Check for
updates

The Embodied Teaching of Spatial Terms: Gestures Mapped to Morphemes Improve Learning

Natasha Janzen Ulbricht*

Department of Philosophy and Humanities, Freie Universität Berlin, Berlin, Germany

Learning spatial terms in a second language is often an arduous task which learners perform with varying levels of success. While classroom-based studies of gesture have shown the importance of embodied learning, predictions about which teaching gestures are most effective remain rare. In the context of learning and performing a play, this study investigates two English language teaching methods, one with teacher gestures at the level of morphology and one with gestures at the sentence level. This experiment with a diverse group of primary-school-age children from Germany and Poland ($N = 76$) shows that although over time both groups made similar gains in understanding and using spatial terms, this gain was more immediate for learners exposed to one gesture per morpheme. For beginning learners spatial terms are frequent, important and abstract, hence this research may have important implications for understanding the nature of effective methods for teaching and testing abstract concepts.

Keywords: second language acquisition, testing procedures, spatial terms, gesture, linguistic units, abstract concepts

INTRODUCTION

When observing the position or trajectory of objects in space, we are usually unaware that categorical distinctions are imposed on the scene. However, talking about movement and position requires that space be divided into discrete basic spatial categories. While this process may seem effortless in a language we know well, learning to use spatial terms in a second language (L2), where space may be partitioned very differently is often a difficult task. At the same time, the semantic categories associated with words like *in*, *on*, *under*, and *to* are highly relevant for describing not only objects, actions, and events but also creating narrative space (Lütke, 2011). Moreover, in addition to relating real or imaginary scenes, physical spatial configurations also lead to abstract non-spatial meanings (Lakoff and Johnson, 1999; Tyler and Evans, 2003). Clearly understanding the notion of physical support in *Your keys are on the table* makes understanding the implied offer of emotional support in *You can count on me* much easier. For beginning second language learners, spatial terms are frequent and important. Effective teaching methods for spatial language are thus essential for second language acquisition.

Gestures Play an Important Role in Learning and Teaching

As humans because of our physical and neurobiological architecture, we perceive objects and actions in certain ways. Gestures or symbolic hand movements can represent this conceptual information through form and movement (McNeill, 1992; Stokoe, 2000). During interaction with children, adults regularly combine objects, actions and words, and seem to intuitively recognize that gesture may scaffold children's understanding (Kang et al., 2015; Rohlfing et al., 2016). And in

fact children are often better able to understand spoken messages when these are accompanied by meaningful gestures than when linked to conflicting or no gestures (Goldin-Meadow et al., 1999). Researchers have reported that seeing gestures promotes cognitive development (e.g., McGregor et al., 2008; Cook et al., 2010) and L2 word learning (Macedonia et al., 2019) and that when words and body movements are used in combination, this leads to better retention (Kiefer et al., 2007; Arndt and Sambanis, 2017; Sambanis and Walter, 2019). It has been suggested that gesture used in combination with speech may reduce cognitive demands on processes of learning by allowing two different representational systems, both visual and verbal, to share the load (Goldin-Meadow, 2000; see Pouw et al., 2014 for an overview).

Researchers have recently proposed the Gesture-for-Conceptualization Hypothesis (GfCH) which states that gestures schematize information and are conceptually linked not only to speaking, but also to thinking in general (Kita et al., 2017). Observing gestures triggers semantic processing (Wu and Coulson, 2007; Kelly et al., 2009) and related to L2 learning, iconic gestures can allow linguistic units, such as a new L2 word, to be unambiguously connected to a hand movement (see also Huang et al., 2019). This connection decreases the need for semantic aspects of language comprehension, which allows the brain to save these resources for additional information processing, possibly leading to more robust learning and better retention (Skipper, 2014; Hupp and Gingras, 2016). Zwaan and Radvansky (1998: p. 177) have suggested that words and sentences can be understood as instructions for creating a mental representation of the described situation. More recently, Brouwer et al. (2012) have proposed the term mental representation of what is being communicated (MRC) for the internal representation a listener or reader constructs while comprehending a sentence, story, or scene. They further specify that MRCs are derived not only directly from linguistic input, but also from inferences made on the basis of logical, causal, or pragmatic world knowledge (2012: 136). It follows that if in addition to patterns available in speech, gestures make it easier for a listener to construct a correct MRC, this would translate into more efficient mental processing. If meaningful gestures enable learners to update their MRC with less effort and more clarity, learning would be less tied to contextual familiarity and more prone to consolidation.

