Page Proof Instructions and Queries

Please respond to and approve your proof through the "Edit" tab, using this PDF to review figure and table formatting and placement. This PDF can also be downloaded for your records. We strongly encourage you to provide any edits through the "Edit" tab, should you wish to provide corrections via PDF, please see the instructions below and email this PDF to your Production Editor.

Journal Title: Language Teaching Research

Article Number: 1024364

Thank you for choosing to publish with us. This is your final opportunity to ensure your article will be accurate at publication. Please review your proof carefully and respond to the queries using the circled tools in the image below, which are available in Adobe Reader DC* by clicking **Tools** from the top menu, then clicking **Comment**.

Please use *only* the tools circled in the image, as edits via other tools/methods can be lost during file conversion. For comments, questions, or formatting requests, please use T. Please do *not* use comment bubbles/sticky notes.

*If you do not see these tools, please ensure you have opened this file with **Adobe Reader DC**, available for free at get.adobe.com/reader or by going to Help > Check for Updates within other versions of Reader. For more detailed instructions, please see us.sagepub.com/ReaderXProofs.

No.	Query
GQ1	Please confirm that all author information, including names, affiliations, sequence, and contact details, is correct.
GQ2	Please confirm that the funding and conflict of interest statements are accurate.
GQ3	Please note that we cannot add/amend ORCID iDs for any article at the proof stage. Following ORCID's guidelines, the publisher can include only ORCID iDs that the authors have specifically validated for each manuscript prior to official acceptance for publication.
1	Please confirm corresponding author mailing/postal address, and in particular please confirm the postal code / zipcode. The details here and on the SAGE database differ
2	in this paragraph, 'dissociate' seems to be the incorrect verb; does 'distinguished' maintain the intended sense?
3	We have received Funding information for your article in our peer review system but there is no equivalent Funding statement in your article. Please check if a Funding entry needs to be added, and if so, please provide the required text.
4	Please confirm that acknowledgements, funding, and/or declaration of conflicting interest statements are correct.
5	Fernald & morikawa 1993 is missing from the references list. Please add full reference details here
6	Goldfield 1993 is missing from the references list. Please add full reference details here
7	Goldin-meadow 2012 is missing from the references list. Please add full reference details here
8	Please confirm the note paragraph

Article

LANGUAGE
TEACHING
RESEARCH

Language Teaching Research
1–32

DOI: 10.1177/13621688211024364
journals.sagepub.com/home/ltr
SAGE

Seeing or acting? The effect of performing gestures on foreign language vocabulary learning

Ana B. García-Gámez [GQ: 1]
University of Algarve, Portugal

Óscar Cervilla
University of Granada, Spain

Alba Casado
Jagiellonian University, Poland

Pedro Macizo
University of Granada, Spain

Abstract

We evaluate the impact of gestures during the teaching of vocabulary in a foreign language (FL). Spanish speakers learned words in a FL in four gesture conditions according to the relationship between the meaning of the words and the gestures (congruent gestures, incongruent gestures, gestures without meaning, and no gestures). The participants learned the words by performing gestures ('do' teaching group) or by observing the gestures performed by others ('see' teaching group). Compared to the meaningless gesture condition, the processing of congruent gestures facilitated the recall of second language (L2) words in the 'see' and 'do' teaching groups. However, the interference effect associated with the processing of incongruent gestures was greater in the 'see' teaching group than in the 'do' teaching group. Thus, the performance of gestures seems to mitigate the negative impact that the use of gestures may have on the teaching of vocabulary in a foreign language.

Corresponding author:
Ana B. García-Gámez, Departamento de Neurociências Cognitivas, Centre of Biomedical Research, University of Algarve, Gambelas Campus, Faro, 18005-139, Portugal.
Email: abgamez@ualg.pt [AQ: 1]

Keywords
gestures imitation, gestures observation, iconic gestures, teaching FL vocabulary, verbs learning

I Introduction

Bilingualism has become the rule rather than the exception. To illustrate, in 2016, around two thirds of working-age adults in the European Union knew at least one foreign language (FL) (Eurostat, 2019). However, it is also true that not everyone speaks a second language (L2). Therefore, it is necessary to implement teaching strategies that enhance FL learning. From a theoretical perspective (Kroll & Stewart, 1994), the acquisition of vocabulary in a FL would imply the establishment and reinforcement of connections between semantics and the new words that are learned in the FL. Thus, while novice learners preferably translate words through the use of lexical connections between first language (L1) and FL words, expert learners translate words across languages using connections between the meaning and the lexicon in FL (Talamas et al., 1999).

In the past, different strategies have been implemented to maximize the teaching of FL vocabulary by stressing the semantic processing of the material. Increased vocabulary learning has been demonstrated with FL teaching methods that foster a semantic route of processing; e.g. presenting a word to learn with a picture denoting its content in adults (Altarriba & Mathis, 1997), and children population (Comesaña et al., 2009); or imagining the meaning of a word to be learned (Ellis & Beaton, 1993; Wang & Thomas, 1995). Moreover, the use of gestures that represent the meaning of words to be learned has a positive effect on vocabulary acquisition (for reviews, see, for example, De Grauwe et al., 2014; Gullberg, 2014; Kelly et al., 2008, 2009; Macedonia & Knösche, 2011; Macedonia & Kriegstein, 2012; McCafferty & Stam, 2008; Morett, 2014; So et al., 2012).

In the current study, we focus on the role of gestures in the teaching of FL vocabulary. Specifically, we evaluate whether the mere exposure to gestures is sufficient to show the positive effect of gestures on the learning of FL words. Additionally, we examine if it is necessary for the participants to perform the gestures in order to observe their effect on learning.

1 The role of movement in teaching

In the educational system, the possible advantages associated with teaching through actions relative to the observation-based teaching have been a topic under discussion for decades (Goldin-Meadow, 1999). Forms of movement directly related to language processing are gestures that usually occur with speech. There are different types of gestures associated with language processing (McNeill, 1992), such as one or more fingers directed to a reference (deictic gestures), hand movements that reflect the prosody and emphasize the speech (beat gestures), gestures culturally specific such us thump up and hand in fist to indicate ‘good’, and representational or iconic gestures used to illustrate the meaning of what is being said. In particular, there are a large number of studies showing that the use of movements and, especially, iconic gestures, benefits the acquisition of vocabulary in FL (Feyereisen, 2006; Kelly et al., 2009; Macedonia & Klimesch, 2014; Morett, 2014).

The role of movements in FL vocabulary acquisition has been tested in previous teaching protocols. To illustrate, Asher (1966) proposed the Total Physical Response strategy as an efficient manner of acquiring new words in a FL. This strategy consisted of a guided method where students received indications in the L2. For example, children were taught the word *tobe* in Japanese (meaning 'to jump' in English), and every time they heard this word they had to perform the gesture denoted by the meaning of the word (to jump). In addition, empirical evidence on the role of movement in the teaching of an L2 reveals that better vocabulary learning is found when participants learn FL words accompanied by gestures that reflect the common use of objects whose names have to be learned (Feyereisen, 2006; Kelly et al., 2009; Macedonia & Klimesch, 2014). This beneficial effect associated with the use of gestures in the teaching of FL words has been demonstrated in several educational fields such as online courses, language learning, or technology use (Aleven & Koedinger, 2002; Bessen, 2015; although see Nakatsukasa, 2019). Moreover, the relevance of movement in language processing has been confirmed in many studies (Glenberg et al., 2004; Glenberg et al., 2008; Gluhareva & Prieto, 2017; Koriat & Pearlman-Avnion, 2003). Therefore, the performance of actions determines language comprehension.

On the other hand, the beneficial effect of gestures on the teaching of FL seems to depend on the type of word to be learned. In particular, nouns are easier to learn than verbs when children and adults acquire vocabulary (Childers & Tomasello, 2002; Fernald & Morikawa, 1993; García-Gámez & Macizo, 2019; Goldfield, 1993; Hadley et al., 2016; Tardif et al., 1997). Concrete nouns would have specific perceptual features that enhances the learning of new words while verbs represent dynamic information that enables the extraction of the verbal essence of the action (Golinkoff et al., 2002). In a study conducted by Hadley and colleagues (2016), the effect of gestures on the teaching of different types of words was directly addressed in preschool children. The results revealed that although concrete nouns obtained higher learning rates, the use of gestures while teaching verbs served as scaffold for the verbal information provided by them. This information would explain why most of the research evaluating the impact of gestures when teaching an L2 has used verbs as learning material, because the meaning of many verbs (e.g. verbs that denote actions with manipulable objects) are closely related with movements (e.g. Kelly et al., 2009). In fact, previous studies confirm that the semantic representation of verbs intrinsically contains a gestural or motor component (Boulenger et al., 2009; Hauk et al., 2004). To our knowledge, the study developed by García-Gámez and Macizo in 2019, was the first work in which nouns and verbs were compared when adult individuals learned FL words accompanied by gestures. The results revealed that, in general, nouns were easier to learn than verbs; however, when the meaning of the verbs and the gestures matched (i.e. when the iconic gesture clearly represented the verb meaning), the verbs reached a learning rate similar to that of nouns. Thus, gestures would play a greater role during the learning of verbs and would reduce the increased difficulty associated with the learning of verbs vs. nouns in a FL (see also, Kelly et al., 2009; Quinn-Allen, 1995).

Different theoretical arguments explain the beneficial effect that gestures have in the teaching of a foreign language. The 'learning-by-doing' perspective defends the active participation of the individual in the learning process by performing actions during the

educational development. Learning-by-doing can positively affect the formation of neural networks underlying the acquisition of knowledge and the performance of many cognitive skills (Goldin-Meadow et al., 2012). Moreover, gestures might favor the involvement of the participant in the learning task (Helstrup, 1987), so they could facilitate enhanced attention to the learning material which would increase the retention of words (Craik & Tulving, 1975). In addition, the performance of gestures when individuals process new words would promote the creation of a mental image associated with the meaning of this word, which would reinforce the semantic content of the word to be learned (Denis et al., 1991). Finally, gestures would enrich the encoding of the words to be learned by adding sensorimotor networks and procedural memory to the semantic/declarative memory associated with the meaning of the words (Macedonia & Mueller, 2016).

However, the use of gestures in teaching has some limitations. For example, Zhen and colleagues (2019) found that gestures are only useful in FL learning when the motor and sensory modalities share a common representational mapping. Moreover, a negative effect is found associated with the use of gestures in FL vocabulary learning. Incongruent gestures (i.e. familiar gestures with an easily recognizable meaning that mismatch the meaning of the word to be learned) make the teaching of FL difficult (Feyereisen, 2006; Kelly et al., 2009; Macedonia et al., 2011). The negative effect of incongruent gestures might be due to a semantic interference effect (Bernardis et al., 2008; Yap et al., 2010). To illustrate, Bernardis and collaborators (2008), used a gesture–word priming paradigm in which participants received a gesture prime followed by a word that they had to name. Compared to a baseline condition without gestures, the authors observed faster naming times when the gesture was congruent (i.e. a facilitation effect) and slower latencies when the gesture was incongruent with the meaning of the word (i.e. an interference effect). The semantic interference effect was interpreted as due to the difficulty in integrating the meaning of the gesture and the word when participants processed gesture–word pairs with different meanings.

