role of cognitive abilities

Marie-Josée Bisson*

De Montfort University, Leicester, UK
*Corresponding author. E-mail: marie-josee.bisson@dmu.ac.uk

Abstract

Research suggests new foreign language (FL) words are learned more easily if their phonology follows the phonotactic rules of the native language. Very little is known, however, about the impact of orthography on FL learning. This study investigated the cognitive mechanisms supporting the learning of words with familiar and unfamiliar orthographies. Participants took part in learning and meaning recall tasks, as well as a series of cognitive tasks (short-term and working memory tasks and tasks assessing their phonological and acoustic abilities). Orthographic and phonological familiarity judgments were collected using another sample of participants. Using a mixed-effects model, the results showed that orthographic familiarity impacted FL word learning even after controlling for phonological familiarity. However, there were no interactions with cognitive abilities.

Introduction

Not all words are created equal; some words are easier to learn than others and this can be due to a number of features such as concreteness, word type, frequency, cognate status, length, and phonology (De Groot & Keijzer, 2000; Lotto & De Groot, 1998; Morra & Camba, 2009; Vidal, 2011). For example, new foreign language (FL[1]) words are learned more easily if their phonology follows the phonotactic rules of the native language (Ellis & Beaton, 1993; Kaushanskaya et al., 2011; Morra & Camba, 2009). Another important word-related feature that can impact learning is cross-linguistic orthographic similarity, or how similar FL words look compared to native language words (Bartolotti & Marian, 2017; Bordag et al., 2016; Ellis & Beaton, 1993). What are the mechanisms then that support learning of words with familiar and unfamiliar orthographies? Prior research has highlighted various cognitive mechanisms with a role in FL word learning, for example, short-term and/or working memory (Bisson et al., 2021; Martin & Ellis, 2012; Morra & Camba, 2009), as well as phonological abilities

---

[1]In this article the term "foreign language" is used to indicate a language other than the native language. It therefore could be used interchangeably with second language, additional language, etc.

(Bisson et al., 2021; Hu, 2012; Morra & Camba, 2009; Vijayachandra, 2007). The following section highlights the current findings regarding the role of orthography in FL word learning. Then, the cognitive mechanisms supporting word learning are reviewed. The final section introduces the current study that investigates the interaction between cross-linguistic orthographic similarity (word-related feature) and cognitive abilities (individual differences).

The role of orthography in FL learning

Research on the impact of orthography on FL word learning has not been extensive (Simon & van Herreweghe, 2010); however, several studies suggest it is an important factor (Bartolotti & Marian, 2017; Bordag et al., 2016; Ellis & Beaton, 1993). The cross-linguistic orthographic regularity of FL words is typically measured through lexical variables (e.g., n-gram probabilities, n-gram frequencies, or neighborhood density) or psycholinguistic variables (e.g., judgments/ratings of wordlikeness, typicality, similarity, etc.; Dijkstra et al., 2010; Tokowicz et al., 2002). Therefore, it is a measure of how closely the FL words follow the spelling patterns of the native language, or how similar/typical the FL words look compared to native language words. For example, the Welsh word *llygad* (meaning "eye") would score low on an n-gram probability measure and would be judged dissimilar or atypical to English words, as words in English never start with the bigram *ll*. Ellis and Beaton (1993) used average bigram frequencies and minimum bigram frequency to predict native language to FL and FL to native language translation accuracy. Their measures of cross-linguistic orthographic regularity were not significant predictors in their causal path analyses; however, they found a correlation of .30 between minimum bigram frequency and native language to FL translation. Using a similar classification of cross-linguistic orthographic regularity (neighborhood size, positional segment and bigram frequencies), Bartolotti and Marian (2017) found that orthographically wordlike nonwords were easier to learn during a picture-word paired-associate learning task. Once they controlled for phonological wordlikeness however, their wordlike advantage only remained significant during written word production (as opposed to a meaning recognition task). Another study found some impact of FL orthotactic probabilities on FL word learning (Bordag et al., 2016). Similarly to studies by Ellis and Beaton (1993) and Bartolotti and Marian (2017), part of the target item selection process involved calculating bigram and trigram probabilities, however Bordag et al. (2016) also collected ratings of wordlikeness, and only targets judged either atypical or typical on a scale of 1 to 6 (the end points of a typicality continuum) were included. Target items (German pseudowords) were learned through either an incidental or intentional learning phase where participants either read texts without their knowledge of the word learning aspect of the study or read definitions of novel words for a follow-up test. All participants then completed a self-paced reading task where the reading time on the target items was compared in semantically plausible versus implausible sentences. Results showed orthotactic probabilities only impacted reading times in the intentional learning group. Therefore, orthographic similarity, as measured through lexical or psycholinguistic variables seems to predict different aspects of word learning linked to the recall or processing of the written forms of the FL words in some learning contexts only (i.e., intentional learning).