Related to mental representations, the notion of embodied simulation has been proposed citing research which demonstrates that both physical and imagined manipulation lead to substantial gains in memory and language comprehension (Glenberg, 2011; de Koning et al., 2017). Although different from our everyday integrated perception, human cognitive neuroscience shows that at any given moment only fragments of scenes are available to consciousness, these being guided and filtered by the demands of attention and task relevance (Cichy and Teng, 2017). Following this line of thinking, gestures at the sentence level, where one hand movement corresponds with an entire sentence, could allow more time for learners to simulate the scene connected to the gestures leading to an increase in understanding.

Despite the fact that the benefits of gesture for second language learning are well-documented (Macedonia and von Kriegstein, 2012; Hattie and Yates, 2013; Arndt and Sambanis, 2017), the mechanisms by which gesture facilitates learning are not fully understood. Neuroscientific research shows that perceptual and lexical-semantic spatial information have a parallel organization in the brain (Göksun et al., 2013) and that simple gestures can make meaningful differences in how complex language is understood (Holle et al., 2012), however, the relationship between speech, gesture and language comprehension is complex. Some research suggests that under certain circumstances, for example when cognitive demands are high or skill level is low, gestures may disrupt comprehension (McNeil et al., 2000; Kelly, 2017). Gesture theory, as outlined in the GfCH, makes predictions about the supportive effects of gestures for learning, but how to best use gestures in L2 classrooms is under-researched, leaving many questions unanswered.

While the relationship between gesture and L2 teaching and learning has been examined, few studies have operationalized spatial term learning in classroom settings, and even fewer with primary-school-age learners. This research gap is unfortunate because although L2 spatial language is clearly important, it is often perceived by teachers as challenging to teach (Lütke, 2011). Qualitative and quantitative studies relevant to classroom-based English language spatial term learning are reviewed and summarized in **Table 1**.

This paucity of research raises several more general issues. Knowing meaningful gestures tied to a word or sentence has been shown to enhance learning, however learning gestures in addition to speech initially increases cognitive demands (Macedonia and Klimesch, 2014). Students learn more when their teachers gesture effectively (Alibali et al., 2013), however predictions about which gestures are effective are rare. Iconic gestures, which have a "close formal relationship to the semantic content of speech" (McNeill, 1992: p. 12) have been shown to be beneficial, but there are different kinds of iconicity (Perniss and Vigliocco, 2014). How children mentally represent conceptual information changes over time (Kelly, 2017), suggesting that development might influence which gestures are most effective. Further, as the MRC concept suggests, a substantial amount of the information we use to determine meaning is not associated with a single lexical item (Foster, 2001; Knoeferle et al., 2010). In this article we do not ask if gestures per se "help." For this the interested reader is referred to reviews by Macedonia and von Kriegstein (2012) and Cook (2018); (see also Dargue et al., 2019 for a recent meta-analysis of gesture and comprehension). Building on past research, rather we ask if evidence exists that gestures which connect specific linguistic units with specific hand movements should rather be at the sentence or the morphological level. Researchers have previously called for experiments with more specific predictions about which gestures will support learning and precisely when these gestures will be helpful (Roth, 2001; Alibali et al., 2013; Cook, 2018), and in doing so have specifically mentioned the variable of linguistic units as relevant (Gullberg, 2013: p. 1872).