Similarly, García-Gámez and Macizo (2019) observed both the positive and negative effects of the use of gestures in the teaching of vocabulary in a foreign language. In particular, the authors evaluated the teaching of foreign words (Vimmi, an artificial language) over three sessions (i.e. three consecutive days). During the learning process, each word to be learned was accompanied by a gesture and, afterwards, the participants had to reproduce the gesture previously presented. The results revealed better learning when the words were accompanied by a gesture congruent with the meaning of the word to be learned. Conversely, learning was impaired when the meaning of the gesture was inconsistent with that of the word. Thus, the processing of gestures could benefit or impair FL learning depending on the congruency between the gestures and the meaning of the words (Feyereisen, 2006; Kelly et al., 2009; Macedonia et al., 2011). However, in the study conducted by the authors, it could not be distinguished whether those effects were due to the mere observation of gestures and/or the performance of gestures during training, because the participants saw and reproduced the gestures in the training sessions. In the current study, we will evaluate this point by comparing the conditions of seeing vs. doing gestures during learning.[AQ: 2]

2 Consequences of seeing vs. doing while teaching

Several studies have examined the differences between self-performed tasks and experimenter-performed tasks (Cohen, 1981; Engelkamp & Zimmer, 1983). In 2012, Goldin-Meadow and colleagues directly compared the role of self-performed gestures versus seeing another individual producing them when children learned a mental transformation task. In their study, children were required to perform a mental rotation task in which they had to indicate whether two shapes presented in different orientations were the same figure or not. They used this task because previous studies demonstrate the close relationship between mental rotation and motor processing. When participants are instructed to mentally rotate a target, premotor areas involved in the planning of actions become active (Ganis et al., 2000; Glenberg et al., 2008) and participants spontaneously make gestures when they are required to explain how they solve this task (Chu & Kita, 2008). Goldin-Meadow and colleagues (2012) showed that children obtained better results when they were instructed to produce the gesture needed to solve the transformation task rather than when they observed the experimenter doing the movements. Thus, the performance of movements improved the results obtained in the task compared to the mere exposure to gestures associated with the mental rotation of objects.

Empirical evidence has confirmed also the importance that the self-generation of movements has during the teaching of linguistic material. For example, Morett (2018) showed that spontaneous production of gestures in a dialogic task had a greater impact on FL learning than the viewing of nonspontaneous gestures. On the other hand, Tellier (2008) presented new words in English to five-years-old children coupled with gestures that they had to reproduce or accompanied by pictures depicting the words meaning. Gestures production was more effective for learning than pictures observation. She argued that gestures are able to integrate the visual and motor modalities leaving a strong imprint in memory. Engelkamp and collaborators (1994) asked participants to learn sentences while performing the actions described in the sentences or by simply listening to and memorizing the material. The results revealed that the recall of sentences was higher when participants performed actions during the learning phase. The authors interpreted that the performance of actions favored the formation of a motor trace that benefited the retention of information. Empirical evidence supporting this argument comes from the study conducted by James and Swain (2011). The authors taught children action words associated with concrete toys. Some of the children manipulated the objects while learning and the remaining children observed the experimenter manipulating the same objects. When children listened to the words they had previously learned, motor brain areas were activated only in children who performed the toys manipulation themselves. Thus, the performance of motor actions when teaching vocabulary favors the learning of new words and the benefit associated with the performance of movement during learning seems to be due to the formation of a motor trace that would be activated during the subsequent retrieval of information.

However, other studies have found similar pattern of results when participants produce actions and when they only see actions produced by others (e.g. Rizzolatti & Craighero, 2004). In the lowest levels of linguistic processing, hand gestures observation or production seem to have a limited effect on the learning of FL segmental phonology

or novel phonetic distinctions (Hirata et al., 2014; Kelly et al., 2014). In higher linguistic levels, where hand gestures have proven to positively affect learning, different studies have reflected that self-generated movements and gestures observation yield similar outcomes. For example, Cherdieu et al. (2017) showed that the learning of anatomy lectures was similar when the instructor performed movements related to the lecture content or when the students imitated these movements. Concerning FL acquisition, during the learning of Chinese tones and words, similar results were found when participants observed (Experiment 1) and produced (Experiment 2) pitch gestures (metaphoric gestures mimicking melody in speech production) (Baills et al., 2019). In a recent study, undergraduate English speakers were acoustically presented with 10 Japanese verbs while an instructor performed iconic gestures. Participants obtained similar results when they were instructed to learn the words just by seeing the instructor gestures or by imitating her movements (Sweller et al., 2020). Finally, in a study addressing the effect of movements on sentences reading comprehension in children, Glenberg and colleagues (2004, Experiment 3) found intermediate results. Children were exposed to histories happening in a particular scenario (a farm) where different referents appeared (a sheep or a tractor). For the first group of children, objects referred to in the text were present and they were instructed to simulate the sentence content by manipulating the objects. The second group of children was required to imagine they were manipulating the toys. The results showed a beneficial effect of the manipulation condition while the imagined condition presented a modest improvement compared to an only-read condition. Thus, although movement performance appears to improve learning, it is not clear whether a learning protocol that involves self-generated actions would have an additional benefit to the mere observation of movements.

Taken together, the self-generation of movements during learning seems to have a beneficial effect on both non-linguistic tasks (e.g. Goldin-Meadow et al., 2012) and foreign language teaching (Engelkamp et al., 1994; Morett, 2018; Tellier, 2008). The formation of an enriched semantic representation in memory, containing not only verbal but also motor information, makes more accessible the information previously acquired and this is at the basis of the beneficial effect of performing movements during learning (Engelkamp et al., 1994; James & Swain, 2011). However, other studies seem to suggest that the mere viewing of gestures is enough for learning regardless of whether participants are instructed to perform the gestures themselves (Sweller et al., 2020). This controversial pattern of results could be due to methodological differences between studies such as the sample of participants, children (e.g. Goldin-Meadow, 2012) versus undergraduate students (e.g. Kelly et al., 2014), the type of learning task such as dialogic task (e.g. Morett, 2018) versus segmental phonology (e.g. Hirata et al., 2014), etc. In particular, the vast majority of studies in favor of the positive effect of self-generated gestures use semantically rich material such as words (Tellier, 2008) or sentences (Engelkamp et al., 1994) while some of the studies in which no difference between seeing and producing gestures is observed are focused on non-semantic linguistic levels (e.g. segmental phonology, Hirata et al., 2014; Kelly et al., 2014) or they manipulate the performance or non-performance of gestures in different experiments (e.g. Baills et al., 2019). In our study, we address these aspects in a unified study in which we manipulate between-groups the performance vs. viewing of gestures while participants learn words accompanied by iconic gestures that convey semantic information.

3 The current study

In general, there is agreement about the facilitative role that gestures have in the teaching of FL vocabulary (Feyereisen, 2006; Kelly et al., 2009; Macedonia & Klimesch, 2014). However, there is controversy regarding the role that the performance of self-generated movement ('do' teaching) versus the mere observation of movement ('see' teaching) has on FL vocabulary acquisition. As indicated above, some studies show an improved learning associated with the self-generation of gestures (Engelkamp et al., 1994; Morett, 2018; Tellier, 2008) while in other works no difference is found between the viewing of gestures and the performing of gestures by the learners of a FL (Hirata et al., 2014; Kelly et al., 2014; Sweller et al., 2020). The debate about the advantage associated with the self-generation of gestures during the teaching of FL words could be mediated by differences in the type of foreign language information that is taught (e.g. phonology, words, sentences, dialogues), the type of gesture used during the learning process (e.g. iconic gestures, pitch gestures), or the type of relationship between the learning material and the gestures used during the acquisition of a foreign language. In the present study, we address these aspects following the methodology employed by García-Gámez and Macizo (2019). To be more specific, in a single experiment, we explored the possible differences between FL teaching based on the viewing of gestures vs. FL teaching based on the performance of gestures using the same material (verbs), the same language pairs (L1: Spanish, FL: Vimmi), and the same learning conditions depending on the relationship between the gestures and the FL words.

On the other hand, one important contribution of the current study is that we explored the role of viewing vs. self-performing gestures in adult population unlike the vast majority of previous studies in which this issue is addressed with children (Glenberg et al., 2004; Goldin-Meadow et al., 2012; James & Swain, 2011; Tellier, 2008). It is important to highlight this methodological difference because adult population has much more experience performing actions than children and for adults, the semantic content of gestures and the visual imagery associated with words, is richer than in children (Bauer et al., 2017). Taken this issue into account, the mere observation of movements resembling action verbs would be enough for adult population to take advantage of the positive effect of gestures on FL learning and hence, gestures observation would reinforce the connections between the FL words and the semantic system in a similar manner as if they were performing the gestures themselves (see Sweller et al., 2020). In other words, adult participants may not show any difference between the viewing of gestures and the performance of gestures.

In our study, we directly evaluated these questions. To this end, adult Spanish (L1) speakers learned new words (verbs) in an artificial language (Vimmi, FL) in three consecutive days. This form of continuous training over three consecutive days was similar to the one used by García-Gámez and Macizo (2019) in which participants learned FL words (verbs, Experiment 2) while seeing and reproducing gestures during learning. In addition, the evaluation of the participants through three sessions of FL vocabulary teaching allowed to capture the learning curve throughout the training. Finally, the duration of the study, the number of exposures to FL words, and the number of learning sessions were established according to García-Gámez and Macizo so that the participants

could acquire a correct percentage of FL words without reaching a ceiling effect that would prevent the evaluation of the role of gestures through different learning conditions (e.g. in García-Gámez and Macizo, Experiment 2, the recall percentage was 20% in the first session, 43% in the second session, and 64% in the third session).