Cognitive mechanism supporting word learning

As well as word-level predictors, individual differences are known to impact FL word learning (Bisson et al., 2021; Hu, 2012; Kapa & Colombo, 2014; Martin & Ellis, 2012;

Morra & Camba, 2009; Vijayachandra, 2007). In Bisson et al. (2021), a composite score including short-term and working memory was a significant predictor of FL word learning. The tasks for phonological and visuo-spatial short-term memory included a storage element only (digits and red dots on a grid, respectively) whereas the working memory tasks involved both storage and processing (Conway et al., 2005; Martin & Ellis, 2012). These four tasks shared the underlying requirement to encode serial information about the to-be-remembered items and this was proposed as an important aspect of FL word learning. For example, the syllables that compose a FL word must be recalled in the correct order otherwise this may lead to communication or comprehension errors. As well as short-term and working memory, other important language learning skills include phonological abilities (Bisson et al., 2021; Hu, 2012; Vijayachandra, 2007). In children this is sometimes referred to as phonological awareness or sensitivity and measured using tasks such as rhyme and alliteration awareness, phoneme and vowel deletion as well as syllable and rhyme discrimination (Hu, 2012; Morra & Camba, 2009; Vijayachandra, 2007). It is thought to be important for the perception, manipulation and encoding of speech and nonspeech sounds (Melby-Lervåg et al., 2012; Morra & Camba, 2009; Silbert et al., 2015). Although prior research has shown that phonological abilities are important for children language learners, Bisson et al. (2021) expanded this to adult language learners using phonological and acoustic discrimination tasks. These tasks focused on fine-tuned discrimination abilities (Lengeris & Hazan, 2010; Silbert et al., 2015) and showed that being able to encode precise phonological representations is an integral part of FL learning.

Cognitive mechanisms and orthography

As discussed so far, both individual differences and word-level variables are important predictors of language learning. Much prior research has investigated these in isolation, but interactions between the two may be important to consider (Housen & Simoens, 2016). In their taxonomy of second language learning difficulty, Housen and Simoens (2016) explain that feature-related difficulties, context-related difficulties, and learner-related difficulties are all important to study for the field of language learning to evolve, but this will be restricted unless we also investigate how they interact. Recent research showed an interaction between context-related difficulty (learning conditions) and learner-related difficulty (working/short-term memory; Bisson et al., 2021), but there is no prior research to our knowledge investigating the interaction between feature-related and learner-related difficulties.

The current study

The aims of the current study are therefore to (1) expand the findings on the role of orthography in FL learning and (2) to investigate the interaction between orthography (feature-related variable) and cognitive abilities (memory and phonological abilities; learner-related variables). To classify the FL word's orthography, judgments of familiarity, similarly to the typicality ratings used in Bordag et al. (2016; but see also see Dijkstra et al., 2010; Tokowicz et al., 2002) were collected. A group of native English speakers judged pairs of FL words as to which one was more similar to native language words (the comparative judgment approach; Bisson et al., 2016, 2020; Jones et al., 2019). The word's phonology was controlled for (as in Bartolotti & Marian, 2017) through judgments of phonological familiarity. The raw data for this study was obtained from

Bisson et al. (2021). For individual differences the study focused on the cognitive abilities found to be significant predictors in Bisson et al. (2021), which included a "memory" composite score (short-term and working memory) and a "phonological abilities" composite score (phonological and acoustic abilities).[2] Because individual differences in vocabulary knowledge are linked to better language learning and processing skills, this was controlled in the current study (Bisson et al., 2021; Mainz et al., 2017; Morra & Camba, 2009). Prior research on the role of orthography in word learning measured learning through recall or processing of FL word form. Here prior research was expanded by measuring learning at the level of meaning recall. Based on prior research, it was predicted that orthography would be a significant predictor of meaning recall: The meaning of words with more unfamiliar orthographies would be more difficult to recall. As mentioned earlier, prior research has not addressed the interaction between orthography and cognitive abilities. However, based on the research of Bisson et al. (2021) who found an interaction between cognitive abilities and learning conditions, it was predicted that cognitive abilities would interact with orthography such that participants with better cognitive abilities would be less affected by unfamiliar orthography. In other words, cognitive mechanisms would compensate for the difficulty of the items.

Method

Ethical considerations

Ethical approval was granted by the Faculty of Health and Life Sciences Research Ethical Committee at De Montfort University. All participants gave informed consent to take part in the studies.

Participants

The final participant sample included 132 native English speakers from a UK Midlands University (20 males, M age = 20.29, SD = 3.96). Participants generally reported having some knowledge of FLs and a moderate level of fluency (M number of FLs = 2.39, SD = 1.20; M fluency = 3.59, SD = 1.71 on a scale ranging from 1 = poor to 7 = fluent). One bilingual participant, as well as four participants with prior knowledge of Welsh were excluded. Two participants were excluded for achieving lower than chance on the letter-search task (incidental learning) and one because they reported having cognitive difficulties. A further 12 participants did not complete Session 2 and were therefore excluded, and 7 participants were removed due to technical difficulties during one of the tasks.