To shed more light on this issue, a recent study investigated the influence of teacher gestures on oral fluency in a diverse group of primary school age children (Janzen Ulbricht, 2018). The experiment implemented two methods of teaching English,

Researchers	Participants	Study objective
Johansson Falck (2018)	9 Swedish pupils 12–13 years	Effect of learners applying body-world knowledge categories for *in* and *on* to L2 learning
Nakatsukasa (2016)	48 ESL university students Mage = 20.4 years	Effect of teacher gestured corrective feedback on learner locative preposition production for *above, under, in, on,* and *next to*
Eskildsen and Wagner (2015)	An adult Mexican Spanish-speaking learner of English, his classmates and teacher	To investigate how common L2 gesture-speech combinations are deployed by teachers and reused within the classroom by learners to facilitate production and understanding for *under* and *across*
Rumme et al. (2008)	97 Japanese pupils Mage = 12.1 years	Effect of teacher abstract pointing gestures on preposition distinction learning between *on-under, next to-between, in front of-behind,* and *near-at*

one with teacher gestures at the level of morphology, and one with gestures at the sentence level plus the written text. This experiment showed a difference in long-term fluency gain between the experimental conditions among the high and low performers. Here it was observed that children with a lower initial speech rate benefit more from gestures at the level of morphology, while children with an initially higher speech rate benefit more from reading plus sentence-level gestures, suggesting that the initial fluency level of learners is predictive of which type of gesture benefits fluency the most. One limitation of the previous study was that in the measure of oral fluency used (speech rate), all syllables, regardless of word or phrase complexity, were treated equally.

The present study extends this research, and examines in more detail the role of these same teaching conditions in learning English spatial terms. Since gesture has the potential to embody spatial information, gesture may be especially helpful for teaching spatial terms, as has been explored by others in L1 (McGregor et al., 2008) and L2 learning (Eskildsen and Wagner, 2015; Nakatsukasa, 2016; Ahlberg et al., 2018). Understanding how spatial language performance in one domain contributes to the development of performance in another may lead to findings that can enhance educational practice. As outlined in the GfCH (Kita et al., 2017), gesture theory makes predictions about the supportive effects of gestures on learning, but guidelines about which gestures teachers should use remain underspecified. At a symbolic level gestures can be paired with different units of language. As such, gestures at the sentence level provide an interesting comparison to gestures at the level of morphology and allow us to identify the circumstances under which gestures which vary in this way may be differentially beneficial to classroom-based learning. While not the only valid approach to classroom research, experiments involving complete teaching methods are essential because they can establish how different elements, such as gesture type and access to text, work in combination. Thus, such experiments can provide more ecologically valid grounds for generalization than experiments which differ in one variable alone.

The present study reports the results of a 7-week experiment that tested the effects of gesture-based L2 instruction on long-term spatial term learning. Children from two primary schools, one in Germany ($n = 29$) and one in Poland ($n = 47$), were tested on their use of English spatial terms in week 1, week 3, and week 7 to measure initial learning and retention. In week

2 of the experiment, two sets of matched codified gesture (CG) and scenic learning (SL) text-learning phases were designed for a common English theater project. While learning the play (for a total of 3 h over 4 days), the children were randomly placed in the CG or SL conditions where they learned and memorized the same text[1]. To control for teacher effects, two teachers at each school taught the same text to both groups in each condition. In the codified CG group, the teacher provided one gesture per morpheme for all the words of the play, meaning that words, and gestures were learned together. Consistent with the SL method, the teacher taught the children the play supported by gestures at the sentence level and the written text. The sample size ($N = 76$) was based on convenience, but as can be inferred from **Table 1**, is above the mid-range value of similar experiments.

Background on Gestures in the Experiment

Codified gestures refer to specific hand or arm movements which have a "dictionary meaning" within a particular group (Poggi, 2013). This group can have many members, such as the number of people who understand the European *What an idiot* forehead tap. This group can also be as small as the students of a particular teacher who has a special sign to prompt using the past tense. Codified gestures may be iconic, such as meaning *fire* by wiggling fingers to suggest flames, but may also be arbitrary, as when tapping the back of the right hand into the palm of the left to represent *dlaczego* meaning *why* in Polish sign language. Although there are important differences between codified gestures and the hand movements which make up sign languages (see McNeill, 1992; Crystal, 2007), compared to spoken language, sign languages have more potential for iconic forms because they are produced with the hands, face and body (Perniss and Vigliocco, 2014). When meaningful hand movements are combined with new words, learners may benefit since gestures can be perceptually similar to the object or event being referenced and can add semantic information, which in turn can prime lexical representations (Roth, 2001). We should note that in gesture studies there is wide agreement that hand movements

