Sweller, N., Shinooka-Phelan, A., & Austin, E. (2020). The effects of observing and producing gestures on Japanese word learning. *Acta Psychologica*, 207, 103079.

Title: The Effects of Observing and Producing Gestures on Japanese Word Learning

Authors: Naomi Sweller[1], Aya Shinooka-Phelan[1], Elizabeth Austin[1,2]

1 Department of Psychology, Macquarie University, Sydney, Australia

2 Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Please address correspondence to: Naomi Sweller, Department of Psychology, Macquarie University, Sydney, 2109, Australia.

Email: naomi.sweller@mq.edu.au. Phone: +61 2 9850 8084

ORCID: 0000-0003-2269-4763

1. Introduction

Communication occurs through both verbal and non-verbal modalities, such as gesture. Across ages and cultures, people gesture to further illustrate their verbal message (Hostetter, 2011; Macedonia & von Kriegstein, 2012). Gestures are used not only when communicating in one's native language, but also in a foreign or second language (L2), frequently to compensate for speech difficulties by visually representing the meaning of words (Macedonia & Klimesch, 2014).

Gestures can benefit learners across a variety of tasks (see Dargue, Sweller & Jones, 2019 for a review), including in educational settings such as when learning a second language (Gullberg, 1998, 2006, 2014; Gullberg, deBot, & Volterra, 2008; McCafferty, 2002). Observing and producing gestures both promote learning beyond verbal modality learning alone (e.g., Baills, Suárez-Gonzáles, Gonzáles-Fuente, & Prieto, 2018). However, research examining the role of gesture in L2 learning is limited (Macedonia & von Kriegstein, 2012). It is not yet clear whether observing or producing gestures is more effective as a language teaching tool (though see Morett, 2018), and whether these effects vary with time to recall. Further, examining gesture production at test, and how such gesture production may mediate the relationship between gesture observation/production during learning and recall at test is yet to be examined. This study examined the effects of observing and reproducing gestures during foreign word learning, whether such effects vary with time, and the potential mediation of observing or producing gestures on word learning via gesture production at test.

1.1 Foreign Word Learning

Vocabulary is important for effectively communicating meaning in any language. One of the difficulties that foreign language learners often face is the acquisition of novel vocabulary terms (Kelly, McDevitt, & Esch, 2009). Despite learners' efforts to acquire these words, memory for foreign language words often decays shortly after learning (Macedonia, Müller, & Friederici, 2011). Given the beneficial effects noted above of gestures on a variety of learning tasks, it is possible that encoding and retaining novel vocabulary terms may be facilitated by the presence of gesture during learning.

Examination of the effect of gesture in learning foreign words or sentences has been limited. Notably, the effect of gesture on L2 learning is often assessed in written formats (e.g. Kelly et al., 2009). Examinations of written recall only are restricted in their generalisability however, as language is used in both written and verbal formats. The extent to which gestures can benefit verbal recall remains unclear.

Recent research has examined novel L2 word learning in a dialogic task (Morett, 2018). Morett (2018) found that producing, but not viewing gestures affected communication and learning, and concluded the effects of producing gestures were stronger than the effects of observing gestures. However, the study only examined the effects of producing gestures while explaining words to another learner, and did not examine the effects of producing gestures while participants learned the words themselves. Furthermore, participants were not instructed to produce any specific type of gesture. Not all gestures are equivalent, and a consideration of gesture types is warranted.

1.2 Gesture Classification

Gestures can be classified into many non-mutually exclusive formats. Iconic gestures represent a concrete concept, visually recreating an aspect of the referent such as shape, size and movement (McNeill, 1992). For instance, a hand gesture which holds an imaginary glass and moves it towards the mouth can represent an action to drink. Given their link to concrete imagery, iconic gestures are those which may best facilitate the link between novel words and meanings. Such gestures are the focus of the current study. Other gesture types include metaphoric gestures (which represent abstract concepts), beat gestures (rhythmic hand movements) and deictic (pointing) gestures (McNeill, 1992).

Gestures can either be produced by others and observed by listeners, or produced by speakers themselves. The effect of both types of gesture on foreign word learning has been examined (e.g., Allen, 1995; De Nooijer, Van Gog, Paas, & Zwaan, 2013; Rowe, Silverman, & Mullan, 2013; Tellier, 2007), but with no attention to the potential mediating effect of producing gestures at test on the relationship between gesture observation and learning at training, and performance at test. We turn now to a discussion of observing and producing gestures on language learning, and proposed mechanisms underlying each function.

1.3 Observation of Gesture and Learning

Observing gestures can benefit foreign word learning, tone identification and pronunciation (Gluhareva & Prieto, 2016; Allen, 1995). For example, observing syllabic beat gestures accompanying foreign words improves tone identification and word pronunciation (Baills et al., 2018). Mapping words to meaning is an important part of foreign language vocabulary acquisition (Jiang, 2002). The semantic representations that iconic gestures present can enable learners to build and access connections between words and meanings in their native language, making the memory of words more enduring (Kelly et al., 2009).

Foreign language learners benefit from observing iconic gestures, which can physically provide semantic representations of novel words (see Huang, Kim & Christianson, 2018; Kelly et al., 2009). Establishing the connection between novel words and their meaning is a necessary process to integrate foreign language vocabulary (Allen, 1995). As such, observing iconic gestures can aid learners to internalise foreign words.

1.3.1 Mechanisms Underlying Observing Gesture. There are a number of methods through which observing gesture may enhance foreign language learning. Speech and gesture may work together to convey information through differing modalities. Dual coding theory and multimedia learning theory argue that listeners construct separate mental representations of speech and gesture content (Mayer, 2009; Paivio, 1990). Referential connections integrate the two mental representations, allowing the verbal and nonverbal systems to trigger activity in each other. If one mental representation is degraded or forgotten, the other may still be accessible to memory (Paivio, 1990).

It is also possible that observing gestures may elicit gesture production by learners, in turn enhancing learning. Observing a teacher's gestures can benefit learners by increasing the likelihood of their own spontaneous gesture production (Cook & Goldin-Meadow, 2006). That is, observing a teacher's gestures may play two important roles: not only through enhancing learners' comprehension of the spoken information, but also through encouraging learners to spontaneously produce gestures themselves. Producing gestures can itself lead to internalisation and enhanced retention of spoken information.

1.4 Production of Gesture and Learning

Producing or imitating gestures during learning can enhance word learning (Baills et al., 2018; Macedonia, Bergmann & Roithmayr, 2014; Macedonia & Klimesch, 2014; Macedonia & Knösche, 2011). For example, words or sentences of an artificial language that participants learned by reproducing an instructor's gestures were recalled better over both the short-term for sentences (six days) (Macedonia & Knösche, 2011) and the long-term for words (444 days) (Macedonia & Klimesch, 2014) than those taught verbally. Participants in the 2011 study received daily training sessions, however the effect of producing gestures over verbal only learning did not appear until the third day of learning (Macedonia & Knösche, 2011), indicating that the effect of producing gestures may not emerge in the short-term.

Producing gestures has similarly found to benefit learning of non-artificial languages, such as Chinese tones (Baills et al., 2018). Furthermore, Hirata, Kelly, Huang and Manansala (2014) explored whether observing and/or producing similar types of beat gestures enhanced learning of Japanese length contrasts. Observing and/or producing gestures improved vowel length identification, although it should be noted that the Hirata et al. (2014) study did not include a condition where participants neither observed nor produced gestures, meaning conclusions regarding the efficacy of gestures on foreign word learning remain tentative. Similarly, Kelly, Hirata, Manansala and Huang (2014) examined the effects of observing and producing gestures on learning Japanese moras and syllables (but not whole words) and found similar results for both observing only and observing and producing gestures. Again however, this study only compared both observing and producing gestures with observing gestures only, without any no-gesture condition.