Another sample of 16 native English speakers was recruited to provide familiarity data (4 males M age = 21.23). These participants were also students at a UK Midlands University and they had no prior knowledge of Welsh.

Material

Three lists of 40 depictable Welsh words were used for the purpose of the learning and recall tasks (see Stimuli.txt on https://doi.org/10.21253/DMU.14170832.v1). This

---

[2]Bisson et al. (2021) also found orthographic abilities to be a significant predictor; however, in view of the limitations of this task (see the discussion section of Bisson et al., 2021), this predictor is omitted here.

included auditory and written word forms as well as a picture depicting the meaning of the words. Auditory word forms were recorded by a Welsh native speaker, and pictures were selected from Brodeur et al. (2010) or Moreno-Martínez and Montoro (2012). One list of words was used for the incidental learning task and one list was used for the intentional learning task (one list was use for an additional task in Bisson et al., 2021) counterbalanced across participants. All computer tasks were presented using Psy-choPy (Peirce et al., 2019).

Procedure

Participants completed all tasks over two sessions on consecutive days (Session 1: incidental learning followed by intentional learning tasks, language background questionnaire, short-term memory tasks, phonological and acoustic tasks, and Session 2: meaning recall, working memory tasks, native language vocabulary knowledge; see Figure 1). Participants also completed additional tasks not included here (see Bisson et al., 2021).

Learning tasks

Participants completed both an incidental and an intentional learning task. For incidental learning, participants were asked to complete a letter-search task (as in Bisson et al., 2013, 2014, 2015). Participants were not informed about the upcoming vocabulary test and they were not asked to learn the meaning of the words, thus creating
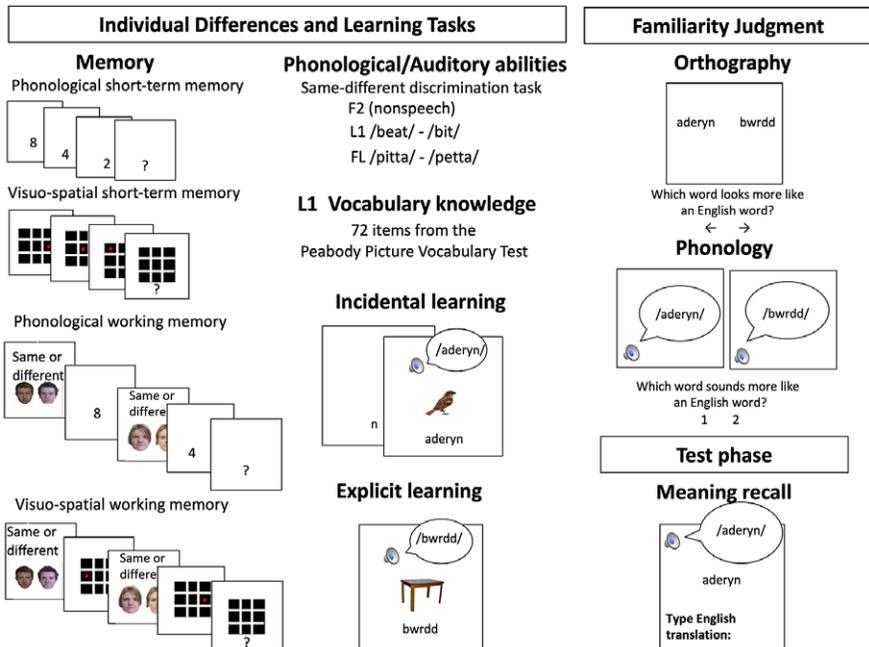
Figure 1. Graphical representation of the individual differences task, learning and recall tasks as well as the familiarity judgment tasks (the latter completed by a separate sample of participants). Pictures and faces taken from Brodeur et al. (2010), Moreno-Martínez and Montoro (2012), and Burton et al. (2010).

the conditions in which incidental learning could occur (see Hulstijn, 2001). On each trial, participants first saw a blank screen for 500 ms followed by a letter presented for 1 second, and finally the written word appeared. Participants were instructed to indicate with a button press whether the letter they saw was present in the written word. Even though these were not necessary for the letter-search task, at the same time as the written word onset, participants were also exposed to the auditory word form and to a picture depicting the meaning of each word. The written word and picture remained onscreen until the end of the trial and the auditory word form was played once. Stimuli presentation duration was set to 3 seconds to allow for encoding of information without being too lengthy. Participants first completed 10 practice trials with feedback before completing three blocks of 80 trials without feedback. Each Welsh word was presented twice in each block, once with a letter that was in the word and once with a letter that was not. The order of the trials in each block was randomized. For the intentional learning task, participants were now specifically instructed to learn the meaning of 40 Welsh words. The stimuli were presented exactly in the same manner as in the incidental learning, except that there was no letter to search and no buttons to press. However, participants were now informed that they would be tested on the words the next day. Each learning task lasted 20 minutes.