In the current study, participants were randomly assigned to two teaching conditions; the 'see' and the 'do' teaching groups. The 'see' teaching group was required to read aloud Spanish–Vimmi word pairs (L1–FL) to observe and to imagine themselves mimicking the gestures that were presented on a video at the same time. The 'do' teaching group was instructed to read aloud the word pairs in Spanish and Vimmi (L1–FL) and to imitate the gestures that were presented on the screen. Furthermore, in order to compare the effect of viewing vs. generating gestures depending on the type of gesture, we used the same four learning conditions previously employed by García-Gámez and Macizo (2019): the L1–FL word pair to be learned was (1) presented alone (no gesture condition), (2) coupled with a gesture that matched the meaning of the word (congruent gesture condition), (3) coupled with a gesture which was semantically unrelated to the word meaning (incongruent gesture condition), or (4) coupled with a gesture that did not denote any specific meaning (meaningless gesture condition). Macedonia et al. (2011) compared the effect of iconic gestures versus a baseline condition composed of meaningless gestures when participants learned FL vocabulary (i.e. Vimmi words). Thus, following these authors, in our study, we also considered the meaningless gesture condition as a baseline to evaluate both the benefit and the potential cost associated to the use of congruent and incongruent gestures in FL vocabulary learning. Note that the no gesture condition vs. the congruent/incongruent conditions differ in two aspects, the involvement of semantic information and the motor component. On the contrary, the meaningless gesture condition vs. the congruent/incongruent conditions only differ in the semantic component (only present in the congruent/incongruent condition but not in the meaningless gesture condition). Thus, we thought it would be suitable to use the meaningless gesture condition as a baseline in our study. However, the no gesture condition was also relevant in the current experiment to examine the effect of the motor activity during the encoding of FL words that was present in all gesture conditions (congruent, incongruent and meaningless gesture conditions) but not in the no gesture condition.

In this study, action verbs were used as learning material because of the close relationship between the meaning of verbs and movements (Boulenger et al., 2009; Hauk et al., 2004). To evaluate the FL learning rate across training sessions, participants performed a forward (L1–FL) and a backward (FL–L1) translation task. There were two reasons for using this evaluation task: (1) The forward and backward translation task has been used in many studies to evaluate the development of lexical and semantic proficiency in bilinguals (e.g. Kroll et al., 2002; Kroll & Stewart, 1994; Kroll & Tokowicz, 2001). (2) According to the Revised Hierarchical Model (i.e. Kroll & Stewart, 1994), there are differences in the way bilinguals perform this task depending on the direction of translation. Backward translation would involve lexical connections (FL–L1). On the contrary, forward translation would demand more cognitive resources because it would include an additional stage of semantic processing.

Based on previous studies about the role of gestures on FL vocabulary teaching, we expected to find a positive effect from the use of congruent gestures during the acquisition of FL words. At the same time, the processing of gestures not related to the meaning

of words would impair FL learning. This pattern of results would confirm both the benefit and the cost of using gestures during vocabulary acquisition in a FL (Feyereisen, 2006; Kelly et al., 2009; Macedonia et al., 2011). However, the most relevant predictions referred to the role of performing vs. observing gestures during the teaching of FL. If the mere observation of gestures was sufficient to modulate vocabulary acquisition, the pattern of outcomes would not depend on the type of training. On the contrary, if active training involving the performance of gestures maximizes learning, the learning rate would be higher in the 'do' teaching group compared to the 'see' teaching group. Regarding the direction of the translation task, as in previous studies, we expect to observe an asymmetric effect with a more efficient performance in the backward direction than in the forward direction (e.g. Kroll & Stewart, 1994). This prediction is due to the fact that forward translation entails more semantic processing than backward translation which would increase the difficulty when translating from L1 to FL than vice versa. Concerning the possible interactions between the direction of the translation task and the gesture conditions (congruent, incongruent, meaningless and no gestures) it would be expected that the forward translation, which is more semantically mediated compared to the backward translation, would have a greater impact on the gesture conditions that are semantically congruent and incongruent with the FL words relative to the meaningless gesture condition.

II Method

1 Participants

Thirty-one Spanish speakers we recruited for this study (28 women and 3 men). At present, it is difficult to consider Spanish speakers as monolinguals due to the L2 and even third language (L3) instruction at school and higher education. However, we looked for participants as less proficient in any FL as possible. To this end, we used the following inclusion criteria to select the participants in our study: On a daily basis, they had to report that (1) they had no contact with any other language (oral language or signed language) different from Spanish, (2) their last contact with a FL had to be at high school, (3) they had never received any FL instruction apart from regular education and, (4) they had never obtained a FL certification. All participants were selected from the same pool (they were students from the University of Granada, Spain) and received course credits as reward. Sixteen of them (15 women and 1 man) were randomly assigned to the 'do' teaching group, their mean age was 21.12 years ($SD = 2.53$). The remaining 15 participants (13 women and 2 men) were randomly assigned to the 'see' teaching group, their mean age was 21.13 years ($SD = 2.72$). Each participant provided written informed consent before performing the experiment. None of the participants reported a history of language disabilities, and they had normal or corrected-to-normal visual acuity.

2 Design and materials

The participants were randomly divided into two experimental groups. Half of them learned words with the 'see' training and the rest with the 'do' training. Four FL vocabulary teaching conditions were manipulated within-participants: No gesture condition: Spanish

(L1) – Vimmi (FL) word pairs had to be learned without gestures (e.g. *peinar*, 'to brush' in Spanish, and *tola*, a Vimmi word). Meaningless condition: L1–FL word pairs to be learned were coupled with meaningless gestures (e.g. *peinar–tola* and the gesture of touching the forehead and the right ear with the right forefinger). Congruent condition: L1–FL word pairs were accompanied with gestures that reflected the common use of objects whose names had to be learned in FL (e.g. *peinar–tola* and the gesture of holding an imaginary comb in the right hand and comb the hair from the front to the back). Incongruent condition: L1–FL word pairs were coupled with a gesture associated with an action verb different from that denoted by the L1 word (e.g. *peinar–tola* and the gesture of moving both hands fingers as if typing on a keyboard) (see Figure 1).

The material (word pairs and gestures) used in the current study were the same as those described in García-Gámez and Macizo (2019, Experiment 2). The congruent and incongruent gestures presented along with the L1–FL word pairs were iconic gestures (McNeill, 1992), which have been also called representational gestures (Kendon, 1983), that usually illustrate a concrete physical object or movements associated with a known action. For example, for the meaning of eating, the gesture would involve holding an imaginary spoon and putting it to the mouth. The gestures used in the meaningless condition were small movements performed with the hand that did not have iconic or metaphoric associations with the meaning of physical items (for example, to form a fist with one hand and raise the fingers of the other hand). We took care to select meaningless gestures with similar properties to those of meaningful gestures (e.g. hand configuration, the use of simple movement trajectory and spatial location). To ensure that the congruent, incongruent and meaningless gesture conditions were actually different at the level of association between the semantics of the words and the gesture, a pilot study was conducted. A group of 15 Spanish participants who did not participated in the main experiment was selected. The task consisted on the presentation of a video showing a gesture on the screen coupled with a Spanish word. Participants were instructed to rate the degree of semantic correspondence between the word and the gesture meaning from 1 (low match) to 9 (high match). The results revealed that the three gesture conditions differed in terms of the association between the words and the meaning of the gestures (a detailed explanation of this pilot study and the pattern of results is described in García-Gámez & Macizo, 2019).

In addition, 40 words were selected in Spanish. These words were verbs denoting familiar actions performed with hands and face (e.g. to kiss, to pray, etc.) (for the complete set of materials, see Appendix 1). Forty words were also selected from an artificial language, Vimmi (Macedonia et al., 2011; Macedonia & Knösche, 2011). The corpus of Vimmi words is constructed so that it avoids factors that might favor the learning of specific items (co-occurrence of syllables, similarity with words from languages such as Spanish, English, and French). Vimmi words were carefully selected so that they were pseudowords with legal orthography and phonology in Spanish but without meaning. The forty Spanish words were paired randomly with the forty Vimmi words. The resulting 40 word pairs were randomly sorted into 4 sets of 10 word pairs. Four lists of 40 word pairs were created. Each list was composed of 10 word pairs in each of the four training conditions (no gesture condition, meaningless condition, congruent condition, and incongruent condition). Each participant received one list, with all participants being

Figure 1. Teaching conditions used in the study. Spanish (L1) – Vimmi (FL) words (verbs) are coupled with different gesture conditions. In the example, *peinar* ('to brush' in Spanish) – *tola* (a Vimmi word) was accompanied by (a) the gesture of holding an invisible comb in the right hand and comb the hair from the front to the back (congruent condition); (b) the gesture of moving both hands fingers as if typing on a keyboard (incongruent condition); (c) the meaningless gesture of moving the hand from the forehead to the ear (meaningless condition); (d) the word pairs were presented without a gesture (no gesture condition).

randomly assigned to one of the four lists. Across lists, the 40 words were counterbalanced over the four training conditions, so that all word pairs appeared in all training conditions.

The forty word pairs were randomly sorted into 4 sets (10 word pairs in each) and randomly assigned to one of the four teaching conditions. The Spanish verbs across the 4 sets were equated in lexical variables (Davis & Perea, 2005). There were no differences

across word sets in the number of graphemes, $F < 1$ ($M = 6.53$, $SD = 1.48$), number of phonemes, $F < 1$ ($M = 6.30$, $SD = 1.44$), number of syllables, $F < 1$ ($M = 2.50$, $SD = 0.68$), lexical frequency, $F < 1$ ($M = 15.48$, $SD = 23.62$, per one million count), familiarity, $F < 1$ ($M = 3.50$, $SD = 3.09$), and concreteness, $F < 1$ ($M = 2.90$, $SD = 2.53$). Finally, we controlled for the similarity between the L1 and FL words. The Spanish and Vimmi words across conditions shared the same number of phonemes in the same position, $F < 1$ ($M = 0.37$, $SD = 0.63$), and irrespective of the position within the word, $F < 1$ ($M = 1.45$, $SD = 0.96$).

3 Procedure

FL vocabulary teaching involved three training sessions conducted on three consecutive days. In each session, participants performed, firstly, the FL training and, afterwards, the assessment of the FL learning. The two phases were separated by a 15-minute break. E-prime experimental software was used for stimulus presentation and data acquisition (Schneider et al., 2002). Participants were informed that the training sessions would be recorded on video to be sure that they followed the instructions provided by the experimenter. The procedure performed in this study was approved by the Ethical Committee on Human Research at the University of Granada (Spain) associated with the research project (Grant PID2019-111359GB-I00; number issued by the Ethical committee: 957/CEIH/2019) awarded to Pedro Macizo, and in accordance with the 1964 Helsinki declaration and its later amendments.