---

[1]In this experiment, as in others, variability in participant characteristics may affect individual learning outcomes. While it is known that linguistic and socioeconomic variables often influence language learning processes (Krifka et al., 2014), these confounding variables are commonly dealt with by randomly assigning participants to experimental conditions to ensure even distribution across conditions (e.g., Novack et al., 2014). Children in this experiment were also randomly assigned to the experimental groups.

	Codified gesture condition	Scenic learning condition
A hand movement	Per morpheme	For sentences
Gestures for sample sentence: It's dark out there.	It + is + dark + out + there	It is dark out there [prep **stroke**] stroke = *open hands pass in front of the face covering the eyes*
Gestures in sample sentence	Total number = 5	Total number = 1
The play was learned	**Without** written text	**With** written text
Time in experimental conditions	12 sessions of 15 min long spread over 4 days	

can be categorized into different subtypes (Kita et al., 2017). Although the gestures used in this study could be categorized in other ways (e.g., McNeill, 1992), the term codified gestures has been used to emphasize the one-to-one relationship between movement and meaning. We should also note that research on L2 learning has used different terms for similar movements-meaning relationships at some times simply referring to gestures (e.g., Goldin-Meadow, 2000; Cook, 2018) and at others creating novel terms such as Voice Movement Icons or VIMs (Macedonia, 2020). In summary, experimental conditions in this experiment were different in that teacher hand movement referred to fixed morphemes (e.g., {rock} + {- s}) and were the only form of input in the CG condition, and referred to fixed sentences (e.g., *Let's get out of bed!*) where learners had access to the written text in the SL condition. Conditions were the same in that both used fixed movement-meaning pairs to reinforce learning.

All of the spatial terms tested were embedded in the text of the play (for testing materials and procedures, see section Instruction). Consistent with stories and the English language in general (Crystal, 2007), some words were more frequent than others. O*ut* (as in *out of bed* and *out the window)* was mentioned seven times; *over* five times; *under* and *in* four times; whereas *around*, *between* and *through* were used three times. *To* (as in *Let's go to the window!*) and *on* (*the owl was sitting on a tree*) were only mentioned once. This difference in frequency, because inherent in the text, was the same for both experimental conditions. To conclude, there were two experimental conditions, as shown in **Table 2**.

Research Questions

Much research on gesture and L2 learning has focused on whether gesture-based instruction benefits learners. These experiments, while necessary, lack the precision necessary to provide guidance on which gestures might support learning best. With this study we move beyond this question by testing the effects of teaching methods involving different teacher gestures at the level of linguistic units on spatial term learning outcomes. We hypothesize that during second language acquisition gestures can support the mental representation of what is being said (MRC), reducing uncertainty and resulting in more efficient language processing. We make no prior claims about one condition being more efficient than another. Matched codified gesture and scenic learning units for beginning English learners were developed and their effects on L2 spatial term learning were tested. Following a repeated-measures design, which quantifies changes over time,

analyses of a gain in spatial term ability were carried out. This study is consistent with the premise that meaning is embodied and that learning occurs as a result of collaboration with others in familiar socially constructed settings (Bruner, 1983; Tomasello et al., 2012; Rohlfing et al., 2016) and addresses the following research questions:

1. In the context of learning and performing a play, can a long-term gain in L2 spatial term ability be measured?
2. If the same text is learned in different ways, using a gesture for every word without the written text (CG) or using a gesture for the most important sentences with access to the written text (SL), are there measurable differences between experimental groups?