Meaning recall task
Participants were presented with each Welsh word from the incidental and intentional learning tasks in random order. They were asked to type the English translation for each word. Participants were encouraged to try to answer each trial, but they could also just press "enter" to proceed to the next trial. Welsh auditory word forms were presented once on each trial, however the written word forms remained on screen until participants completed their answer. There was no time limit to complete this task and on average participants took approximately 5 minutes to complete it.

Short-term/working memory tasks
Participants completed both a verbal and a visuo-spatial short-term and working memory task. For the verbal short-term memory task, participants were presented with digits from 1 to 9 and upon the presentation of a question mark had to recall the digits in the correct order using the keyboard. For the visuo-spatial short-term memory, the to-be-remembered stimuli was a red circle on a 3×3 grid and participants recalled using mouse clicks on an empty grid. For both verbal and visuo-spatial working memory tasks, in between each storage element described in the preceding text, participants had to judge whether pairs of faces were from the same or different people (Hubber, 2015). Spans ranged from three to nine elements presented in random order and participants completed three trials of each span length. For each memory task, the proportion of correctly recalled items in the correct order was calculated on each trial (Conway et al., 2005).

Phonological/acoustic abilities
A same-different A-X discrimination task was used (see Lengeris & Hazan, 2010; Silbert et al., 2015) with FL (/pita/ - /peta/) and native language (/beat/ - /bit/) continuums for phonological abilities as well as a nonspeech continuum (F2) for acoustic abilities. Participants had to indicate on each trial whether two words or sounds (beeps) were the same or different. Stimuli varied in their similarity to the end point of the continuums in

nine steps such that on some trials, stimuli were easier/harder to discriminate. There was one block of 32 trials (16 "same" and 16 "different" trials) for each continuum and the order of trials and blocks was randomized.

Language background

Participants completed 72 trials of the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) and received one point for each correct answer. The most difficult items (157–228) were used. Participants also completed a self-reporting language background questionnaire where they were asked to indicate what FLs they knew and to rate their fluency in each of reading, written, listening, and speaking on a scale of 1 = poor to 7 = fluent (see Bisson et al., 2021).

Familiarity judgments

The comparative judgment approach (Bisson et al., 2016, 2020; Jones et al., 2019) was used to rate the Welsh words from most unfamiliar to most familiar. For orthographic familiarity, participants were shown two of the 120 Welsh words on each trial on the left and right side of the screen (see Figure 1). They were asked to select which one looked more similar to an English word. Following this, participants rated the words for phonological familiarity using the same method. They heard 2 of the 120 items one after the other on each trial and were asked to indicate which one sounded more similar to an English word by pressing "1" or "2." Each participant received a different trial file to ensure that each item was compared against different items across participants. Each participant completed 120 judgments in random order and each item was judged 32 times across participants.

Results

The data and analysis script for this study are available at https://doi.org/10.21253/DMU.14170832.v1.

Orthographic and phonological familiarity judgments

The judgments were highly reliable (Scale Separation Reliability index = .90 and .82 for orthographic and phonological familiarity judgments, respectively; Verhavert et al., 2018). Using the Bradley–Terry statistical model (Firth, 2005) a z-score parameter estimate was calculated for orthographic and phonological familiarity for each item. Figure 2 shows a ranking order from most unfamiliar (low score) to most familiar (high score) item. For example, the three items with the lowest ranking and hence deemed most unfamiliar were "ymennydd," "cyfrifiadur," and "cwmwl" (brain, computer, and cloud, respectively) and the three items with the highest ranking were "crys," "haul," and "clust" (shirt, sun, and ear, respectively).

Cognitive predictors

Table 1 shows the means (*SD*) for the raw scores in each task. A principal component analysis (see Principal_Component_Analysis.docx on https://doi.org/10.21253/DMU.14170832.v1) was conducted to group the cognitive tasks together into a memory composite (visuo-spatial and verbal short-term and working memory tasks) and a
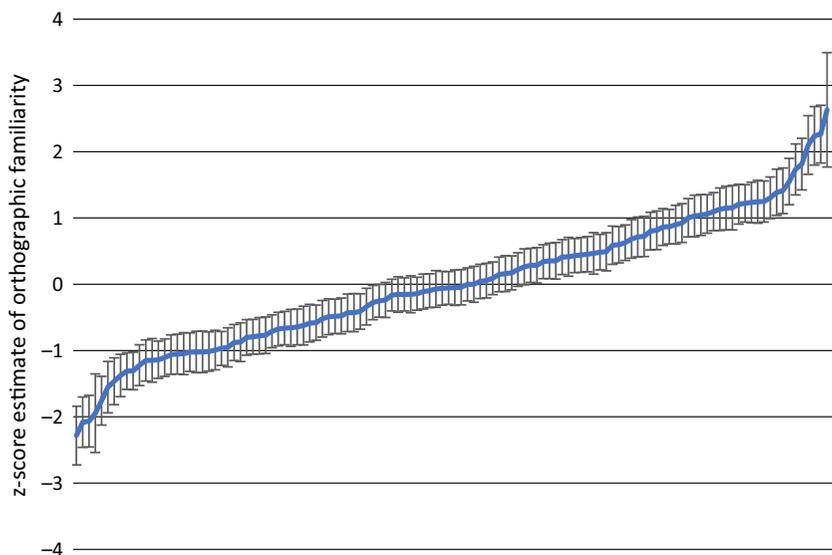
Figure 2. Z-score rank order for the orthographic judgments with standard error of the estimates.