a 'See' FL training. Participants were presented with a block of 40 Spanish–Vimmi word pairs. These word pairs were grouped (10 word pairs in each group) according to the four teaching conditions (no gestures, meaningless gestures, congruent gestures, and incongruent gestures). This block was repeated 12 times. Hence, a participant received 480 trials where the 40 word pairs were presented 12 times. A short break was introduced between teaching blocks. The word pairs were randomly presented within each condition. In addition, the order in which the teaching conditions were presented within a block was counterbalanced. On each trial, the participant received a Spanish–Vimmi (L1–FL) word pair visually presented at the bottom of the screen. These word pairs were presented with a video where an actor performed the iconic gestures (see Figure 1). Gestures were recorded on video by the experimenter and they were congruent, incongruent, and meaningless, depending on the teaching condition. The duration of each recorded gesture was five seconds and the gesture was repeated twice. The participants were instructed to read aloud each L1–FL word twice. In the three gesture conditions (congruent, incongruent and meaningless), participants were instructed to imagine themselves imitating the actor gestures but they did not have to do any movement. They had to mentally produce the gesture presented each time they said aloud the L1–FL word pairs, so they repeated the gestures twice as the actor does in the videos. In the no gesture condition, the actor did not perform any movement and participants only repeated aloud the words pairs twice (L1 Spanish–FL Vimmi). For example, when participants received the word pair *peinar–tola* along with the congruent gesture, they had to say aloud this word pair at the time they mentally produced the gesture of holding an imaginary comb with

the right hand and to comb the hair from the front to the back. Once the participants had produced the word-pair twice, they had to press the space bar to continue to the next trial. Each training session lasted approximately 1 hour. The reason why we asked the participants of the 'see' teaching group to imagine themselves mentally producing the gesture was to prevent the participants from ignoring the gesture and focusing on the learning of the words in Vimmi. We recognize that, a priori, we had no way to confirm that the participants were indeed mentally producing the gesture. However, a posteriori, the differences observed in the 'see' teaching group across the gesture conditions (see Section III) seem to confirm that, actually, the participants followed the instructions provided by the experimenter and they mentally performed the required gesture.

b 'Do' FL training. As in the 'see' training, participants were presented with a block of 40 Spanish–Vimmi written word pairs. The learning material was exactly the same used in the 'see' FL training. The participants were instructed to read aloud each L1–FL word twice. In the three gesture conditions, participants had to produce the gesture presented in the video each time they said aloud the L1–FL word pair. Hence, they repeated the words pairs and the gesture twice. Participants started the production of the gesture when they began the production of the L1–FL word pair. Once the participants had produced the word-pair twice, they had to press the space bar to continue to the next trial. The training lasted approximately 1 hour.

c FL learning assessment. Two tests were used to evaluate the acquisition of FL words in the 'see' and 'do' teaching groups: Translation from Spanish into Vimmi (forward translation from L1 to FL) and translation from Vimmi into Spanish (backward translation from FL to L1). These tasks have been used in previous studies to evaluate FL learning (Kroll & De Groot, 2005; Poarch, Van Hell, & Kroll, 2015) and in studies about the role of gestures in FL vocabulary acquisition (García-Gámez & Macizo, 2019).

The order in which the translation tests were presented was randomized across the three training sessions and across participants to avoid any kind of order effect. In each translation task, the 40 Spanish words and the 40 Vimmi words were presented in the forward and backward translation tasks, respectively. On each trial, a word was presented in the middle of the screen until the participant produced its translation. Oral translations were recorded for later analyses of recall accuracy. Response times (RTs) from the presentation of the word until the beginning of the oral translation were also registered. The learning assessment lasted approximately 10 minutes depending on the participants' performance.

III Results

Translation direction (forward translation, backward translation), training session (first session, second session, third session) and teaching condition (no gestures, meaningless gestures, congruent gestures, incongruent gestures) were considered as within-participants factors while the teaching group ('see', 'do') was considered a between-groups variable. Analyses of variance (ANOVAs) were conducted with participants (F_1) and items (F_2, word pairs to be learned) as random factors. Bonferroni-adjusted pairwise

comparisons can be also consulted in the footnote.[1] Mixed model analyses are also reported as online supplementary material.

Although recall percentages (Recall %) are the main index of vocabulary acquisition, reaction times (RTs) were also analysed in this study. The RTs associated with correct translations were trimmed following the procedure described by Tabachnick and Fidell (2001) to eliminate univariate outliers. Raw scores were converted to standard scores (*z*-scores). Data points which, after standardization, were 3 *SD* outside the normal distribution, were considered outliers. After removing outliers from the distribution, *z*-scores were calculated again. The filter was applied in recursive cycles until no observations were outside 3 *SD*. In all analyses, we adopted a significance level of $\alpha = 0.05$. The percentage of outliers was 10.62% in the 'see' and 11.51% in the 'do' teaching groups. Only correct responses were included in the analyses of the RTs. Data points were excluded from the RT analyses if: (1) the participants produced nonverbal sounds that triggered the voice key, (2) the participants stuttered or hesitated in producing the word, (3) the participants produced something different than the word required.

Some small errors were allowed and considered correct responses depending on the length of the correct word to be produced (for the same coding system, see García-Gámez & Macizo): (1) For monosyllabic words, the replacement of a vowel; (2) for disyllabic words, the replacement of a vowel or a consonant but not both; (3) for words with three or more syllables, the inversion of a vowel and a consonant or the replacement of a vowel or a consonant. We considered as correct answers those low-level (sublexical) errors (e.g. replacement of vowels that did not involve new semantic content). Since the FL words were in an artificial language (Vimmi), when the participants replaced a vowel in Vimmi (e.g. 'rel') to create a legal word (e.g. 'rol') this type of response was considered an error (less than 5% of total errors).

1 Recall performance

Recall percentages were submitted to a Session × Teaching Condition × Teaching Group × Translation Direction analysis of variance (ANOVA). Table 1 shows the complete pattern of results obtained in the analysis. The mean of recall in the 'see teaching group' was 48.39% (47.00% in the forward translation direction and 49.78% in the backward translation direction) while in the 'do' teaching group the mean recall was 59.98% (59.32% in the forward translation direction, and 60.63% in the backward translation direction). There was a significant main effect of session, $F_1(2, 58) = 153.65$, $p < .001$, $\eta_p^2 = .84$, $F_2(2, 156) = 919.30$, $p < .001$, $\eta_p^2 = .92$. The recall percentage was 30.19% ($SE = 4.02$) in the first session, 59.61% ($SE = 3.89$) in the second session, and 72.74% ($SE = 3.68$) in the third session. Linear trend analysis revealed that the recall of FL words was higher in the final session relative to the beginning of the training, $F_1(1, 29) = 177.52$, $p < .001$, $\eta^2 = .86$, $F_2(1, 78) = 1315.19$, $p < .001$, $\eta^2 = .94$. The main effect of teaching condition was significant too, $F_1(3, 87) = 18.96$, $p < .001$, $\eta_p^2 = .40$, $F_2(3, 234) = 19.84$, $p < .001$, $\eta_p^2 = .21$. Mean recall percentage was 63.72% ($SE = 3.64$) in the congruent condition, 50.07% ($SE = 4.11$) in the incongruent condition, 47.73 ($SE = 3.81$) in the meaningless condition, and 55.20% ($SE = 3.84$) in the no gesture condition. Compared to the meaningless condition, the recall percentages were higher in the

Table 1. Statistical analyses performed on recall percentages and response times (RTs): Main effects and interactions between the 'see' and 'do' teaching groups.

Effects	Recall (%)				RT (ms)			
	F_1	p_1	F_2	p_2	F_1	p_1	F_2	p_2
Teaching group	2.61	.12	9.04	.00*	6.72	.01*	98.01	.00*
Translation direction	2.80	.10	5.30	.02*	17.71	.00*	39.57	.00*
Translation direction × teaching group	0.37	.55	0.46	.50	1.18	.29	13.78	.00*
Session	153.65	.00*	919.3	.00*	34.26	.00*	146.51	.00*
Session × teaching group	2.11	.13	135.88	00*	5.90	.00*	29.33	.00*
Teaching condition	18.96	.00*	19.86	.00*	6.91	.00*	20.21	.00*
Teaching condition × teaching group	2.24	.09**	1.60	.18	1.48	.23	7.14	.00*
Translation direction × session	12.45	.00*	28.69	.00*	1.90	.16	4.37	.01*
Translation direction × session × teaching group	0.02	.98	0.80	.45	0.11	.90	2.39	.10
Translation direction × teaching condition	9.37	.00*	21.31	.00*	.92	.44	1.22	.30
Translation direction × teaching condition × teaching group	4.98	.00*	10.04	.00*	1.36	.26	3.35	.02*
Session × teaching condition	1.32	.25	1.63	.14	2.13	.05**	2.75	.01*
Session × teaching condition × teaching group	1.55	.16	0.88	.51	2.13	.05**	7.17	.00*
Translation direction × session × teaching condition × teaching group	0.52	.79	0.90	.50	0.85	.53	2.98	.01*

Notes. *$p < .05$. **$p < .10$.

Table 2. Comparison between teaching conditions in the 'see' and 'do' teaching groups on recall percentages.

	'see' group				'do' group			
	L1 to FL		FL to L1		L1 to FL		FL to L1	
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Meaningless vs. congruent	8.03	.00*	3.28	.01*	6.28	.00*	2.61	.02*
Meaningless vs. incongruent	2.75	.02*	1.34	.20	0.36	.73	1.82	.09
Meaningless vs. no gestures	3.13	.01*	3.48	.00*	1.07	.30	2.85	.01*
Congruent vs. no gestures	2.51	.03*	1.29	.22	5.49	.00*	0.38	.71
Congruent vs. incongruent	3.31	.01*	3.50	.00*	4.84	.00*	3.62	.00*
Incongruent vs. no gestures	0.20	.85	0.84	.41	0.78	.45	4.85	.00*

Notes. FL = foreign language. L1 = first language. Comparison of recall percentages across teaching conditions obtained in the translation tasks. *$p \leq .05$.

congruent condition, $t_1(30) = 6.91, p < .001$, $t_2(79) = 5.73, p < .001$, and the no gesture condition, $t_1(30) = 3.55, p = .001$, $t_2(79) = 3.68, p = .007$. The difference between the meaningless condition and the incongruent condition was not significant, $t_1(30) = .86, p = .40$, $t_2(79) = .79, p = .43$. Finally, the congruent gesture condition showed significant advantages compared to the no gesture condition, $t_1(30) = 3.69, p < .001$, $t_2(79) = 2.12, p = .04$. Thus, the more participants trained the better results they obtained, however, the meaningless gesture condition seemed to be the most detrimental to the learning process, while the performance of congruent gestures improved the learning process compared to the no gesture condition.