1.4.1 Mechanisms Underlying Producing Gesture There are a number of mechanisms through which the production of gesture may enhance foreign language learning. Producing gestures may increase available cognitive resources (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wilson, 2002). According to theories of embodied cognition, physical movements, including gesture, play an important role in shaping cognitive processes, and the body can be used to offload cognitive work onto the environment (Wilson, 2002). In this way, it may be possible to extend our limited cognitive capacity by externalising or embodying mental representations through gesture production (Wilson, 2002). If producing gesture lightens the load on cognitive resources, more resources can be allocated to the task, potentially improving memory performance (Goldin-Meadow, 2010).

It is also possible that producing gestures may facilitate access to stored information by providing a stable, external physical memory cue (Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). Alibali, Kita, and Young (2000) argued that physically producing what is in one's mind through gestures enhances speech production by conceptualising thoughts. Gestures can perhaps alter the way we conceptualise information through activating, manipulating, packaging and exploring information (Kita, Alibali & Chu, 2017). Given their link to concrete concepts, iconic gestures may be particularly well placed to provide such external memory cues. Iconic gestures can enhance word retrieval by cross-modally activating a word's concept, due to their physical representations of the semantic features of words (Rauscher, Krauss, & Chen, 1996). It is clear that observing and producing gestures both have the potential to enhance novel word learning. What is less clear however is whether producing gestures during learning has any beneficial effects *beyond* those of simply observing gestures, or whether producing gestures at test can mediate the relationship between observing or producing gestures at training and performance at test.

1.5 Comparing Observation and Production of Iconic Gestures in Word Learning.

It has been proposed that there are perceptual, attentional, linguistic, spatial, memory and embodied/situated mechanisms underlying the effects of observing and producing gesture on learning (see Cook, 2018 for a review). Although the current paper cannot and does not attempt to distinguish between these mechanisms, motor imagery (see 1.3.1 above) and physical movement (see 1.4.1) may play key roles. A comparison of observing only (motor imagery) with both observing and producing gestures (motor imagery plus physical movement) will help to distinguish these mechanisms. Whether observing and producing gestures perform different functions and therefore have differing, or perhaps additive, effects, or they perform similar functions and therefore have similar effects, is not yet clear.

Faster processing time and enhanced memory performance have been found in action sentence recall in a native language (e.g., "lift the pencil") when participants produced the corresponding action than when they learned the sentences verbally (Cohen, 1981; Masumoto et al., 2006; Mohr, Engelkamp, & Zimmer, 1989). Action and gesture both involve physical movement, and iconic gestures can have a clear resemblance to actions. Producing actions and iconic gestures may, therefore, be similarly effective (Macedonia & Knösche, 2011).

Indeed, some studies have found beneficial effects of producing actions on recall of action sentences beyond effects of only observing an experimenter perform the actions (Engelkamp & Zimmer, 1997; Mulligan & Hornstein, 2003). Similarly, the work of Morett

(2018) suggested that the effects of producing gestures were stronger than those of observing gestures in L2 word learning (see 1.1 above). Conversely, other studies have found no beneficial effects of producing gestures beyond observing or imagining gestures (e.g., Baills, 2018; Brooks, Barner, Fran & Goldin-Meadow, 2018; Cohen, 1981; Kamermans, Pouw, Fassi, Aslanidou, Paas & Hostetter, 2019; Kormi-Nouri, 2000). Kormi-Nouri (2000) proposed that physical movement is not crucial, but rather that motor imagery, a mental simulation of performing the action, is sufficient to promote recall.

The Gesture as Simulated Action (GSA) framework holds that both observing and producing gesture elicits motor imagery (Hostetter & Alibali, 2008; see also Wu & Coulson, 2014, for a discussion of iconic gestures promoting image-based simulations). Motor imagery, therefore, occurs regardless of whether or not gestures are overtly produced, although the GSA framework does not preclude an additional beneficial effect of producing gestures beyond that of motor imagery alone. In sum, it is unclear whether it is motor imagery rather than physical movement that plays a role in the positive effect of gesture production beyond verbal only learning (Kormi-Nouri, 2000), or whether physical movement through producing gestures could enhance Japanese word learning beyond only observing gestures (Engelkamp & Zimmer, 1997).

1.6 Present Study

The present study examined the effects of observing and reproducing iconic gestures on learning Japanese verbs. It examined whether learning words with accompanying iconic gestures at encoding promotes learning beyond hearing the words without gestures, as well as whether there is a larger effect of reproducing an instructor's gestures, than of only observing the gestures, both in the short- and longer-term. As past studies indicate that the beneficial effect of gesture production takes time to emerge, we used a one week gap between two memory tests.

The study also examined whether observing or reproducing gestures at encoding affects learners' spontaneous gesture production during retrieval, and whether rates of spontaneous gesture production change with time. Iconic gestures are of particular interest given their conceptual relationship with concrete concepts such as the verbs presented in the current study. Finally, we examined the relationship between gesture observation, spontaneous gesture production at retrieval and recall.

It was expected that: 1) observing or reproducing iconic gestures at encoding would enhance word recall to a greater extent than speech only learning; 2) word recall would be greater for participants who reproduced iconic gestures at encoding than those who only observed gestures; 3) while recall would decrease with time, this decrease would be smaller for participants who observed or reproduced iconic gestures at encoding than those who received speech only; 4) spontaneous gesture production at retrieval would be greater for participants who observed or reproduced the instructor's gestures at encoding than those who received speech only and 5) a positive association between the number of spontaneous gestures produced at retrieval and words recalled would be found, as well as a positive mediating effect of gestures produced at test on the relationship between gesture condition at encoding and recall.

2. Method

2.1 Experimental Design

The present study was a 3 x (2) mixed-subjects design, with gesture condition as the between-subject factor and memory test at two time points as the within-subject factor. Participants were randomly allocated to one of the three gesture conditions: speech only (no gesture accompanying the spoken words), observe gesture (spoken words accompanied by the instructor's iconic gestures), or reproduce gesture (spoken words accompanying the instructor's iconic gestures as well as instructions to reproduce them). Each learning condition was followed by a test immediately following training (total 30-minutes duration for the initial session), and a delayed test one week later (10 minutes). The dependent variables were the number of correct words recalled and the number and types of spontaneous gestures produced by participants at recall.

2.2 Participants

Sixty-three participants (9 males, 54 females) were recruited through the Macquarie University Psychology Participant Pool (16 first year introductory psychology students, 44 second year cognition students). Participants ranged from 18 to 35 years (*M* = 20.51 years, *SD* = 3.49). One participant did not report their age. Participants provided written consent for the inclusion of their data in analysis and publication, and acknowledged that they and their fully anonymised data could not be identified via the paper.

All participants were native English speakers with no prior Japanese language learning. Forty five participants reported speaking a language other than English, with a mean self-rated competency level of 4.13 out of 9 (*SD* = 2.94; see 2.3.3 for details of the LOTE demographics questionnaire). After completing both parts of the study, participants were awarded course credit. Two participants did not complete the second part of the study and one participant did not correctly follow the experimental instructions. Data from these participants were excluded from analyses. A total of 60 participants' data was analysed, with 20 participants per condition.