phonological abilities composite score (phonological abilities in L1 and FL and acoustic abilities). A z-score was then calculated for each task and these were averaged together to create the memory and phonological abilities composite scores for the mixed-effects model in the following text. The vocabulary test score loaded on the same factor as the memory and acoustic abilities tasks. This is likely due to the vocabulary test using acoustic abilities and short-term/working memory abilities to process each target word and keep it active while selecting an answer. These were therefore regressed from vocabulary test score and the residuals were used as a measure of vocabulary knowledge for the mixed-effects model mentioned in the following text (see Bisson et al., 2021; Morra & Camba, 2009). Using residuals in an analysis is a contested practice in psycholinguistics (see Wurm & Fisicaro, 2014), therefore a second analysis using the unresidualized vocabulary test score (see Table 4) is also provided, and we come back to this issue in the "Discussion" section.

Learning scores

The raw accuracy scores in the recall task were used for the mixed-effects model using the binomial family. However, by-participant and by-item mean recall accuracies were calculated for the correlation matrices (see Tables 1 and 2).

Mixed-effects models

Analyses were conducted using *R* (R Core Team, 2019) version 3.6.2 glmer function of the lme4 package (version 1.1-27.1) with participants and items as random factors and maximum likelihood estimation. The model started with a maximal random effect structure with random slopes for all repeated-measure variables (Barr et al., 2013), and the "bobyqa" optimizer (Linck & Cunnings, 2015). As the fit was singular,

Table 1. By-participant correlation matrix

	Means (SD)	α	Recall	Acoustic	Phono L1	Phono FL	VSTM	SSTM	VWM	SWM
Recall	0.16 (0.12)	.91								
Acoustic	0.58 (0.10)	.66	.31***							
Phono L1	0.57 (0.08)	.47	.08	.30**						
Phono FL	0.58 (0.09)	.58	.32***	.29**	.37***					
VSTM	0.78 (0.11)	.69	.21*	.20*	.03	.06				
SSTM	0.73 (0.12)	.75	.30***	.23**	.22*	.11	.43***			
VWM	0.78 (0.14)	.82	.31***	.34***	.11	.18*	.58***	.48***		
SWM	0.64 (0.19)	.88	.26**	.31***	.15	.21*	.38***	.59***	.55***	
Vocab	41.65 (9.69)	.87	.49***	.36***	.18*	.16	.30**	.38***	.46***	.41***

Note: Phono = phonological, VSTM = verbal short-term memory, SSTM = visuo-spatial short-term memory, VWM = verbal working memory, SWM = visuo-spatial working memory, vocab = vocabulary, α = Cronbach's alpha.

*= p < .05;

**= p < .01;

***= p < .001.

Table 2. By-item correlation matrix

	Means *(SD)*	Recall	Phono Estimates
Recall	0.16 (0.12)		
Phono Estimates	0 (1)	.41***	
Ortho Estimates	0 (1)	.47***	.52***

Note: Phono = phonological, ortho = orthographic.
***= *p* < .001.

the model was simplified by removing the random slopes for the fixed-effects contributing the least variance (ibid.). The model investigated the main effects of memory, phonological abilities, and cross-linguistic orthographic similarity and controlled for cross-linguistic phonological similarity and native language vocabulary (using residuals). The model also included interactions between both cognitive abilities (memory and phonological abilities) and orthography. The following *R* syntax was used in the final model: glmer(accuracy~memory*Ortho_estimates + phono*Ortho_estimates + Phono_estimates + ZRE_vocab + (1 + Ortho_estimates| participant) + (1 + ZRE_vocab|item), data = data, family = binomial, control = glmerControl(optimizer = "bobyqa")). As can be seen from Table 3, all main effects were significant, but the interactions were not.

As explained earlier, the use of residual is criticized, and as such, a model using the same parameters as mentioned previously but with the unresidualized native language vocabulary test score is provided in the following text. As can be seen in Table 4, except for the main effect of memory, which is now only approaching significance, all other main effects remain significant and the interactions are not significant.