The overall recall percentage was higher in the 'do' teaching group (59.98%, $SE = 4.99$) than in the 'see' teaching group (48.39%, $SE = 5.15$). The main effect of teaching group was not significant in the analyses by participants, $F_1(1, 29) = 2.60, p = .12, \eta_p^2 = .08$, but it was by items, $F_2(1, 78) = 9.04, p = .003, \eta_p^2 = .10$. Moreover, the Teaching Group × Teaching Condition × Translation Direction interaction was significant, $F_1(3, 87) = 4.98, p = .003, \eta_p^2 = .15, F_2(3, 234) = 10.06, p < .001, \eta_p^2 = .11$. No other three-way interactions showed significant differences (all *p*s > .05). The Teaching Condition × Translation Direction significant interaction was analysed further for each teaching group separately (see Table 2 and Figure 2).

a 'See' teaching group. The teaching condition (congruent, incongruent, meaningless, and no gestures) and the translation direction (forward, backward) were entered for ANOVA analysis. The main effect of teaching condition was significant, $F_1(3, 42) = 8.97, p < .001, \eta_p^2 = .39, F_2(3, 117) = 2.74, p = .04, \eta_p^2 = .07$. The recall percentage was 57.44% ($SE = 5.61$) in the congruent condition, 47.22% ($SE = 5.72$) in the incongruent condition, 39.00% ($SE = 5.71$) in the meaningless, and 49.89% ($SE = 5.55$) in the no gesture condition. Compared to the meaningless condition, there was a facilitation effect with congruent gestures so the recall percentage was higher in the congruent condition, $t_1(14) = 5.08, p < .001$, $t_2(39) = 3.33, p = .002$. In addition, the difference

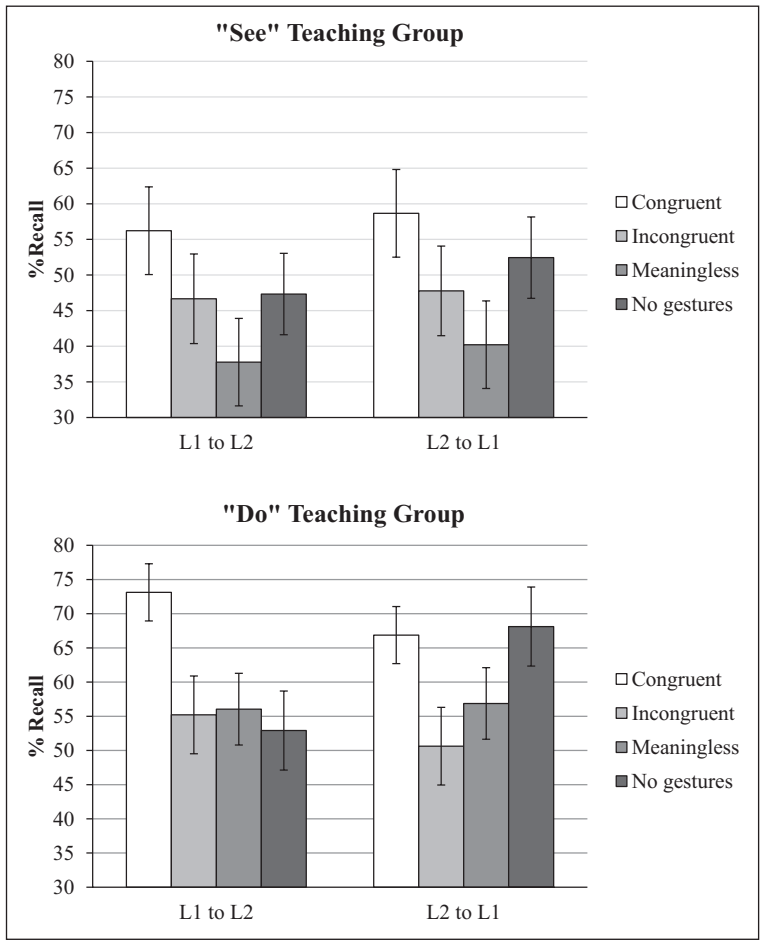

Figure 2. Recall percentage (% Recall) of the 'see' (upper graph) and 'do' (bottom graph) teaching groups as a function of translation direction (L1 to FL, FL to L1) and the gesture conditions (congruent, incongruent, meaningless, and no gestures).
Note. Standard error is plotted in vertical lines.

between the meaningless gesture condition and the no gesture condition was significant, $t_1(14) = 3.70$, $p = .002$, $t_2(39) = 2.13$, $p = .04$, thus, the processing of a meaningless gesture reduced the recall of FL words compared to the learning of words without gestures. There was a trend towards significance when the meaningless gesture condition was compared to the incongruent gesture condition by participants, $t_1(14) = 2.06$, $p = .06$, $t_2(39) = 1.31$, $p = .20$. Finally, when the congruent and no gesture conditions were compared, congruent gestures revealed a trend towards significance by participants, $t_1(14) = 1.97$, $p = .07$, $t_2(39) = .55$, $p = .58$. The remaining main effects or interactions did not reach significance (all ps $> .05$).

b 'Do' teaching group. As in the 'see' teaching group, the teaching condition and the translation direction were subjected to analysis of variance (ANOVA). There was a main effect of teaching condition, $F_1(3, 45) = 12.76, p < .001, \eta_p^2 = .46, F_2(3, 117) = 23.77, p < .001, \eta_p^2 = .38$, which was modulated by the Translation Direction × Teaching Condition interaction, $F_1(3, 45) = 13.94, p < .001, \eta_p^2 = .48, F_2(3, 117) = 26.12, p < .001, \eta_p^2 = .40$. This interaction was analysed further by examining the teaching condition effect in each translation direction separately.

In the forward translation direction (L1–FL), the teaching condition effect was significant, $F_1(1, 15) = 146.21, p < .001, \eta^2 = .91, F_2(3, 117) = 29.22, p < .001, \eta^2 = .43$. There was a facilitation effect so participants were more accurate in the congruent gesture condition (73.13%, $SE = 4.18$) relative to the meaningless gesture condition (56.04%, $SE = 5.24$), $t_1(15) = 6.28, p < .001, t_2(39) = 6.73, p < .001$. No differences were found between the meaningless and the incongruent (55.21%, $SE = 5.68$) conditions, $t_1(15) = .36, p = .73, t_2(39) = .25, p = .80$. No differences were found between the meaningless and the no gesture conditions (52.92%, $SE = 5.78$), $t_1(15) = 1.07, p = .30, t_2(39) = 1.02, p = .31$. Therefore, the interference associated with the processing of meaningless gestures vs. no gesture condition obtained in the 'see' teaching group was not observed in the 'do' teaching group. Finally, the performance of congruent gestures showed a higher advantage compared to the no gesture condition, $t_1(15) = 5.49, p < .001, t_2(39) = 7.51, p < .001$.

In the backward translation direction, the differences between teaching conditions were significant too, $F_1(1, 15) = 139.35, p < .001, \eta^2 = .90, F_2(3, 117) = 20.61, p < .001, \eta^2 = .35$. The mean of recall was 66.87% ($SE = 5.67$) in the congruent condition, 50.63% ($SE = 6.38$) in the incongruent condition, 56.88% ($SE = 5.39$) in the meaningless condition, and 68.13% ($SE = 5.00$) in the no gesture condition. The difference between the meaningless and the congruent conditions was significant, $t_1(15) = 2.61, p = .02, t_2(39) = 3.73, p < .001$, as well as the comparison between the meaningless and no gesture conditions, $t_1(15) = 2.84, p = .01, t_2(39) = 4.82, p < .001$. The comparison between meaningless and incongruent gestures was not significant by participants, $t_1(15) = 1.82, p = .09$, but it was by items, $t_2(39) = 2.38, p = .02$. The comparison between the congruent and no gesture conditions was not significant $t_1(15) = 0.38, p = .70, t_2(39) = .11, p = .91$.

2 Response times

As in the recall analyses, the Session × Teaching Condition × Teaching Group × Translation Direction variables were entered in analyses of variance (ANOVA) with participants (F_1) and items (F_2) as random factors (for the complete set of results, see Table 1; and for the mean RTs in each treatment, Table 3).

The session effect was significant, $F_1(2, 58) = 34.26, p < .001, \eta_p^2 = .54, F_2(2, 156) = 146.51, p < .001, \eta_p^2 = .65$. The mean response time in the first session was 2,536 ms ($SE = 64$), 2,252 ms ($SE = 65$) in the second session, and 2,020 ms ($SE = 58$) in the last session. Linear trend analyses showed significant differences between the first and the third session, $F_1(1, 29) = 49.53, p < .001, \eta^2 = .63, F_2(1, 78) = 254.27, p < .001, \eta^2 = .77$. The main effect of translation direction was significant too, $F_1(1, 29) = 17.71,$

Table 3. Response times (RTs) in the 'see' and 'do' teaching groups.

	Congruent	Incongruent	Meaningless	No gestures
'see' teaching group:				
L1–FL (1st session)	2,546 (181)	3,006 (176)	3,061 (184)	2,802 (213)
L1–FL (2nd session)	2,464 (156)	2,505 (141)	2,501 (169)	2,536 (162)
L1–FL (3rd session)	1,961 (142)	2,083 (123)	2,482 (191)	2,140 (140)
FL–L1 (1st session)	2,696 (193)	2,783 (131)	2,930 (146)	2,494 (203)
FL–L1 (2nd session)	2,068 (91)	2,207 (155)	2,464 (169)	1,997 (190)
FL–L1 (3rd session)	1,740 (125)	2,075 (156)	1,220 (110)	1,879 (140)
'do' teaching group:				
L1–FL (1st session)	2,474 (175)	2,326 (170)	2,543 (178)	1,756 (206)
L1–FL (2nd session)	2,055 (151)	2,328 (136)	2,278 (164)	2,469 (157)
L1–FL (3rd session)	1,902 (137)	2,130 (119)	2,069 (185)	2,071 (136)
FL–L1 (1st session)	2,262 (187)	2,538 (224)	2,325 (141)	2,030 (196)
FL–L1 (2nd session)	1,720 (88)	2,237 (150)	2,100 (164)	2,116 (184)
FL–L1 (3rd session)	1,729 (121)	2,256 (152)	1,753 (106)	1,835 (136)

Notes. Mean RTs (in milliseconds) obtained in the 'see' and 'do' teaching groups as a function of the translation direction (L1 to FL translation, FL to L1 translation), the training session (first session, second session, third session) and the teaching condition (congruent gestures, incongruent gestures, meaningless gestures, no gestures). Standard errors are in brackets.

$p < .001$, $\eta_p^2 = .38$, $F_2(1, 78) = 39.57$, $p < .001$, $\eta_p^2 = .34$, with participants responding faster in the backward (2,185 ms, $SE = 55$) than in the forward (2,354 ms, $SE = 54$) translation direction. The main effect of teaching condition was significant, $F_1(3, 87) = 6.91$, $p < .001$, $\eta_p^2 = .19$, $F_2(3, 234) = 20.21$, $p < .001$, $\eta_p^2 = .21$. Mean response times were 2,135 ms ($SE = 68$) in the congruent condition, 2,372 ms ($SE = 62$) in the incongruent condition, 2,393 ms ($SE = 57$) in the meaningless condition, and 2,177 ms ($SE = 80$) in the no gesture condition. Participants responded significantly faster in the congruent condition compared to the meaningless condition, $t_1(30) = 3.35$, $p = .002$, $t_2(79) = 5.21$, $p < .001$. The comparison between the meaningless and the no gesture conditions was marginal, $t_1(30) = 1.97$, $p = .06$, $t_2(79) = 1.91$, $p = .06$. No differences were observed between the meaningless and the incongruent gesture conditions, $t_1(30) = .15$, $p = .89$, $t_2(79) = .58$, $p = .57$. The difference between the congruent and no gesture conditions was not significant by participants, $t_1(30) = 1.45$, $p = .16$ but it was by items, $t_2(79) = 3.66$, $p < .001$. Therefore, participants were faster at the end of the training compared to the beginning of the teaching process. They showed the regular translation direction effect with slower responses in the forward direction compared to the backward direction (e.g. see also Kroll & De Groot, 2005); and they also exhibited faster processing times in the congruent gesture condition compared to the meaningless and the no gesture conditions (see also Macedonia et al., 2011).