2.3 Materials and Procedure for the Learning Stage

Ethics approval was obtained from the Macquarie University Faculty of Human Sciences Research Committee (Reference Code: 5201600134). All participants were individually tested in a quiet room on campus. Participants were randomly assigned to one of the three gesture conditions. Participants were told they would be learning Japanese verbs by watching the videos and their task was to remember as many words as possible.

2.3.1 Japanese words. The Japanese words used in the videos were common verbs with either two or three moras. The current study partially replicates that of Kelly et al. (2009), and the Japanese words and the word learning scripts used were used previously by Kelly et al. (2009). See Appendix A for a complete list of words.

2.3.2 Stimulus videos and training. All videos consisted of three learning blocks, with each of the 10 Japanese verbs presented in each block. To control for order effects, 15 versions of the video were created, all with the same female narrator. That is, five videos were created within each of the three gesture conditions for the purpose of counterbalancing the presentation of words. There was a one minute break between blocks. Words were presented in a predetermined random order in each learning block and all three blocks for each participant had different randomised word orders. Videos were created using iMovie for Mac (Version 10.1.1) from Apple Inc. The instructor was a native Japanese speaker to ensure words were presented with the correct accents.

In the videos for the speech only condition, the instructor kept her hands still and presented no gestures. In the videos for the observe and the reproduce conditions, the instructor produced iconic gestures accompanying presentation of the Japanese words. For example, an iconic gesture representing an action to drink (holding an imaginary glass and moving it towards the mouth) was simultaneously produced with the Japanese word “No-mu” (to drink). Total duration of the training phase was approximately 17 minutes.

At the beginning of the videos, short written instructions were presented on screen, outlining the experimental procedure. In the speech only condition, participants were instructed to verbally repeat the Japanese and English word pairs. In the observe condition, participants were instructed to verbally repeat the word pairs and observe the instructor’s gestures. In the reproduce condition, participants were asked to verbally repeat the word pairs as well as reproduce the instructor's gestures. The videos used in the observe and the reproduce conditions were identical except for the initial written instructions.

Each video had the same example word at the beginning as a practice trial. The instructor introduced the example word once, saying "The next word is Tsu-ne-ru" and continued by defining it: "Tsu-ne-ru means pinch", twice. Underlined words represent the point at which the iconic gesture was performed in the observe and the reproduce conditions. After the instructor defined each word, there was a five-second gap for participants to repeat the word pair. In the videos for the observe and the reproduce conditions, the instructor demonstrated the gesture once simultaneously with introducing the Japanese word, and twice in each defining sentence with both the Japanese and English words. This sequence of word presentation was identical throughout the videos.

All participants verbally repeated each defining word pair twice in each learning block, totalling six repetitions during the learning session. Participants in the observe and the reproduce conditions observed the gestures five times per Japanese word in each block, totalling 15 times during the whole word learning session. In addition, participants in the reproduce condition reproduced the gestures in each defining sentence simultaneously with watching the experimenter, totalling four times per Japanese word in each block, i.e. 12 times in total.

Participants in the speech only condition were initially only told to observe the videos, with no explicit instructions regarding their own gesture production. However, as our research question was in regards to the effects of observing or reproducing gestures during training, for the purposes of the cleanest comparison of observe and/or reproduce with speech only, if a participant in either the speech only or observe conditions began to spontaneously produce gestures during the practice session, they were verbally asked not to gesture. Two participants in the observe condition spontaneously produced gestures at the beginning of their session, and were asked only once not to gesture. Following this single instruction, neither participant gestured again. No participants in the speech only condition spontaneously produced gesture at any point during training.

2.3.3 Demographic questionnaire. Following training, participants completed a brief demographic questionnaire including age, gender and competency in speaking languages other than English (LOTE). Participants indicated their competency in speaking LOTE on a Likert scale ranging from 1 ("Not at all") to 9 ("Native level").

2.4 Materials and Procedure for the Testing Stage

Following the learning stage, participants completed the immediate memory test.

2.4.1 Verbal response items. All recall tests were conducted in verbal format. Participants were asked three free recall questions followed by 20 cued recall questions.

2.4.1.1 Immediate memory test: free recall. On completion of the demographic questionnaire, the experimenter asked the free recall question: "First, can you tell me as many Japanese words and their English translation word pairs you have learned from the video as possible, like Tsu-ne-ru means pinch?" After participants were unable to recall any more word pairs, the experimenter asked: "Would you like to list all the English words you have heard in the video? This might help you to recall more word pairs". On completion of this task, the experimenter asked participants: "Would you like to add any Japanese words? You may not recall a Japanese-English word pair, but do you recall any Japanese word by itself?" Participants received no feedback during any of the test phases.

The purpose of asking participants to recall the list of either English or Japanese words alone was to elicit recall of Japanese-English word pairs. The scores for these items were not included in the main analyses if only the Japanese or only the English word was recalled. If participants recalled a correct Japanese-English word pair while recalling English or Japanese words alone, the answer was marked as correct and included in their score.

2.4.1.2 Immediate memory test: cued recall. On completion of the free recall test, the experimenter asked 20 cued recall questions: "I am going to tell you a series of Japanese (or English) words. I would like you to tell me each English (or Japanese) translation" and verbally presented one word at a time. Ten questions were cued in Japanese to recall in English and 10 were cued in English to recall in Japanese. For example, participants were given a Japanese word (e.g., Tsu-ne-ru) and asked to recall its English translation (Pinch) and vice versa for English to Japanese cued recall. After completing 10 cued recall items in one language (e.g., from Japanese to English), 10 cued recall items in the other language (e.g., from English to Japanese) were presented, with the order of the two tasks counterbalanced between participants. Within each task the order of the words was randomised for each participant using an online random number generator (https://www.randomizer.org).

2.4.1.3 Delayed memory test. The second session was conducted one week after the first session. The procedure for the delayed memory test was identical to the immediate memory test. However, the order of language category (Japanese to English first or English to Japanese first) in the cued recall tests was reversed from participants' first memory test. On completion of the second session, all participants were given a brief explanation of the purpose of the study.

2.4.2 Gesture production. Participants were given no instruction about their gesture production at retrieval in both memory tests, as our primary interest was on spontaneous gesture production.

2.4.3 Recording devices. Participants' verbal responses and gesture production during the memory tests were audio and video recorded for subsequent coding.

2.5 Coding and Analyses

2.5.1 Word scoring. All correctly articulated and recalled Japanese-English word pairs for free recall, and Japanese and English words for cued recall were given a score of 1.

A score of 1 was given to the Japanese words when participants articulated them with the same number of moras and more than half of the phonemes were the same as the correct pronunciation. For instance, "Ha-shi-ru" (run) was learned during training, however a score of 1 was also given to participants who articulated it as "Ha-se-ru". Specifically, in order to obtain a score of 1: 1) the response had to contain the same number of moras as the correct word; 2) two of the three moras had to match if there were three moras in the correct word; 3) both of the two moras had to match if there were only two moras in the correct word and finally 4) the response had to end in the same mora as the correct word.

Non-response, incorrect response (e.g., Ha-shi-ru means drink) and incomplete word pairs (missing either Japanese or English) received a score of zero. Scores were summed to create a score out of 10 for each recall test (free recall and each cued recall test) for each of the immediate and delayed memory tests.