Discussion

The first aim of the current study was to investigate the impact of orthography on FL word learning. The results supported prior research, as even after controlling for how different a word's phonology is, it remained that how a word looked was important

Table 3. Final model showing main effects and interactions

						Random effects	
		Fixed effects				By participant	By Item
	Log odds (logit)	95% CI	Odds Ratio	*z*-value	*p*-value	*SD*	*SD*
(Intercept)	−2.23	[−2.46, −2.00]	0.11	−19.06	< .001		
STWmemory	0.48	[0.26, 0.70]	1.62	4.20	< .001	–	–
Phono abilities	0.39	[0.16, 0.62]	1.48	3.27	< .01	–	–
ZRE_vocab	0.37	[0.20, 0.54]	1.45	4.22	< .001	–	.04
Phono_est	0.23	[0.03, 0.42]	1.25	2.29	< .05	–	–
Ortho_est	0.41	[0.21, 0.61]	1.50	3.97	< .001	.25	–
STWmemory:Ortho_est	−0.06	[−0.16, 0.04]	0.94	−1.16	= .245	–	–
Phono abilities:Ortho_est	−0.07	[−0.18, 0.03]	0.93	−1.42	= .155	–	–

Note: Phono = phonological, STWmemory = Short-term and working memory composite score, Est = estimates, Ortho = orthographic; ZRE_vocab = standardized residual vocabulary test score; number of observations = 10,560, number of participants = 132, number of items = 120, VIFs < 1.34.

Table 4. Final model using unresidualized vocabulary test score

	Log odds (logit)	95% CI	Odds Ratio	z-value	p-value	By participant SD	By Item SD
		Fixed effects				Random effects	
(Intercept)	−2.24	[−2.47, −2.01]	0.11	−19.17	< .001		
STWmemory	0.22	[−0.02, 0.47]	1.25	1.79	= .074	–	–
Phono abilities	0.29	[0.05, 0.53]	1.33	2.38	< .05	–	–
Z_vocab	0.45	[0.25, 0.65]	1.57	4.49	< .001	–	.17
Phono_est	0.22	[0.03, 0.42]	1.25	2.28	< .05	–	–
Ortho_est	0.41	[0.21, 0.61]	1.50	3.98	< .001	.25	–
STWmemory:Ortho_est	−0.06	[−0.17, 0.04]	0.94	−1.17	= .243	–	–
Phono abilities:Ortho_est	−0.07	[−0.18, 0.03]	0.93	−1.43	= .153	–	–

Note: Phono = phonological, STWmemory = Short-term and working memory composite score, Est = estimates, Ortho = orthographic; Z_vocab = standardized vocabulary test score number of observations = 10,560, number of participants = 132, number of items = 120, VIFs < 1.34.

(Bartolotti & Marian, 2017; Bordag et al., 2016; Ellis & Beaton, 1993). In particular, participants performed better on the meaning recall task for words that were judged to have a similar orthography to native language words. This result expanded prior research as learning was measured at the level of meaning recall whereas prior research assessed the impact of orthography on recall and processing of *word form* (Bartolotti & Marian, 2017; Bordag et al., 2016; Ellis & Beaton, 1993). Having a dissimilar orthography impacted participant's ability to recall the meaning of the words either because it increased the difficulty in encoding new word form representations or because it increased the difficulty in linking these to meaning representations (or both). Future studies could probe the locus of the effect further by using a word form recognition task.

The second aim of the study was to investigate the interaction between orthography (a psycholinguistic variable) and cognitive abilities (individual differences variables). Here results showed no significant interactions. However, this could be due to how individual differences and word-level variables were conceptualized in this study (memory and phonological abilities for the former and cross-linguistic orthographic similarity for the latter) as well as how they were measured. Therefore, future studies should consider other individual differences and psycholinguistic variables as well as interactions between the two. For example, prior research investigated the role of phonological (phonotactic) similarity on FL learning (Ellis & Beaton, 1993; Kaush-anskaya, et al., 2011; Morra & Camba, 2009), but this research could be expanded to include the interaction between this and phonological abilities (an individual differences variable). Similarly, combinations of native/foreign languages that are more phonologically/orthographically distant (see Gamallo et al., 2017) may require more input from cognitive mechanisms to support learning. For language families that are closely related (e.g., similar phonology and orthography), long-term language knowledge can support learning (Majurus et al., 2008) and word processing (Akamatsu, 1999), and therefore there may be less reliance on cognitive mechanisms. In the current study, the combination of Welsh FL and English native language was used. The Welsh alphabet includes eight digraphs (*ch*, *dd*, *ff*, *ng*, *ll*, *ph*, *rh*, and *th*), which is dissimilar to the English alphabet and may have contributed to the difficulties experienced by participants with orthographically dissimilar words. However, Welsh is not considered

a difficult language to learn because of its transparent orthography. Every letter is pronounced and the grapheme to phoneme mapping is consistent as each letter corresponds to one sound (Davis, 2014; Ellis & Hooper, 2001). It would therefore be important to investigate other combinations of languages with more/less orthographic distance and also for language combinations with more/less transparent orthographies. Furthermore, the current study did not investigate interactions with learning situations (context-related difficulties in Housen & Simoens, 2016), and this should be addressed in future research. Bisson et al. (2021) found, for example, that memory was more important for intentional than incidental learning and this finding could be expanded by considering word-level predictors.