MATERIALS AND METHODS

Participants

Our study was conducted with 76 learners between the ages of 8 and 13 from two primary schools ($M = 10.9$ years, $SD = 0.96$, 42 females), one in urban Germany and one in rural Poland. In both locations, the instruction during week 2 was a week-long joint theater project, in Germany between members of a grade 5 class ($n = 19$) and a class of refugee children ($n = 10$) from the same school, and in Poland between two different grade 5 classes ($n = 21$) and between two grade 6 classes ($n = 26$). Of the grade 5 German children, 15 (79 percent) identified an L1 other than German as their primary home language. All the refugee children had an L1 other than German as their home language. At the time of the study the refugee children had spent between 1 month and 3 years in Germany, but 9 (90 percent) had been in Germany for <2 years. In Poland all children reported Polish as their primary home language. All children reported having previously learned English in Germany, Poland or in their country of origin. Polish and German children began learning English in school in grade 3, meaning grade 5 learners were in their third year and grade 6 learners in their fourth year of English instruction. Refugee children reported between 1 and 3 years ($M = 1.7$, $SD = 0.95$) of instruction. Children who participated had submitted written consent from their parents prior to the study and agreed to participate.

Instruction Materials

Two sets of text-learning phases were developed, each resulting in a total of 3 h of instruction. The content of the play to be taught during the project was segmented into 12 units of 15 min

FIGURE 1 | Sample gestures for the sentence It is dark out there in the codified gesture (CG) condition.

each. For each teaching phase both a version that utilized scenic learning (SL) forms of instruction and a codified gesture (CG) version of instruction were designed. As previously mentioned, in the SL condition the focus of the first six units was on understanding and fluently reading the play, whereas sessions 7–12 focused on using sentence-level gestures to speak together as a group and memorize the character parts.

Codified Gesture Condition

In both the CG and the SL conditions, the children had instruction in which they separately learned the same text. In the CG condition, the teachers taught a set of gestures, one for every morpheme in the play. In this condition, most words such as *under* had a single gesture, but some words such as *bears* had two gestures, one for {bear} and one to show the plural {-s}. The children were seated in a semicircle facing the instructor throughout all text learning phases. While reading the text, the teacher spoke and gestured the play, meaning that words and hand movements were learned simultaneously. (For sample gestures in the CG condition, see **Figure 1**). The children were instructed to speak as soon as they recognized a gesture, but were not instructed to gesture. In Germany once the children could recognize and speak the words, they began to imitate the accompanying gestures. In Poland, although given the same

instructions, surprisingly, the children in the CG grade 6 group hardly gestured. Because the focus of this experiment is on the effects of teacher gestures on spatial language learning and children are compared to themselves, this difference, although interesting, does not influence our results[2].

Scenic Learning Condition

Scenic learning is an approach which combines movement and choral repetition of words, lexical chunks, or sentences. These movements, although simple, reinforce associations between words and mental images or scenes taken from daily life, hence the name scenic learning (Böttger and Sambanis, 2017: p. 62). In previous classroom-based experiments the scenic learning approach has shown an advantage over traditional teaching methods for both vocabulary and pronunciation (Hille et al., 2010). Because the focus of the current experiment was not on

---

[2]When asked to use gesture participants often produce responses that are more strategic and thoughtful (Hattie and Yates, 2013: p. 142). Especially in group learning situations, however, there can be pedagogical reasons for encouraging but not requiring learners to perform certain behaviors (Sambanis and Walter, 2019). Given the short time teaching time and diverse learners in this experiment (refugee learners), pedagogical reasons were the decisive factor in modeling and thus encouraging but not requiring learners to perform gestures. In most groups (all in Germany and all in Poland except the mentioned grade 6 CG group) classroom observers indicated that learners reliably gestured of their own accord.

whether gesture-based instruction is beneficial to learners but compares two different gesture-based methods, the SL condition was adapted. In this condition, the emphasis of the first six sessions was on understanding and fluently reading the text. Children were initially told to relax, close their eyes and listen to the teacher read, and listen for words they recognized. After listening, the teachers were instructed to work through the text using techniques they had found successful in the past, such as reading the play in roles and in small groups.