2.5.2 Gesture scoring. The number of gestures that participants spontaneously produced shortly before and simultaneously with the articulation of recalled words during each memory test was counted. A gesture was counted if it was clear that the participant intended the gesture to relate to the recalled word. There was no specific timing or other cut-off for a gesture to be coded; it could commence in silence or accompanied by speech. Gestures were categorised as either iconic, beat, deictic or metaphoric. Although gestures may take more than one form, in this case gestures were coded as iconic if they had an iconic element to them that corresponded with the word to be recalled. Beat gestures were those without any semantic relationship to the accompanying speech content, and which mirrored simple rhythmic hand movements in form. Any gestures which appeared to be grooming movements were discounted. All gestures were coded as distinct instances, including repetitions of the same gestures. That is, if a participant produced a "drinking" gesture twice, they received a score of two.

2.5.3 Reliability. A second coder (a native Japanese speaker) who was blind to the purpose of the study independently coded 25% of verbal responses and gestures produced during the two tests. Intra-class correlations (ICC) were obtained to evaluate reliability using an absolute agreement model. The single measure ICC was reported as only the first coder's scores were used for analysis. ICC scores were produced for free recall, Japanese to English and English to Japanese cued recall, the number of spontaneous gestures produced and their types in each memory test. All ICCs ranged between .926 and 1, with all $p < .001$.

2.5.4 Analysis Plan. As some dependent variables were not normally distributed, differences between experimental conditions on both verbal word recall and spontaneous gestures produced at test were analysed using two-way mixed models with 2000 bootstrapped samples. Orthogonal contrasts for the gesture condition factor tested 1) the observe and reproduce conditions combined compared with the speech only condition, and 2) the observe condition vs the reproduce condition.

The indirect effect of gesture condition on recall via gestures spontaneously produced at recall was analysed using a mediation analysis with gesture condition dummy coded as the independent variable, iconic gestures spontaneously produced at each test phase as the mediating variable, and recall at each corresponding test phase as the dependent variable. Gesture condition dummy codes indicated the observe and reproduce conditions, making the speech only condition the reference category. Standard errors and confidence intervals for the indirect effects were bootstrapped with 2000 replications. Iconic gestures were included if they took any form, either that displayed during the training phase, or a different iconic gesture taking an alternate form, but which nevertheless depicted the same verb.

3. Results

3.1 Preliminary analyses

Words that were not articulated with correct pronunciation but articulated to meet certain standards (see 2.5 above) were marked as correct. To evaluate the use of this standard, results were compared with separate analyses which applied stricter standards such that only complete correct articulation was coded as constituting accurate recall. The results for all inferential tests did not differ between the two criteria, therefore results using the original criteria are presented below. See Appendix B for descriptive statistics for recall scores using the stricter scoring protocol. It is worth noting that the observe and reproduce conditions appear to drop more in performance than the speech only condition with the application of the stricter scoring protocol. It is possible that co-speech gestures may interfere with phoneme processing. Kelly and Lee (2012) discuss how gesture may hinder language learning with high phonemic demands, and phonemic demands may be seen as higher with the stricter protocol. However, as performance in the speech only condition was much lower using the original scoring criteria than the other two conditions, scores in this condition had less capacity to fall, and firm conclusions regarding phonemic demands are not possible here.

A set of mixed models was run to examine the robustness of the observed effects at the level of the individual word. That is, are the effects of gesture and time equivalent across words? Six mixed-effects logistic regressions were run, with fixed effects of gesture condition and time (and the associated interaction), and random effects of word nested within participant. The dependent variables were scored as 1 for correct and 0 for incorrect for each individual word. We ran three models for each dependent variable (free recall, Japanese to English cued recall and English to Japanese cued recall) with a random slope for time that varied between words, and a further three models for each dependent variable with random slopes for gesture condition that varied between words. For all six models, the estimate of the random slope was as close to zero as could be estimated, meaning no confidence intervals or associated significance levels could be calculated. It appears therefore that the beneficial effects of gesture did not differ across words or across time.

A between-subjects ANOVA was conducted to examine differences in participants' competency in speaking a LOTE between conditions. There was no significant difference in participants' competence in LOTE between conditions, $F(2, 57) = 0.11$, $p = .892$, $\eta_p^2 < .01$. There were no significant correlations between participants' LOTE competency and their performance in each recall test, with Pearson's r ranging between -.04 and .02, all $p > .790$.

3.2 Main Analyses

3.2.1 The effect of gesture condition and time on recall. Three 3 (gesture condition: speech only, observed gesture, reproduce gesture) x 2 (time: immediate, delayed) bootstrapped mixed models were conducted to assess the effect of gesture condition and time on free recall and the two cued recall tests (see Table 1 for descriptive statistics).

Table 1

Mean and Standard Deviation Verbal Recall in all Recall Tests

	Gesture Condition							
	Speech		Observe		Reproduce		Total	
Variable	M	SD	M	SD	M	SD	M	SD
Immediate Test								
Free Recall	1.35	1.39	4.10	2.65	3.45	2.74	2.97	2.59
Cued Recall (J-E)	3.25	2.00	6.60	2.66	5.95	2.98	5.27	2.93
Cued Recall (E-J)	1.60	1.35	4.75	2.90	4.50	3.30	3.62	2.98
Delayed Test								
Free Recall	0.75	1.07	3.30	2.72	3.05	2.48	2.37	2.46
Cued Recall (J-E)	2.45	2.01	5.25	2.83	5.60	2.76	4.43	2.89
Cued Recall (E-J)	1.00	1.17	3.80	2.86	4.25	2.65	3.02	2.73

Note: Japanese to English (J-E), English to Japanese (E-J)

3.2.1.1 Free recall. There was a significant main effect of gesture condition, $\chi^2(2) = 102.47$, $p < .001$. The number of words recalled was significantly greater for participants in the observe and reproduce conditions on average than the speech only condition, $z = 10.10$, $p < .001$, 95% CI: 1.95, 2.90, but did not differ between the observe and the reproduce conditions, $z = 1.16$, $p = .25$, 95% CI: -0.31, 1.21. There was a significant main effect of time, $\chi^2(1) = 6.09$, $p = .014$, such that the number of words recalled was greater in the immediate than the delayed memory test. There was no interaction between gesture condition and time, $\chi^2(2) = 0.35$, $p = .842$.

3.2.1.2 Japanese to English cued recall. There was a significant main effect of gesture condition, $\chi^2(2) = 91.73$, $p < .001$. The number of words recalled was significantly greater for participants in the observe and reproduce conditions on average than the speech only condition, $z = 9.58$, $p < .001$, 95% CI: 2.39, 3.61, but did not differ between the observe and the reproduce conditions, $z = 0.35$, $p = .728$, 95% CI: -0.70, 1.00. There was a significant main effect of time, $\chi^2(1) = 6.58$, $p = .010$, such that the number of words recalled was greater in the immediate than the delayed memory test. There was no interaction between gesture condition and time, $\chi^2(2) = 1.43$, $p = .489$.

3.2.1.3 English to Japanese cued recall. There was a significant main effect of gesture condition, $\chi^2(2) = 131.66$, $p < .001$. The number of words recalled was significantly greater for participants in the observe and reproduce conditions on average than the speech only condition, $z = 11.41$, $p < .001$, 95% CI: 2.51, 3.54, but did not differ between the observe and the reproduce conditions, $z = 0.23$, $p = .820$, 95% CI: -0.96, 0.76. There was a significant main effect of time, $\chi^2(1) = 4.20$, $p = .04$, such that the number of words recalled was greater in the immediate than the delayed memory test. There was no interaction between gesture condition and time, $\chi^2(2) = 0.85$, $p = .653$.