This study used familiarity judgments which can be qualified as a psycholinguistic measure compared to, for example, more lexical measures such as orthotactic probability or neighborhood density. Judgments are quick to obtain and do not necessitate lengthy calculations. In addition, they allow a glimpse into how native English speakers react to FL words' orthography. The comparative judgment approach is novel in psycholinguistics, but it is ideal for situations in which the criteria is difficult to define (Bisson et al., 2020). In the present study, participants judged the words as to their familiarity or similarity with native language word orthography. A Likert-scale could have been used instead, for example, from 1 = very unfamiliar to 7 = very familiar (similarly to Bordag et al., 2016; Dijkstra et al., 2010; Tokowicz et al., 2002). However, this kind of measure seems arbitrary. For example, how unusual does a word need to be to rate a 2 on the scale? Conversely, with comparative judgment a decision is made as to which out of two words is more familiar. It is therefore not necessary to decide the degree of unfamiliarity of each word, but rather to make a holistic judgment. This method has been used successfully in many domains, such as educational assessments and research (Bisson et al., 2016, 2020; Humphry & Heldsinger, 2019; Jones et al., 2019; Pollitt, 2012). Importantly, the current study shows it can also be applied in psycholinguistics.

As mentioned in the "Results" section, residual vocabulary test scores were used instead of the raw vocabulary test scores for the main analysis. This decision was based on the principal component analysis, which indicated that the native language vocabulary test score loaded onto the same factor as the memory tasks and, to a smaller extent, the acoustic task. This is probably due to some extent to the way native language vocabulary knowledge is tested in the PPVT. During the PPVT, participants hear a word such as "terpsichorean" once. They then have to keep this word active in memory while searching the four choices of pictures to select the correct depiction of the word. Therefore, to perform well on a vocabulary test such as the PPVT, one requires good short-term/working memory and acoustic abilities. It would therefore be important in future research to use a vocabulary test that relies less on additional cognitive abilities to perform the test and therefore is a purer measure of actual vocabulary knowledge. As using residuals in multiple regression analysis is a contentious issue in psycholinguistics (Wurm & Fisicaro, 2014), a second analysis was also provided with the unresidualized vocabulary test scores. The results of the two analyses were very similar overall confirming the main effect of orthographic similarity on FL word learning. However, the composite score for memory only approached significance in the unresidualized analysis. Based on all the prior research showing a role for short-term and/or working memory in word learning (e.g., Kapa & Colombo, 2014; Martin & Ellis, 2012; Morra & Camba, 2009), this is probably due to vocabulary test used in the current study as well as its relationship with cognitive abilities as explained in the preceding text. Finally, this

study involved secondary data analysis and it would therefore be important to replicate the results on another dataset.

In conclusion, the current study is unique in that it investigated both individual differences and word-level predictors concurrently (the interaction between learner-related and feature-related variables mentioned in Housen & Simoens, 2016). However, results did not show an interaction between orthography and the cognitive abilities (short-term/working memory and phonological abilities) included in this study. The current study expanded prior research showing that orthography impacted word learning even at the level of meaning recall. It is hoped this study will encourage researchers to pursue similar lines of enquiries (e.g., with combination of more/less distant languages and interactions between variables) to refine models of FL word learning.

Data availability statement. The experiment in this article earned an Open Data badge for transparent practices. The materials are available at https://doi.org/10.21253/DMU.14170832.v1

References

Akamatsu, N. (1999). The effects of first language orthographic features on word recognition processing in English as a second language. *Reading and Writing*, *11*, 381–403. https://doi.org/10.1023/A:1008053520326

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

Bartolotti, J., & Marian, V. (2017). Orthographic knowledge and lexical form influence vocabulary learning. *Applied Psycholinguistics*, *38*, 427–456. https://doi.org/10.1017/S0142716416000242

Bisson, M.-J., Gilmore, C., Inglis, M., &. Jones, I. (2016). Measuring conceptual understanding of mathematics using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education*, *2*, 141–164. https://doi.org/10.1007/s40753-016-0024-3

Bisson, M.-J, Gilmore, C., Inglis, M. & Jones, I. (2020). Teaching using contextualised and decontextualised representations: Examining the case of differential calculus through a comparative judgement technique. *Research in Mathematics Education*, *22*, 284–303. https://doi.org/10.1080/14794802.2019.1692060

Bisson, M-J., Kukona, A., & Lengeris, A. (2021). An ear and eye for language: Mechanisms underlying second language word learning. *Bilingualism: Language and Cognition*, *24*, 549–568. https://doi.org/10.1017/S1366728920000723

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2013). Incidental acquisition of foreign language vocabulary through brief multi-modal exposure. *PLoS ONE*, *8*, e60912. https://doi.org/10.1371/journal.pone.0060912