Importantly, the main effect of teaching group was significant, $F_1(1, 29) = 6.72$, $p = .01$, $\eta_p^2 = .19$, $F_2(1,78) = 98.01$, $p < .001$, $\eta_p^2 = .55$. Participants in to the 'do' teaching group responded faster (2,137 ms, $SE = 71$) than the 'see' teaching group (2,401 ms, $SE = 73$). Finally, the Session × Teaching Group interaction was significant, $F_1(2, 58) =$

5.90, $p = .005$, $\eta_p^2 = .17$, $F_2(2, 156) = 29.33$, $p < .001$, $\eta_p^2 = .27$. Participants were significantly faster in the 'do' teaching group (2,282 ms, $SE = 89$) compared to the 'see' teaching group (2,790 ms, $SE = 92$) in the first teaching session, $F_1(1, 29) = 15.86$, $p < .001$, $\eta^2 = .35$, $F_2(1, 78) = 43.80$, $p < .001$, $\eta^2 = .36$. The difference between the teaching groups was not significant in the second teaching session by participants, $F_1(1, 29) = 1.92$, $p = .18$, $\eta^2 = .06$, but it was by items, $F_2(1, 78) = 11.04$, $p = .001$, $\eta^2 = .13$. Finally, at the end of the training, the teaching groups did not differ by participants, $F_1(1, 29) = .80$, $p = .38$, $\eta^2 = .03$, but the teaching effect was significant by items, $F_2(1, 78) = 9.65$, $p = .002$, $\eta^2 = .11$.

IV Discussion

Movements seem to play a role in many cognitive processes. A facilitative effect has been observed with different types of movements not only in educational settings, but also in clinical contexts (e.g. developmental disorders, aphasia treatments, etc.) (Botting et al., 2010; Hogrefe et al., 2013; Kelly et al., 2008). For example, pointing movements, defined as deictic gestures, and beat gestures, that reflect the prosody and emphasize the speech, have showed positive effects on language learning and development (Kushch., 2018; Morett, 2014; So et al., 2012). Iconic gestures that make reference to concrete entities or actions are especially remarkable in this context. These gestures have been used in many studies to explore how acts enhance memory consolidation in language production and comprehension (Goldin-Meadow & Alibali, 2013; Straube et al., 2012) and in FL vocabulary acquisition (So et al., 2012; Tellier, 2008). Previous works have addressed the comparison between gestures imitation and pictures observation (Tellier, 2008), some studies have compared the observation vs. the production of non-iconic gestures (e.g. pitch gestures, beat gestures) (Morett, 2014; Baills et al., 2019), and Sweller and colleagues (2020) directly evaluated the impact of self-performed gestures and gestures observation with iconic gestures that were implemented as learning material together with acoustically presented FL words. However, to our knowledge, the effects of observation vs. production of gestures have not been evaluated taking into account the different semantic relation that can be established between gestures and words (congruent, incongruent, meaningless).

In our study, the main interest was the direct comparison between the consequences of self-performed iconic gestures relative to the observation of gestures while teaching vocabulary in FL. To this end, participants learned FL verbs accompanied by different types of gestures (congruent, incongruent, meaningless, no gesture) in two different conditions. The 'do' teaching group was instructed to perform the gestures that accompanied the FL words while the 'see' teaching group observed the gestures without performing any movement.

The results found in the study revealed higher recall of FL words in the 'do' teaching group (60%) than in the 'see' teaching group (48%), however, the difference between types of training was significant in the analyses conducted by items but not in the analyses conducted by participants. Nevertheless, the participants retrieved FL words more rapidly in the 'do' teaching group (2,137 ms) than in the 'see' teaching group (2,401 ms). Thus, the training based on the self-generation of gestures facilitated, to some extent, the

retrieval of vocabulary in a FL. It is possible that requesting the participants in the 'see' teaching group to imagine the gesture would have reduced the possibility of observing a more consistent teaching group effect. However, as noted in Section II, the instruction to imagine gestures was implemented to prevent the participants in the 'see' teaching condition from focusing only on the learning of FL words while ignoring the processing of the gestures. Nevertheless, previous studies comparing the direct manipulation of objects against an imagined manipulation condition show better outcomes when participants actually perform the motor activity (Glenberg et al., 2004).

Concerning the effect of the different gesture conditions used in our study, the results revealed a facilitation effect on FL teaching when the words to be learned were accompanied by congruent gestures compared to the meaningless gesture condition. Specifically, the congruent gesture condition was associated with higher recall percentage and faster response time relative to the meaningless gesture condition. This improvement in FL teaching due to the processing of gestures whose meaning is congruent with that of the words to be learned confirms the results observed in previous research (Feyereisen, 2006; Kushch et al., 2018; Macedonia & Klimesch, 2014; Macedonia & Knösche, 2011). For example, in Macedonia and colleagues' (2011) study, a group of German speakers learned Vimmi words presented with iconic gestures (e.g. the word *suitcase* appeared with the gesture of an actor lifting an imaginary suitcase) or meaningless gestures (e.g. the word *suitcase* and the gesture of touching one's own head). The results revealed better recall for words learned with iconic gestures relative to words accompanied by meaningless gestures.

In addition to the facilitative effect of gestures in the congruent condition compared to the meaningless gesture condition, a controversial pattern of results was observed in the 'see' teaching condition of the current study. Participants remembered fewer words in the meaningless gesture condition compared to the no gesture condition. Overall, this outcome has been found in previous studies and seems to suggest that the use of gestures when teaching a foreign language can have both a positive and a negative effect (e.g. García-Gámez & Macizo, 2019; Kelly et al., 2009). The use of gestures in the learning of a foreign language makes the student face a dual task (the learner has to code both the gesture and the FL word). This dual task situation would interfere with the learning process. In fact, previous studies confirm the difficulty in the coding of information in a foreign language if, at the same time, students are instructed to undertake a concurrent task (Van Patten, 1990; Wong, 2001). Later, we will discuss in depth the negative effect observed in some of the gesture conditions of our study. Therefore, the use of gestures would have a negative effect during learning (i.e. when teachers perform meaningless gestures). However, this negative effect dissipates and, instead, a facilitation effect emerges, when the semantic information contained in the gesture converges with the meaning of the word to be learned (i.e. the congruent condition). This positive effect of the use of congruent gestures in teaching seems to be associated with a more efficient semantic processing compared to the meaningless gesture condition (Denis et al., 1991).

In the present study, and in line with previous findings, the FL learning facilitation effect due to the processing of congruent gestures was independent of the type of training ('see' teaching vs. 'do' teaching) which suggests that mere exposure of gestures is sufficient to observe the beneficial effect of gestures on the acquisition of vocabulary in a

FL (Sweller et al., 2020). Although our behavioral results are not directly linked to brain activity measures, this pattern of results is also in line with the outcomes of different studies in which it is demonstrated that mere observation of actions produces a pattern of brain activation similar to that found during the performance of motor actions (motor cortex activation, Stefan et al., 2005). Thus, the processing of gestures, both observing and performing them, might enrich the encoding of the words to be learned by adding sensorimotor networks and procedural memory to the semantic/declarative memory associated with the meaning of the words (Macedonia & Mueller, 2016). Hence, gestures would enhance semantic processing of words.

However, as we have discussed above, not all gesture conditions improved the learning process. Specifically, in the 'see' teaching condition, the performance of gestures incongruent with the meaning of FL words hindered vocabulary acquisition. Previous studies have shown the cost associated with processing incongruent gestures (Feyereisen, 2006; Kelly et al., 2009; Macedonia et al., 2011). The negative effect of incongruent gestures during FL vocabulary teaching might be due to semantic interference produced by the mismatch between the gesture and the meaning of the word to be learned (Bernardis et al., 2008; Yap et al., 2010). Semantic interference with incongruent gesture–word pairs would reflect the difficulty to integrate the meaning of the word and the gesture in working memory (Bernardis et al., 2008, p. 1126). In other words, the lack of correspondence between the information activated by the gesture and the word would lead to a conflict situation in working memory, which would hinder the learning and later recall of FL words.

On the other hand, the 'see' teaching group showed an additional interference effect, with lower recall of words in the meaningless gesture condition than in the no-gesture condition. The gestures used in the meaningless condition were small movements performed with the hand that did not have iconic or metaphoric associations with the meaning of physical items (for example, the gesture of moving the hand from the forehead to the ear). Thus, the interference found in the meaningless gesture condition would stem from the conflict between motor traces activated by the observation of meaningless movements and the processing of action verbs. In fact, previous studies have shown that both the processing of action verbs and the observation of actions produce activation of the motor cortex similar to the one that would be expected if the participants were performing a movement. Regarding the processing of verbs, Hauk et al. (2004) found that the passive reading of action verbs (e.g. to lick, pick or kick) differentially activated areas along the associated motor strip that overlapped with areas activated by actual movement of the tongue, fingers and feet, respectively. Concerning the observation of movements, Buccino and collaborators (2001) reported that the observation of both object and non-object related actions produced a somatotopically organized activation of premotor cortex similar to that of the classical motor cortex homunculus. Thus, the results found in the 'see' teaching group suggest that the processing of incongruent and meaningless gestures produced interference due to the mismatch between the semantic and motor information associated with these gestures and the words to be learned in a FL (Huang et al., 2019).

However, the interference effects found in the 'see' teaching group with meaningless gestures compared to the no gesture condition partially disappeared when participants

performed the gestures during teaching; that is, the 'do' teaching group. The attenuation of this interference effect in the 'do' teaching group was evident in the forward translation task but not in the backward translation task. We might ask about the cognitive mechanism responsible for minimizing the negative impact of gestures meaning mismatch in the 'do' teaching group. The benefit associated with the production of gestures while learning has been reported in previous studies. To illustrate, Cook and Goldin-Meadow (2006) found advantages in solving problems associated with self-performed gestures in children. They argued that the facilitative role attributed to acting while learning would be due to the reduction of working memory load. Thus, in the current study, the performance of gestures during teaching would reduce cognitive effort in working memory which would attenuate the conflict produced by the mismatch between the meaning and the motor trace of the gestures and the words in the incongruent and meaningless condition. In fact, it has been broadly confirmed that conflict resolution capacity strongly depends on the availability of resources in working memory (e.g. in bilingual population; Morales et al., 2013).