3.2.2 The effect of gesture condition and time on gesture production at retrieval. Six 3 (gesture condition) x 2 (time) bootstrapped mixed models were conducted to assess the effect of gesture condition and time on spontaneous gesture production (iconic and beat) in each recall test (see Table 2).

Table 2

Mean and Standard Deviation Gesture Production in all Recall Tests

	Gesture Condition							
	Speech		Observe		Reproduce		Total	
Variable	M	SD	M	SD	M	SD	M	SD
Immediate Test								
Free Recall								
Iconic	0.00	0.00	1.30	2.00	3.70	2.54	1.67	2.40
Beat	0.05	0.22	0.00	0.00	0.30	0.66	0.12	0.42
Cued Recall (J-E)								
Iconic	0.00	0.00	0.65	2.03	1.95	2.76	0.87	2.11
Beat	0.30	0.66	0.10	0.31	0.45	1.19	0.28	0.80
Cued Recall (E-J)								
Iconic	0.00	0.00	0.80	1.47	2.90	1.23	1.23	2.17
Beat	0.15	0.37	0.50	1.10	0.35	0.81	0.33	0.82
Delayed Test								
Free Recall								
Iconic	0.00	0.00	0.85	1.31	2.70	2.18	1.18	1.84
Beat	0.05	0.22	0.10	0.45	0.40	0.88	0.18	0.60
Cued Recall (J-E)								
Iconic	0.00	0.00	0.50	1.28	1.70	2.32	0.73	1.67

Beat	0.30	0.73	0.20	0.52	0.50	0.89	0.33	0.73
Cued Recall (E-J)								
Iconic	0.00	0.00	0.85	1.79	3.05	3.10	1.30	2.41
Beat	0.10	0.31	0.30	0.92	0.10	0.31	0.17	0.59

Note: Japanese to English (J-E), English to Japanese (E-J)

3.2.2.1 Free recall. There was a significant main effect of gesture condition on iconic gesture production, $\chi^2(2) = 119.33$, $p < .001$. Participants in the observe and reproduce conditions on average produced more iconic gestures than those in the speech only condition, $z = 10.86$, $p < .001$, 95% CI: 1.75, 2.52. Participants in the reproduce condition in turn produced a greater number of iconic gestures than participants in the observe condition, $z = 5.51$, $p < .001$, 95% CI: 1.37, 2.88. There was no main effect of time, $\chi^2(1) = 2.62$, $p = .105$ and no interaction between gesture condition and time, $\chi^2(2) = 2.64$, $p = .267$.

There was a significant main effect of gesture condition on beat gesture production, $\chi^2(2) = 8.37$, $p = .015$. Participants in the observe and reproduce conditions on average produced more beat gestures than those in the speech only condition, $z = 2.34$, $p = 0.019$, 95% CI: 0.02, 0.28. Participants in the reproduce condition in turn produced a greater number of beat gestures than participants in the observe condition, $z = 2.72$, $p = .007$, 95% CI: 0.08, 0.52. There was no main effect of time, $\chi^2(1) = 0.57$, $p = .45$ and no interaction between gesture condition and time, $\chi^2(2) = 0.59$, $p = .745$.

3.2.2.2 Japanese to English cued recall.

There was a significant main effect of gesture condition on iconic gesture production, $\chi^2(2) = 40.04$, $p < .001$. Participants in the observe and reproduce conditions on average produced more iconic gestures than those in the speech only condition, $z = 6.12$, $p < .001$, 95% CI: 0.82, 1.58. Participants in the reproduce condition in turn produced a greater number of iconic gestures than participants in the observe condition, $z = 3.27$, $p = .001$, 95% CI: 0.50, 2.00. There was no main effect of time, $\chi^2(1) = 0.20$, $p = .659$ and no interaction between gesture condition and time, $\chi^2(2) = 0.20$, $p = .907$.

There were no significant main effects of gesture condition, $\chi^2(2) = 5.89$, $p = .053$, or time, $\chi^2(1) = 0.14$, $p = .707$, on beat gesture production. There was no interaction between gesture condition and time, $\chi^2(2) = 0.18$, $p = .915$.

3.2.2.3 English to Japanese cued recall.

There was a significant main effect of gesture condition on iconic gesture production, $\chi^2(2) = 88.54$, $p < .001$. Participants in the observe and reproduce conditions on average produced more iconic gestures than those in the speech only condition, $z = 9.38$, $p < .001$, 95% CI: 1.50, 2.30. Participants in the reproduce condition in turn produced a greater number of iconic gestures than participants in the observe condition, $z = 5.54$, $p < .001$, 95% CI: 1.39, 2.91. There was no main effect of time, $\chi^2(1) = 0.05$, $p = .818$ and no interaction between gesture condition and time, $\chi^2(2) = 0.05$, $p = .974$.

There were no significant main effects of gesture condition, $\chi^2(2) = 3.72$, $p = .156$, or time, $\chi^2(1) = 1.70$, $p = .193$, on beat gesture production. There was no interaction between gesture condition and time, $\chi^2(2) = 0.84$, $p = .659$.

3.2.3 The association between gesture production at retrieval and words recalled.

A series of mediation analyses as outlined above assessed the indirect effects of gesture condition on recall, mediated by spontaneous iconic gesture production, for each measure of recall (free recall, Japanese-English and English-Japanese cued recall), at both immediate and delayed recall. See Table 3 for a full list of effects. All direct and indirect effects can be found in Table 3.

Table 3

Coefficients, Standard Errors (SE) and Confidence Intervals for Mediation Analyses of Gesture Condition via Spontaneous Gesture Production on Recall

		Immediate				Delayed			
DV	Effect	Coeff-icient	SE	Lower CI	Upper CI	Coeff-icient	SE	Lower CI	Upper CI
Free Recall	Observe on iconic (direct)	1.30*	0.58	0.17	2.43	0.85	0.45	-0.04	1.74
	Reproduce on iconic (direct)	3.70*	0.58	2.57	4.83	2.70*	0.45	1.81	3.59
	Iconic on recall (direct)	.61*	0.14	0.33	0.88	0.29	0.19	-.08	.067
	Observe on recall (direct)	1.96*	0.66	0.67	3.26	2.30*	0.69	0.95	3.65
	Reproduce on recall (direct)	-0.14	0.82	-1.75	1.47	1.50	0.84	-0.15	3.16
	Gesture condition on recall (total direct)	1.82	1.31	-.74	4.38	3.80*	1.34	1.17	6.43
	Observe on recall (indirect)	0.79*	0.34	0.25	1.65	0.25	0.23	-0.10	0.84
	Reproduce on recall (indirect)	2.24*	0.73	1.01	3.88	0.80	0.74	-0.47	2.52
	Gesture condition on recall (total indirect)	3.03*	0.95	1.27	5.09	1.05	0.93	-0.65	3.12
	Observe on iconic (direct)	0.65	0.61	-0.55	1.85	0.50	0.47	-0.42	1.42