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, *64*, 855–877. https://doi.org/10.1111/lang.12085

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *The Quarterly Journal of Experimental Psychology*, *68*, 1306–1326. https://doi.org/10.1080/17470218.2014.979211

Bordag, D., Kirschenbaum, A., Rogahn, M., & Tschirner, E. (2016). The role of orthotactic probability in incidental and intentional vocabulary acquisition L1 and L2. *Second Language Research*, *33*, 147–178. https://doi.org/10.1177/0267658316665879

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS ONE*, *5*, e10773. https://doi.org/10.1371/journal.pone.0010773

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*, 286–291. https://doi.org/10.3758/BRM.42.1.286

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786. https://doi.org/10.3758/BF03196772

Davis., J. (2014). *The Welsh Language a History*. University of Wales Press.

De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*, 1–56. https://doi.org/10.1111/0023-8333.00110

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*, 284–301. https://doi.org/10.1016/j.jml.2009.12.003

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* 4th *edition* (PPVT-4). Pearson.

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign-language vocabulary learning. *Language Learning*, *43*, 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Ellis, N. C., & Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics*, *22*, 571–599. https://doi.org/10.1017/S0142716401004052

Firth, D. (2005). Bradley-Terry models in R. *Journal of Statistical Software*, *12*, 1–12. https://doi.org/10.18637/jss.v012.i01

Gamallo, P., Pichel, J. R., & Alegria, I. (2017). From language identification to language distance. *Physica A: Statistical Mechanics and Its Applications*, *484*, 152–162. https://doi.org/10.1016/j.physa.2017.05.011

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*, 163–175. https://doi.org/10.1017/S0272263116000176

Hu, C.-F. (2012). Fast-mapping and deliberate word-learning by EFL children. *The Modern Language Journal*, *96*, 439–453. https://doi.org/10.1111/j.1540-4781.2012.01357.x

Hubber, P. (2015). "Understanding the role of visuo-spatial working memory in adult mathematics." [Unpublished dissertation thesis]. The University of Nottingham.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

Humphry, S. M., & Heldsinger, S. (2019). A two-stage method for classroom assessments of essay writing. *Journal of Educational Measurement*, *56*, 505–520. https://doi.org/10.1111/jedm.12223

Jones, I., Bisson, M.-J., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal*, *45*, 662–680. https://doi.org/10.1002/berj.3519

Kapa, L. L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, *76*, 237–252. https://doi.org/10.1016/j.jml.2014.07.004

Kaushanskaya, M., Marian, V., & Yoo, J. (2011). Gender differences in adult word learning. *Acta Psychologica*, *137*, 24–35. https://doi.org/10.1016/j.actpsy.2011.02.002

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustic Society of America*, *128*, 3757–3768. https://doi.org/10.1121/1.3506351

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185–207. https://doi.org/10.1111/lang.12117

Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69. https://doi.org/10.1111/1467-9922.00032

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, *8*, 1–14. https://doi.org/10.3389/fpsyg.2017.01164

Majerus, S., Poncelet, M., Ven der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, *107*, 395–419. https://doi.org/10.1016/j.cognition.2007.10.003

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *7*, 379–413. https://doi.org/10.1017/S0272263112000125

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*, 322–352. https://doi.org/10.1037/a0026744

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PloS ONE*, *7*, e37527. https://doi.org/10.1371/journal.pone.0037527

Morra, S., & Camba, R. (2009). Vocabulary learning in primary school children: Working memory and long-term memory components. *Journal of Experimental Child Psychology*, *104*, 156–178. https://doi.org/10.1016/j.jecp.2009.03.007

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice*, *19*, 281–300. https://doi.org/10.1080/0969594X.2012.665354

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short-term memory, and word learning. *Journal of Phonetics*, *50*, 99–119. https://doi.org/10.1016/j.wocn.2015.03.001

Simon, E., & van Herreweghe, M. (2010). The relation between orthography and phonology from different angles: Insights from psycholinguistics and second language acquisition. *Language and Speech*, *53*, 303–306. https://doi.org/10.1177/0023830910372486

Tokowicz, N., Kroll, J. F., De Groot, A. M. B., & Van Hell, J. G. (2002). Number-of-translation norms for Dutch-English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, and Computers*, *34*, 435–451. https://doi.org/10.3758/bf03195472

Verhavert, S., De Maeyer, S., Donche, V., & Coertjens, L. (2018). Scale separation reliability: What does it mean in the context of comparative judgment? *Applied Psychological Measurement*, *42*, 428–445. https://doi.org/10.1177/0146621617748321

Vijayachandra, A. (2007). "The relationship between phonological working memory, phonological sensitivity, and incidental word learning." [Unpublished dissertation thesis]. Bowling Green State University. https://doi.org/10.1017/CBO9781107415324.004

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48. https://doi.org/10.1016/j.jml.2013.12.003