Moreover, in the 'do' teaching group, the attenuation of the interference effects was more evident in the forward translation task (L1–FL translation) than in the backward translation task (FL–L1) (see Figure 1). To illustrate, interference effects were found in the meaningless condition compared to the no gesture condition when participants performed the backward translation task; however, in forward translation, the percentage of recall was similar in the incongruent gesture condition (55.21%), the meaningless gesture condition (56.04%), and the no-gesture condition (52.92%). The main difference between forward and backward translation is the task difficulty. As we noted in Section I, the forward translation compared to the backward translation involves more semantic processing which increases the cognitive load of the task especially in early stages of foreign language learning (e.g. Kroll & Stewart, 1994). Similarly, in the no gesture condition of our study, the recall of FL words was lower and the response latencies higher in the L1–FL translation than in FL–L1 translation; thus, confirming the greater difficulty associated with forward vs. backward translation even when no gestures were used in teaching FL words.

Therefore, the effect of gestures on learning and retrieval of information seems to depend on task difficulty. For example, Marstaller and Burianová (2013) observed that a letter memorization task was more difficult to perform and the recall was lower in participants with low vs. high working memory amplitude. Importantly, the use of gestures facilitated the recall of letters compared to a condition without gestures only in those participants in which the task was more difficult (participants with low working memory amplitude). In our study, the performance of gestures during learning reduced the interference associated with the processing of incongruent and meaningless gestures to a greater extent when the task demands were high (forward translation task). Thus, the facilitative effect of performing gestures during learning seems to be more evident when the retrieval task involves high cognitive effort.

At this point, we might ask why the interaction between the translation direction and the teaching condition found in the 'do' teaching group (i.e. an attenuation of the interference when participants recalled FL words in the forward translation but not in the backward translation) was not observed in the 'see' teaching group (we thank an anonymous

reviewer for pointing out this issue). Thus, in the 'see' group of participants, the recall of FL words was not modulated by the translation direction nor did this variable interact with the teaching condition. In fact, it could be argued that the 'see' teaching group would exhibit an advantage in the translation task because they did not show the usual greater difficulty in the forward vs. backward translation (e.g. Kroll & Stewart, 1994). We acknowledge that we do not have a conclusive explanation for this question. However, it is possible that due to the way of presenting the verbal material in our study (the Spanish word and its translation in Vimmi (L1–FL word pairs), the participants would have used mainly a lexical coding strategy regardless of the translation direction (L2–L1 association learning, Comesaña et al., 2009; Lotto & De Groot, 1998). On the contrary, the performance of gestures in the 'do' teaching group would favor the semantic processing of the material (e.g. Denis et al., 1991; Macedonia & Mueller, 2016). This increased semantic processing in the 'do' group would be more evident in the forward vs. backward translation task (i.e. the translation direction effect) since the L1–FL translation involves the retrieval of conceptual information to a greater extent than the FL–L1 translation (e.g. Kroll & Stewart, 1994, Kroll & Tokowicz, 2001). Nevertheless, we acknowledge that this explanation must be taken with caution.

On the other hand, the pattern of results found in the 'do' teaching group of the current study might be compared with those reported by García-Gámez and Macizo (2019, Experiment 2) since the same procedure, experimental design, and learning material (i.e. verbs) were used in both works. In general, the main effects and interactions between variables were observed in both studies. However, some differences were noted in the set of outcomes. To be more specific, in García-Gámez and Macizo, the comparison between the congruent and no gesture conditions in the backward translation direction revealed an advantage associated with the performance of congruent gestures. This effect appeared in the latency and accuracy analysis. However, in the present study, the 'do' teaching group did not show differences between the congruent and the no gesture condition in the backward translation direction. Although tentative, one possible explanation for these differences may be due to variations in the overall rate of FL learning across studies. Specifically, the participants in the study by García-Gámez and Macizo revealed worse performance across the three learning sessions (i.e. slower response times and lower accuracy) compared to the participants in the current study. Thus, the participants in the García-Gámez and Macizo study had more room to benefit from performing congruent gestures during FL learning than the participants of the current experiment.

Future works might explore in deep the relationship between incongruent and meaningless gestures. Although these two conditions differ in terms of semantic content (i.e. meaningful-meaningless gestures), they seem to show a similar pattern of processing when they are associated with new words to be learned. Moreover, it would be interesting to plan a long-term evaluation program in order to verify if the recall of FL words is maintained over time when learners acquired words in the congruent gesture condition. In addition, it could be evaluated the effect of gestures imitation and gestures observation strategies on the teaching of nouns since these are the first words learned by children in language development.

To conclude, the results obtained in our study show the beneficial effects of processing gestures in FL vocabulary acquisition when they are congruent with the words to be

learned. This positive effect is enhanced when the learners are required to perform the gestures. However, gestures production has an additional effect on teaching: They seem to mitigate the adverse effect of processing gestures not related to the meaning of the words to be learned. Thus, taken together, if one had to choose between different teaching programs (with or without producing different types of gestures), the recommended alternative would be a training protocol in which participants produce gestures congruent with the words to be learned.[AQ: 3][AQ: 4]

Funding [GQ: 2]

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Spanish Ministry of Science and Innovation. Reference of the project: PID2019-111359GB-I00 / SRA (State Research Agency /10.13039/501100011033).

ORCID iDs [GQ: 3]

Ana B. García-Gámez https://orcid.org/0000-0001-5369-1691

Óscar Cervilla https://orcid.org/0000-0002-4011-9810

Supplemental material

Supplemental material for this article is available online.

Note [AQ: 8]

1. Bonferroni-adjusted pairwise comparisons were applied to the experimental analyses with participants (p_1) and items (p_2) as random factors. Regarding the accuracy analyses, the main effect of teaching condition was significant. Compared to the meaningless gesture condition, the participants recall was higher in the congruent condition, $p_1 < .001$, $p_2 < .001$, and the no gesture condition, $p_1 = .01$, $p_2 = .03$. The comparison between the meaningless and incongruent conditions was not significant, $p_1 = 1$, $p_2 = 1$. Finally, the congruent and the no gesture conditions showed significant differences by participants, $p_1 = .002$, but not by items, $p_2 = .20$. Concerning the significant teaching condition effect in the 'see' group, the congruent gesture condition showed a significant advantage compared to the meaningless condition, $p_1 < .001$, $p_2 = .04$. Compared to the meaningless condition, the no gesture condition showed significant differences by participants, $p_1 = .02$, but not by items, $p_2 = .20$. Neither the comparison between meaningless and incongruent conditions, $p_1 = .16$, $p_2 = .80$, nor the one between congruent and no gesture conditions, $p_1 = .25$, $p_2 = 1$, were significant. The Teaching Condition × Translation Direction interaction was significant in the 'do' teaching group. Considering the forward translation direction, the percentage of recall was higher in the congruent condition compared to the meaningless gesture condition, $p_1 < .001$, $p_2 < .001$, and the no gesture condition, $p_1 < .001$, $p_2 < .001$. There were not significant differences between the meaningless and no gesture conditions, $p_1 = 1$, $p_2 = 1$, or between the meaningless and incongruent conditions, $p_1 = 1$, $p_2 = 1$. In the backward translation direction, the congruent gesture condition showed a trend toward a significant advantage compared to the meaningless condition by participants, $p = .06$, that was significant by items, $p > .001$. There were also significant differences between the meaningless and the no gesture conditions, $p_1 = .03$, $p_2 < .001$. No significant differences were found between the meaningless and the incongruent gesture conditions, $p_1 = .63$, $p_2 = .29$, or between the congruent

and no gesture conditions, $p_1 = 1, p_2 = 1$. Finally, the RT analyses showed also a significant effect of teaching condition. There were significant differences between the meaningless and the congruent conditions, $p_1 = .005$, $p_2 < .001$. Neither the comparison between the meaningless and the no gesture, $p_1 = .22$, $p_2 = .051$, nor the one between meaningless and incongruent conditions, $p_1 = 1, p_2 = 1$, were significant. The difference between the congruent and the no gesture conditions was not significant in the analyses by participants, $p_1 = 1$, but it was in the analyses by items, $p_2 = .03$. Taken together, the general pattern of facilitation and interference effects associated with the use of congruent and meaningless gestures are maintained compared to the ANOVA results reported in Section III.

References

Aleven, V.A., & Koedinger, K.R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, *26*, 147–179.

Altarriba, J., & Mathis, K.M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language*, *36*, 550–568.

Asher, J.J. (1966). The learning strategy of the total physical response: A review. *The Modern Language Journal*, *50*, 79–84.

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*, 33–58.

Bauer, P.J., Pathman, T., Inman, C., Campanella, C., & Hamann, S. (2017). Neural correlates of autobiographical memory retrieval in children and adults. *Memory*, *25*, 450–466.

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology*, *25*, 1114–1128.

Bessen, J. (2015). *Learning by doing: The real connection between innovation, wages, and wealth*. Yale University Press.

Botting, N., Riches, N., Gaynor, M., & Morgan, G. (2010). Gesture production and comprehension in children with specific language impairment. *British Journal of Developmental Psychology*, *28*, 51–69.

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*, 1905–1914.

Buccino, G., Binkofski, F., Fink, G.R., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, *13*, 400–404.

Cherdieu, M., Palombi, O., Gerber, S., Troccaz, J., & Rochet-Capellan, A. (2017). Make gestures to learn: Reproducing gestures improves the learning of anatomical knowledge more than just seeing gestures. *Frontiers in Psychology*, *8*, 1689.

Childers, J., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*, 967–978.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*, 706–723.

Cohen, R.L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*, 267–281.

Comesaña, M., Perea, M., Piñeiro, 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*, 22–33.

Cook, S.W., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development*, *7*, 211–232.

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.

Davis, C., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*, 665–671.

De Grauwe, S., Willems, R.W., Rueschermeyer, S., Lemhöfer, H.S., & Schriefers, H. (2014). Embodied language in first- and second-language speakers: Neural correlates of processing motor verbs. *Neuropsychologia*, *56*, 334–349.

Denis, M., Engelkamp, J., & Mohr, G. (1991). Memory of imagined actions: Imagining oneself or another person. *Psychological Research*, *53*, 246–250.

Ellis, N., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*, 559–617.

Engelkamp, J., & Zimmer, H.D. (1983). Zum Einfluss von Wahrnehmen und Tun auf das Behalten von Verb–Objekt-Phrasen [The influence of perception and performance on the recall of verb–object phrases]. *Sprache & Kognition*, *2*, 117–127.

Engelkamp, J., Zimmer, H.D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition*, *22*, 34–39.

Eurostat. (2019). *Foreign language skills statistics*. Eurostat. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_skills _statistics (accessed June 2021).**[AQ: 5]**

Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, *18* 185–205.

Ganis, G., Keenan, J.P., Kosslyn, S.M., & Pascual-Leone, A. (2000). Transcranial magnetic stimulation of primary motor cortex affects mental rotation. *Cerebral Cortex*, *10*, 175–180.

García-Gámez, A.B., & Macizo, P. (2019). Learning nouns and verbs in a foreign language: The role of gestures. *Applied Psycholinguistics*, *40*, 473–507.