Japanese-English	Reproduce on iconic (direct)	1.95*	0.61	0.75	3.15	1.70*	0.47	0.78	2.62
	Iconic on recall (direct)	-0.42*	0.16	-0.73	-.10	-0.43*	0.21	-0.84	-0.03
	Observe on recall (direct)	3.62*	0.76	2.13	5.11	3.02*	0.77	1.51	4.52
	Reproduce on recall (direct)	3.51*	0.82	1.91	5.11	3.89*	0.84	2.24	5.54
	Gesture condition on recall (total direct)	7.13*	1.37	4.45	9.81	6.91*	1.40	4.17	9.65
	Observe on recall (indirect)	-0.27	0.24	-0.91	0.02	-0.22	0.17	-0.72	0.00
	Reproduce on recall (indirect)	-0.81*	0.40	-1.61	-0.27	-0.74*	0.34	-1.49	-0.15
	Gesture condition on recall (total indirect)	-1.08*	0.47	-2.09	-0.26	-0.96*	0.45	-1.99	-0.19
English-Japanese	Observe on iconic (direct)	0.80	0.56	-0.29	1.89	.85	0.64	-0.40	2.10
	Reproduce on iconic (direct)	2.90*	0.56	1.81	3.99	3.05*	0.64	1.80	4.30
	Iconic on recall (direct)	-0.11	0.19	-0.48	0.26	-0.16	0.15	-0.45	0.12
	Observe on recall (direct)	3.24*	0.83	1.61	4.86	2.94*	0.73	1.51	4.37
	Reproduce on recall (direct)	3.21*	0.98	1.28	5.14	3.75*	0.84	2.10	5.40
	Gesture condition on recall (total direct)	6.45*	1.58	3.36	9.54	6.69*	1.37	4.01	9.36
	Observe on recall (indirect)	-0.09	0.22	-0.61	0.29	-0.14	0.18	-0.69	0.08
	Reproduce on recall (indirect)	-0.31	0.73	-1.46	1.34	-0.50	0.53	-1.50	0.58

	Gesture condition on recall (total indirect)	-0.40	0.93	-1.91	1.69	-0.64	0.68	-1.96	0.68

NB: Standard errors for indirect effects are bootstrapped, and confidence intervals are bias-corrected. The speech only condition is the reference category for the dummy codes for gesture condition. * $p < .05$

3.2.3.1 Immediate recall.

Replicating the analyses in 3.2.2, there were significant positive direct effects of the observe and reproduce conditions on gesture production at recall. These positive direct effects combined with the effects of gesture production on word recall in two very different ways. First, there was a significant positive direct effect of iconic gesture production on immediate free recall. The positive direct effects of gesture condition on gesture production, combined with the positive direct effect of gesture production on free recall, resulted in significant positive indirect effects of observe vs speech only and reproduce vs speech only on free recall via spontaneous gestures produced.

Conversely, there was a significant negative direct effect of iconic gesture production on cued Japanese-English recall. The positive direct effects of gesture condition on gesture production noted above, combined with the negative direct effect of gesture production at recall on recall performance, resulted in a significant negative indirect effect of reproduce vs speech only via spontaneous gestures produced. There was no significant direct effect of iconic gesture production on cued English-Japanese recall, and no significant indirect effects.

***3.2.3.2 Delayed recall*.**

There was no significant direct effect of iconic gesture production on delayed free recall, and no significant indirect effects. As with immediate recall, there was a significant negative indirect effect of reproduce vs speech only via spontaneous gestures produced on cued Japanese-English recall. This negative indirect effect was a product of a positive direct effect of reproducing gestures at encoding on spontaneous gesture production at test, and a negative direct effect of gesture production on recall. There was no significant direct effect of iconic gesture production on cued English-Japanese recall, and no significant indirect effects.

4. Discussion

The current study examined the role of observing and reproducing gesture in foreign word learning. It examined whether only observing or reproducing an instructor's iconic gestures impacted foreign word learning, as well as whether the difference in the effect of observing and reproducing iconic gestures at encoding on recall differed across time. The study examined whether only observing or reproducing iconic gestures at encoding affected learners' gesture production during retrieval, and whether gesture production changed with time. Finally, the study explored whether gesture condition affected recall indirectly, mediated by gestures produced spontaneously at test.

4.1 The Effect of Gesture Condition at Encoding on Word Recall

Consistent with past studies using foreign (Allen, 1995; Baills et al., 2018; Huang et al., 2018; Kelly et al., 2009) and artificial languages (Macedonia & Klimesch, 2014; Macedonia & Knösche, 2011), a greater number of words were recalled by participants who either observed or reproduced the instructor's iconic gestures at encoding than participants who learned the words by speech alone in all memory tests. The present finding was in line with the assertion that iconic gestures aid learners to ground the link between novel words and their meanings in learners' native languages (Kelly et al., 2009; McNeill, 1992).

Conversely, the current result contrasts with that of Krönke et al. (2013) of no recall difference of pseudowords with or without accompanying iconic gestures. This inconsistency could be due to differences in task difficulty. Krönke et al. (2013) argued that the beneficial effect of gesture may be reduced if the task is comparatively easy, mirroring Riseborough (1981). Participants in the 2013 study verbally repeated 42 words 21 times across three days, whereas participants in the current study verbally repeated each of 10 Japanese words six times in one day. Increased repeated rehearsal may have made the earlier task easier than the current task, perhaps preventing the detection of gesture effects.

4.2 Observation vs Reproduction of Gesture on Word Recall

Consistent with previous findings which demonstrated no recall difference between learning through only observing or producing gestures (Kormi-Nouri, 2000), there was no recall difference between participants who only observed gestures during learning and those who reproduced iconic gestures during learning. Kormi-Nouri (2000) proposed that the non-significant results were due to observing the actions produced by an instructor eliciting motor imagery. The present finding supports the view that physical movement is not essential for the beneficial effect of gestures to enhance recall over speech only learning.

The GSA framework states that motor imagery is elicited when people observe or spontaneously produce gestures (Hostetter & Alibali, 2008). In the current study, gestures that participants produced at encoding were not spontaneous, but rather were a reproduction of the instructor's gestures. Motor imagery may have been elicited from observing the instructor's gestures rather than planning to produce gestures. If motor imagery plays a role in the beneficial effect of both only observing as well as producing gestures in learning, then the current result suggests that reproducing an instructor's gestures may be no more beneficial than observing the same gestures.

However, if the production of gesture enhances learning by reducing conceptualisation demands on working memory, these beneficial effects may be more pronounced when conceptualisation demands are higher. For example, Engelkamp and Zimmer (1997) and Mulligan and Hornstein (2003) used action sentences rather than single words. In De Nooijer et al. (2013), children reproduced gestures accompanying verbs depicting unfamiliar concepts. These tasks may have had higher conceptualisation demands than the current task, which required participants to reproduce gestures for foreign, but highly familiar, words. Future studies should compare the effects of reproducing vs observing gestures with words/sentences of varying difficulty (e.g., concrete vs abstract verbs).

4.3 Interaction Between Gesture Condition and Time on Word Recall

Although as expected, the number of words recalled decreased with time, contrary to expectation, this decrease did not differ between conditions. This lack of an interaction suggests that the beneficial effect of observing or reproducing gestures over speech only learning does not change with time. The present finding indicates that both only observing and reproducing an instructor's iconic gestures at encoding can equally aid learners to recall learned words compared with speech only learning in both the short and longer term.

4.4 The Effect of Gesture Condition and Time on Gesture Production at Retrieval

As predicted, participants who either observed or reproduced iconic gestures at encoding spontaneously produced a greater number of iconic gestures at retrieval than participants who did not see any gestures. Similarly, spontaneous iconic gesture production at retrieval was greater for participants who reproduced iconic gestures at encoding than participants who only observed them. In contrast, there were fewer consistent group differences in the number of spontaneous beat gestures produced at retrieval. This is of particular interest, as participants neither had beat gestures modelled to them during training, nor were given any explicit instruction regarding beat gestures in any condition.

The finding that observing iconic gestures at encoding increased participants' spontaneous iconic gesture production at retrieval was in line with that of Cook and Goldin-Meadow (2006) that observing a teacher's gestures during learning encouraged children to produce their own gestures at test. The lack of consistent effects for beat gesture production at retrieval also supports the finding of Cook and Goldin-Meadow (2006) that learners' spontaneous gesture production often reflects the gestures they observed during learning.

The current findings suggest that observing or reproducing an instructor's iconic gestures encourages learners' iconic gesture production at retrieval. A greater production of iconic gestures in participants who reproduced the instructor's gestures compared to those who only observed or those who saw no gestures indicates that previously seen beneficial effects of producing gestures during learning may at least partially be indirect: it may encourage learners' spontaneous gesture production at retrieval, which in turn enhances recall. Although the lack of a significant difference in recall seen between the observe and reproduce conditions implies that physical movement is not a prerequisite for the beneficial effects of gesture on learning as noted above, this does not necessarily mean that learning *cannot* be enhanced by movement. In the current study however, the mediating relationship of iconic gesture production on word recall is notably less clear.

4.5 Mediating Role of Spontaneous Gesture Production on Word Recall

The expected positive indirect effect was found for gesture condition on immediate free recall, mediated by spontaneous gesture production. This finding suggests that as well as observing gestures at encoding benefiting free recall, observing or reproducing gestures at encoding prompted more spontaneous gesture production at recall, which in turn had a positive effect on recall. No significant indirect effects were found for delayed free recall.

Highly unexpected, however, were the significant negative indirect effects of the reproduce condition vs the speech only condition on recall, via iconic gestures produced for cued Japanese to English recall, at both immediate and delayed test. A similar pattern of results, albeit non-significant, was found for cued English to Japanese recall. This unexpected finding suggests that spontaneous gestures at test may have hindered recall in the cued test phases. Issues surrounding co-speech gestures interfering with phonemic processing in tasks with high phonemic demands, as outlined above in 3.1, are possible (Kelly & Lee, 2012). In Japanese to English cued recall, participants were presented with the correct pronunciation of the Japanese words, by the native Japanese speaking experimenter. If these pronunciations did not exactly match those encoded in memory, the phonemic demands of the task may have increased relative to the free recall phase, resulting in the detrimental effect of gesture production. This interpretation is tentative however, in light of the similar pattern of results seen for English to Japanese cued recall (albeit non-significant), where participants were not presented with the correct Japanese pronunciation, but rather with the English word.

It is possible that the positive direct effect of iconic gesture production on free recall performance is more an artefact of the nature of the free recall task: participants who recalled more words would be speaking for longer, and therefore had more opportunities to gesture. This effect will not be present for cued recall however, as all participants were cued with all 10 words. The interference effect noted above of the co-speech gestures therefore may only have appeared for these tasks of fixed length, and been masked for free recall by the confound of task length. Future research should consider examining the proportion of times that each participant gestured for words correctly vs incorrectly recalled. Examining whether people gesture more (or less), as a proportion, for correctly vs incorrectly recalled words, will allow for better interpretation of whether the movement of the gesture aids learning.

These mediation analyses should be interpreted with caution however, as both the mediator (gestures produced) and dependent (recall) variables were measured simultaneously. It is possible here that a bi-directional relationship may exist between gestures produced and recall: gestures spontaneously produced may affect recall, but perhaps recall could also affect gestures spontaneously produced.

4.6 Mechanisms and Implications

While there is not yet a comprehensive framework for explaining the mechanisms of gestures accompanying spoken messages for learners, the current study provides some clarification for how gestures might benefit learners. The present study supports the GSA view that gesture benefits learning by eliciting motor imagery (Hostetter & Alibali, 2008). Dual coding theory and multimedia learning theory argue that gestures accompanying speech benefit learners in the construction of mental representations of speech content, making learning accessible via verbal and nonverbal connections (Mayer, 1997, 2009; Paivio, 1990).

Gestures may help bridge the gap between novel words and their meanings, facilitating information processing, as well as providing learners with referential connections to facilitate recall of novel foreign words.

Theories of embodied cognition suggest that externalising thought processes through the production of gesture can facilitate cognitive processes (Wilson, 2002). The present study does not provide clear support for the view that producing gesture enhances learning, with some positive and some negative effects across analyses. It is possible that the task of consciously reproducing the instructor's iconic gestures may have minimised or negated any benefits to cognitive processes of producing gestures. Producing gestures may help word retrieval by cross-modally activating a word's concept due to the external representation of the word when gesture is self-generated or spontaneous, but not when the gesture itself needs to be recalled as well as the word.

4.7 Limitations and Future Directions

The current study involved a condition in which participants were asked to mimic the gestures produced by the instructor, with no condition in which participants produced their own gestures. Only the cumulative effects of gesture production and observation could be examined therefore, with no direct comparison of gesture production only and gesture observation possible. Future research should replicate the current findings, but include a condition in which participants are instructed to gesture during learning but with no instructor gestures to mimic, to enable a direct comparison of gesture production, gesture observation, and the cumulative benefit of gesture production and observation together.

Future research would similarly benefit from a closer examination of the individual iconic gestures the narrator produced in the stimulus videos. The gestures produced were the iconic gestures deemed by a native Japanese speaker to be the most commonly produced gesture in Japan. As the effect across all gestures on recall of all verbs was of primary interest for the current study, rather than the effects of individual gestures on recall of individual verbs, no norming of the videos was conducted. Similarly, there was no strict control of single/dual handedness of each gesture, or of the expanse of each gesture. Recent research has suggested that not all iconic gestures are equally beneficial for learning (Dargue & Sweller, 2018). Future research should more closely examine the effects of individual gestures and distinct types of iconic gestures on recall of individual verbs.

Finally, it is notable that only verbs were used in the current study. It would be of interest to investigate concrete nouns, adjectives, and concepts that could be represented through metaphoric gestures. Recent research has shown significant effects of observing gesture on narrative recall for narratives that included a variety of key terms and gesture types (e.g., Dargue & Sweller, 2019), but whether such effects extend to adjectives and metaphoric gestures, or to foreign word learning, remains an open question.

4.8 Conclusions

The present study demonstrated that while either observing or reproducing the instructor's iconic gestures enhanced foreign word learning, reproducing the instructor's iconic gestures did not enhance foreign word learning beyond only observing the same gestures. Reproducing the instructor's gestures during word learning increased the rate of participants' spontaneous iconic gesture production at retrieval, however such gesture production at test had varying effects on recall. While a positive indirect effect of gesture condition via gesture production at test was found for free recall, a negative indirect was found for cued Japanese to English recall. Findings suggest that observing iconic gestures may be used to benefit adult learners in foreign word acquisition, but that the role of gesture production itself is somewhat more complex.

5. References

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*, 593-613. doi:10.1080/016909600750040571

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal, 79*, 521-529. doi:10.1111/j.1540-4781.1995.tb05454.x

Baills, F., Suárez-Gonzáles, N., Gonzáles-Fuente, S., & Prieto, P. (2018). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*. Advance online publication. doi:10.1017/S0272263118000074

Brooks, N. B., Barner, D., Frank, M., & Goldin-Meadow, S. (2018). The role of gesture in supporting mental representations: The case of mental abacus arithmetic. *Cognitive Science, 42,* 554–575. doi:10.1111/cogs.12527

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22*, 267-281. doi:10.1111/j.1467-9450.1981.tb00402.x

Cook, S. W. (2018). Enhancing learning with hand gestures: Potential mechanisms. *Psychology of Learning and Motivation, 69*, 107-133.