Glenberg, A.M., Gutierrez, T., Levin, J.R., Japuntich, S., & Kaschak, M.P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*, 424–436.

Glenberg, A.M., Sato, M., & Cattaneo, L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology*, *18*, R290–R291.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*, 609–631. **[AQ: 6]**

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*, 419–429. **[AQ: 7]**

Goldin-Meadow, S., & Alibali, M.W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*, 257–283.

Goldin-Meadow, S., Levine, S.C., Zinchenko, E., et al. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, *15*, 876–884.

Golinkoff, R.M., Chung, H.L., Hirsh-Pasek, K., et al. (2002). Young children can extend motion verbs to point-light displays. *Developmental Psychology*, *38*, 604–614.

Gullberg, M. (2014). Gestures and second language acquisition. In Müller, C., Cienki, A., E. Fricke., et al. (Eds.), *Body language communication: An international handbook on multimodality in human interaction* (pp. 1868–1875). De Gruyter Mouton.

Hadley, E.B., Dickinson, D.K., Hirsh-Pasek, K., Golinkoff, R.M., & Nesbitt, K.T. (2016). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly*, *51*, 181–198.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 307–322.

Helstrup, T. (1987). One, two, or three memories: A problem-solving approach to memory for performed acts. *Acta Psychologica*, *66*, 37–68.

Hirata, Y., Kelly, S.D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, *57*, 2090–2101.

Hogrefe, K., Ziegler, W., Wiesmayer, S., Weidinger, N., & Goldenberg, G. (2013). The actual and potential use of gestures for communication in aphasia. *Aphasiology*, *27*, 1070–1089.

Huang, X., Kim, N., & Christianson, K. (2019). Gesture and vocabulary learning in a second language. *Language Learning*, *69*, 177–197.

James, K.H., & Swain, S.N. (2011). Only self-generated actions create sensori-motor systems in the developing brain. *Developmental Science*, *14*, 673–678.

Kelly, S.D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313–334.

Kelly, S.D., Hirata, Y., Manansala, M., & Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Frontiers in Psychology*, *5*, 673.

Kelly, S.D., Manning, S.M., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, *2*, 569–588.

Kendon, A. (1983) The study of gesture: Some remarks on its history. In Deely, J.N., & M.D. Lenhart (Eds.), *Semiotics* (pp. 153–166). Springer.

Koriat, A., & Pearlman-Avnion, S. (2003). Memory organization of action events and its relationship to memory performance. *Journal of Experimental Psychology: General*, *132*, 435–454.

Kroll, J.F., & De Groot, A.M. (2005). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.

Kroll, J.F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Kroll, J.F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In Nicol, J. (Ed.), *One mind, two languages: Bilingual language processing* (pp. 49–71). Blackwell.

Kroll, J.F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*, 137–171.

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, *33*, 992–1004.

Lotto, L., & De Groot, A.M. (1998). Effects of learning method and word type on acquiring vocabulary in a new language. *Language Learning*, *48*, 31–69.

Macedonia, M., & Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind, Brain, and Education*, *8*, 74–88.

Macedonia, M., & Knösche, T. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain and Education*, *5*, 196–211.

Macedonia, M., & Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistic*, *6*, 393–416.

Macedonia, M., & Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Frontiers in Psychology*, *7*, 1–14.

Macedonia, M., Müller, K., & Friederici, A. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*, 982–998.

Marstaller, L., & Burianová, H. (2014). The multisensory of co-speech gestures: A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics*, *30*, 69–77.
McCafferty, S., & Stam, G. (2008). *Gesture, second language acquisition and classroom research*. Taylor & Francis.
McNeill, D. (1992). *Hand and brain*. The University of Chicago Press.
Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*, 187–202.
Morett, L.M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal*, *98*, 834–853.
Morett, L.M. (2018). In hand and in mind: Effects of gesture production and viewing on second language word learning. *Applied Psycholinguistics*, *39*, 355–381.
Nakatsukasa, K. (2019). Gesture-enhanced recasts have limited effects: A case of the regular past tense. *Language Teaching Research*. Epub ahead of print 26 August 2019. DOI: 10.1177/1362168819870283.
Poarch, G., Van Hell, J., & Kroll, J. (2015). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism: Language and Cognition*, *18*, 357–371.
Quinn-Allen, L. (1995). The effect of emblematic gestures on the development and access of mental representations of French expressions. *Modern Language Journal*, *7*, 521–529.
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Revision of Neuroscience*, *27*, 169–192.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide* (Version 1.1). Psychology Software Tools.
So, W.C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*, 665–681.
Stefan, K., Cohen, L.G., Duque, J., et al. (2005). Formation of a motor memory by action observation. *Journal of Neuroscience*, *25*, 9339–9346.
Straube, B., Green, A., Weis, S., & Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: An FMRI study. *PLoS ONE*, *7*, 1–10.
Sweller, N., Shinooka-Phelan, A., & Austin, E. (2020). The effects of observing and producing gestures on Japanese word learning. *Acta Psychologica*, *207*, 103079.
Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics*. 4th edition. Allyn & Bacon.
Talamas, A., Kroll, J.F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*, 45–58.
Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*, 535–565.
Tellier, M. (2008). The effect of gestures on second language memorization by young children. *Gesture*, *8*, 219–235.
Van Patten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*, 287–301.
Wang, A., & Thomas, M. (1995). Effects of keyword on long-term retention: Help or hindrance? *Journal of Educational Psychology*, *87*, 468–475.
Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition*, *23*, 345–368.
Yap, D., So, W., Yap, J., Tan, Y., & Teoh, R. (2010). Iconic gestures prime words. *Cognitive Science*, *35*, 171–183.
Zhen, A., Van Hedger, S., Heald, S., Goldin-Meadow, S., & Tian, X. (2019). Manual directional gestures facilitate cross-modal perceptual learning. *Cognition*, *187*, 178–187.

Appendix 1. Stimulus material used in the study.

Spanish verbs	English translation	Gestures
afeitar	to shave	Make the movement of grabbing an imaginary razor and passing it by the right cheek
apretar	to squeeze	Stretch the right hand with the palm up and close the fist strongly
barrer	to sweep	Make the movement of taking an imaginary broom with both hands and passing it from front to back
batir	to beat	Put the left hand as if was an imaginary bowl and with the right one perform the movement of a whisk over the bowl.
beber	to drink	Stretch the thumb and the little finger, close the rest of the fingers of the right hand and move it towards the mouth
besar	to kiss	Put the right hand fingers on the lips and whilst separating them, give a kiss
botar	to bounce	Stretch the right hand in horizontal plane with respect to the floor with the palm of the hand facing the floor and move the hand up and down
callar	to silence	Put the right forefinger in the lips in a vertical plane
comer	to eat	Make the imaginary movement of taking a spoon and putting it to the mouth
coser	to cough	Hold an imaginary piece of fabric with the left hand and perform the cough movement with an imaginary needle in the right hand holding it between the thumb and the forefinger
disparar	to shoot	Stretch the thumb and the forefinger and perform a slight downward movement in an horizontal plane with respect to the floor
dormir	to sleep	Put together the palms of the hands, put them to the right and nod the head to the right a little and close the eyes
escribir	to write	Take an imaginary pen with the right hand between the thumb and the forefinger and perform left-right writing movements in front of your face
escuchar	to listen	Stretch the right forefinger and touch the right ear
escurrir	to drain	Fist both hands in front of your chest with 15 centimeters distance between them and spin them in opposite directions
estirar	to stretch	Put in contact the thumb and the forefinger of both hands, put the hands together and separate the hands
estornudar	to sneeze	With the nose between the fingers of both hands perform a slight downward movement of the head.

(Continued)

Appendix 1. (Continued)

Spanish verbs	English translation	Gestures
estrujar	to wring out	Put the palms of the hands one in front of the other separated by approximately 15 centimeters with separation between the fingers, approximate the hands at the same time that you fold the fingers and rotate the hands in opposite directions
fotografiar	to photograph	Take an imaginary camera in front of your face and press an imaginary button with the right forefinger
fumar	to smoke	Draw a 'V' with the right and middle forefingers and put them next to the mouth, breathe, separate the fingers from the mouth and breathe out
guiñar	to wink	Wink with the right eye
llorar	to cry	Close both fists, put the thumb and the forefinger in contact with the cheeks and perform a rotating movement
maquillar	to apply makeup	Hold an imaginary eye liner with the right hand between the thumb and the forefinger and draw a line following the eye line next to the eye
mirar	to look	Stretch the right forefinger and put it under the right eye
morder	to bite	Show the teeth and open and close the jaw
negar	to deny	Perform a lateral movement with the head
peinar	to brush	With an imaginary comb in the right hand, comb the hair from the front to the back
pintar	to draw	Stretch the left hand, hold a brush with the right hand between the thumb and the forefinger and draw with slow movements on the left hand
rechinar	to grind	Show the teeth and move the lower jaw from right to left
recortar	to trim	Draw a 'V' with the right forefinger and middle finger, put them in an horizontal plane, approach and separate them
relamer	to lick your lips	Stick out the tongue and pass it by the top lip
rezar	to pray	Put the palms of the hands together in a vertical plane in front of your chest
romper	to break	With the palms of the hands facing the floor, close the fists and put them together sideways. Separate the fists sideways
sacudir	to shake	Close the hands with the palms one in front of the other, and move them towards and outwards from the body

(Continued)

Appendix 1. (Continued)

Spanish verbs	English translation	Gestures
silbar	to whistle	Position the lips as if kissing and expel air
sonreir	to smile	Show your teeth in a big smile
soplar	to blow	Expel air through the mouth
teclear	to type	Move both hands fingers as if you were typing on a keyboard
toser	to cough	Put the right fist with the thumb next to the mouth and perform a slight downward movement of the head
untar	to spread	Stretch both hands, and with the right hand perform the movement of a knife spreading something on the left hand
sin significado 1	meaningless 1	Close the right fist and move the arm from the left to the right
sin significado 2	meaningless 2	Cross the arms in front of the chest
sin significado 3	meaningless 3	Stretch the palms of the hands, and put the right one in front of the right eye and the left one covering the mouth.
sin significado 4	meaningless 4	Draw an imaginary triangle with the right forefinger in front of the face
sin significado 5	meaningless 5	Touch the chin and then the forehead with the right forefinger
sin significado 6	meaningless 6	Stretch the right arm and hand with the palm of the hand facing the floor and move it from left to right
sin significado 7	meaningless 7	With the right forefinger, touch the left cheek and then the right one
sin significado 8	meaningless 8	Touch the forehead and the right ear with the right forefinger
sin significado 9	meaningless 9	Stretch the left hand, clap the hands twice, first with right fist closed and then with the right hand stretched
sin significado 10	meaningless 10	Put the palms of the hands together and rotate them from one side and to another

Notes. Left Column: Spanish verbs used in the experiment. Middle column: English translations. Right column: approximate description of gestures.