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. doi:10.1207/s15327647jcd0702_4

Dargue, N., & Sweller, N. (2018). Not all gestures are created equal: The effects of typical and atypical iconic gestures on narrative comprehension. *Journal of Nonverbal Behavior*, 1-19. doi:10.1007/s10919-018-0278-3

Dargue, N., & Sweller, N. (2019). Learning stories through gesture: Gesture's effects on child and adult narrative comprehension. *Educational Psychology Review*, 1–28. doi:10.1007/s10648-019-09505-0

Dargue, N., Sweller, N., & Jones, M.P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin, 145*, 765. doi:10.1037/bul0000202

De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta psychologica, 144*, 173-179. doi:10.1016/j.actpsy.2013.05.013

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta psychologica, 96*, 43-60. doi:10.1016/S0001-6918(97)00005-X

Gluhareva, D., & Prieto, P. (2016). Training with rhythmic beat gestures favors L2 pronunciation in discourse-demanding situations. *Language Teaching Research, 21*, 609-631. doi: 10.1177/1362168816651463

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*, 1-19. doi: 10.1515/langcog.2010.001

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516-522. doi:10.1111/1467-9280.00395

Gullberg, M. (1998). Gesture as a communication strategy in a second langauge discourse: A study of learners of French and Swedish. Lund, Sweden: Lund University Press

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Homage to Adam Kendon). *IRAL-International Review of Applied Linguistics in Language Teaching, 44*, 103-124. doi: 10.1515/iral.2006.004

Gullberg, M. (2014). Gestures and second language acquisition. In Muller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., & Tessendorf, S. (Eds.), Body, language, communication: An international handbook on multimodality in human interation (Vol.2, pp1868-1875). Berlin, Germany, and Boston, MA: Mouton De Gruyter

Gullberg, M., deBot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture, 8*, 149-179. doi: 10.1075/gest.8.2.03gul

Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second langauge vowel length contrasts. *Journal of Speech, Language, and Hearing Research, 57*, 2090-2101. Doi: 10.1044/2014_jslhr-s-14-0049

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*, 297-315. doi:10.1037/a0022128

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*, 495-514. doi:10.3758/pbr.15.3.495

Huang, X., Kim, N., & Christianson, K. (2018). Gesture and Vocabulary Learning in a Second Language. *Language Learning.* doi: 10.1111/lang.12326

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*, 617-637. doi:10.1017/s0272263102004047

Kamermans, K. L., Pouw, W., Fassi, L., Aslanidou, A., Paas, F., & Hostetter, A. B. (2019). The role of gesture as simulated action in reinterpretation of mental imagery. *Acta Psychologica, 197,* 131–142. doi:10.1016/j.actpsy.2019.05.004

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. doi:10.1080/01690960802365567

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes, 27*, 793–807. doi:10.1080/01690965.2011.581125

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*, 1-11. doi:10.3389/fpsyg.2014.00673

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review, 124*, 245–266. doi:10.1037/rev0000059

Kormi-Nouri, R. (2000). The role of movement and object in action memory: A comparative study between blind, blindfolded and sighted subjects. *Scandinavian Journal of Psychology, 41*, 71-76. doi:10.1111/1467-9450.00173

Krönke, K.-M., Mueller, K., Friederici, A. D., & Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex, 49*, 2553-2568. doi:10.1016/j.cortex.2012.11.016

Macedonia, M., Bergmann, K., & Roithmayr, F. (2014). Imitation of a pedagogical agent’s gestures enhances memory for words in second language. *Science Journal of Education, 2,* 162. doi:10.11648/j.sjedu.20140205.15

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. doi:10.1111/mbe.12047

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education, 5*, 196-211. doi:10.1111/j.1751-228X.2011.01129.x

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human brain mapping, 32*, 982-998. doi:10.1002/hbm.21084

Macedonia, M., & von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics, 6*, 393-416.

Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., & Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain research, 1101*, 102-109. doi:10.1016/j.brainres.2006.05.033

Mayer, R. E. (1997). Multimedia learning:are we asking the right questions? *Educational Psychologist, 32*(1), 1-19. doi:10.1207/s15326985ep3201_1

Mayer, R. E. (2009). *Multimedia learning*. Second edition. Cambridge: University Press.

McCafferty, S. G. (2002). Gesture and creating zones of proximal development for second language learning. *The Modern Language Journal, 86*, 182-203. doi: 10.1111/1540-4781.00144

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Mohr, G., Engelkamp, J., & Zimmer, H. D. (1989). Recall and recognition of self-performed acts. *Psychological Research, 51*, 181-187. doi:10.3758/BF03202759

Morett, L., (2018). In hand and in mind: Effects of gesture production and viewing on second language word learning. *Appled Psycholinguistics, 39,* 355-381. doi: 10.1017/S0142716417000388

Mulligan, N. W., & Hornstein, S. L. (2003). Memory for actions: Self-performed tasks and the reenactment effect. *Memory & cognition, 31*, 412-421. doi:10.3758/BF03194399

Paivio, A. (1990). Mental representations: A dual coding approach. New York, NY: Oxford University Press.

Pouw, W., de Nooijer, J., van Gog, T., Zwaan, R., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology, 5*, 359. doi:10.3389/fpsyg.2014.00359

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226-231. doi:10.1111/j.1467-9280.1996.tb00364.x

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5,* 172-183. doi:10.1007/BF00986134

Rowe, M., Silverman, R., & Mullan, B. (2013). The role of pictures and gestures as nonverbal aids in preshcoolers' word learning in a novel language. *Contemporary Educational Psychology, 38*, 109-117. doi:10.1016/j.cedpsych.2012.12.001

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture, 8*, 219-235. doi:10.1075/gest.8.2.06tel

Wilson, M. (2002). Six views of embodied cognition. Psychonomic Bulletin & Review, 9, 625-636. doi:10.3758/bf03196322

Wu, Y., & Coulson, S. (2014). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychologica, 153,* 39-50. doi:10.1016/j.actpsy.2014.09.002

Appendix A: Japanese Word List

Japanese words and English translations

Item number	Japanese Word	English Translation
Practice	Tsu-ne-ru	Pinch
1	Hi-ku	Pull
2	Ha-na-su	Talk
3	Ta-be-ru	Eat
4	O-su	Push
5	Ka-ku	Write
6	No-mu	Drink
7	Ki-ru	Cut
8	Yo-mu	Read
9	Ha-shi-ru	Run
10	Na-ge-ru	Throw

Appendix B: Descriptive statistics for strict scoring protocol

Mean and Standard Deviation Verbal Recall Scores for the Strict Scoring Protocol in all Recall Tests

	Gesture Condition							
	Speech		Observe		Reproduce		Total	
Variable	M	SD	M	SD	M	SD	M	SD
Immediate Test								
Free Recall	1.25	1.37	3.50	2.59	3.05	2.46	2.60	2.38
Cued Recall (J-E)	3.20	1.99	6.55	2.72	5.95	2.98	5.23	2.95
Cued Recall (E-J)	1.30	1.22	4.15	2.50	4.00	3.26	3.15	2.77
Delayed Test								
Free Recall	0.65	0.99	2.70	2.47	2.70	2.11	2.02	2.16
Cued Recall (J-E)	2.50	1.96	5.10	2.81	5.60	2.76	4.40	2.85
Cued Recall (E-J)	0.65	0.99	3.00	2.77	3.80	2.31	2.48	2.51