While the text of the play remained the same, in contrast to the first six sessions, the focus of sessions 7–12 in the SL condition was on using gestures at the sentence level to memorize and practice speaking together. Following the SL approach, the most central sentences of the play were practiced accompanied by a simple movement. These movements were developed by the teachers at each school to capture the meaning of the most important sentences of the play. As can be seen in **Table 2**, the SL gesture for the sentence *It is dark out there* consisted of a single hand movement. This movement corresponded to the gesture for the word *dark* in the CG condition and is depicted in the dark (beginning) and dark (end) pictures in **Figure 1**. While in the CG condition, all of the words were matched with gestures. In the SL condition, excluding the narrator parts, 78 percent of the words of the play belonged to sentences matched with gestures.

In both Poland and Germany, it was clarified that the goal of practice was for all children to memorize each speaking part independent of the role they would eventually play in the actual performance. In the SL group children had access to the text in written form, but only during the text-learning phases. After the final text-learning phase, the CG and the SL groups were combined at the grade level (meaning grade 5 and grade 6 worked separately), character roles were assigned and a narrator from each group was chosen. For the final 5 h of instruction, the focus moved from learning the text to rehearsing the play on stage in an artistic way. Because of this different focus, during the rehearsal and performance children did not gesture. This is practice of using and then discontinuing gestures once learners have internalized the target language is also consistent with other L2 gesture-based teaching methods (e.g., Macedonia, 2020).

Instruction

Each teacher taught both groups of students in both conditions, with no more than two consecutive sessions being taught by the same teacher. This design allowed for the control of teacher effects. To facilitate continuity of instruction in the SL condition, teachers created lesson plans of the activities in advance. In the CG condition, teachers provided gestures for all the words of the play and wrote brief notes in the teaching materials to document which text sections had been covered. Fidelity of implementation observers were present in each classroom ∼60 percent of the time to ensure that the text was taught as designed in terms of timing, content, and activities. Observers were instructed to note any deviation from the lesson plan as well as any differences in gesture quality within conditions and recorded only little deviation. It is also important to note that before beginning teaching sessions all teachers were tested to ensure gesture proficiency and consistency.

Testing Materials

The stimuli consisted of five objects (teddy bear, box, ball, blanket, and a book) on a table in a room with a chair, window, and a door. Some of the test items were functionally canonical in that the trajector object (e.g., a ball) would commonly go in the landmark object (e.g., a box) in everyday environments. However, many of the test items such as *Put the ball in the blanket.* or *Move the blanket through the chair.* were non-canonical. These items were included in order to determine whether the experimental training phases (learning the text of the play) enabled a less context-dependent understanding of spatial terms. When test items were trialed, combinations which were deemed possible but especially confusing, (e.g., *Put the table on the bear*) or physically difficult, (e.g., *Put the chair on the table*) were removed from the sentence set.

At the beginning of the study, a test using a set of objects not required during subsequent teaching was administered to all initial participants. (For access to online-**Supplemental Materials** and for the actual tests, see the notes section at the end of this article). Retention was measured with follow-up visits the week following instruction and 5 weeks following instruction. In both schools teachers of participating classes were trained in both sets of instructional gestures (∼90 min of training plus access to the filmed gestures) and passed a test before they administered instruction in week 2. In Germany the author administered the baseline and both follow-up tests. In Poland, two teachers of the same school administered the tests. All teachers involved in the project were unaware of the study hypotheses and were only informed that the study aimed to test the effectiveness of gestures for second language learning.

The format of the baseline and both follow-up tests was the same and used three different but equivalent versions of the same test. The test objects used (bear and ball etc.) were the same for each test version, but the order of the spatial terms and the items required for a certain action were randomized and different. Using different but equivalent test versions follows the parallel-forms method for matching statistical reliability (Murphy and Davidshofer, 2005; Hilger and Beauducel, 2017). The order in which the three different test versions (Tests A, B, and C) were administered for the pretest, post, and retest was counterbalanced across all participants.

The format of all testing sessions was a warm-up phase, Part A in which the child heard nine recorded sentences and performed the associated actions, and Part B in which the examiner performed nine actions and the child spoke, meaning each spatial term was tested twice, once in Part A and once in Part B. The test also included part C which was deliberately designed to be difficult to avoid ceiling effects and to make retention challenging. However, since there was no evidence of ceiling effects for parts A and B across participants and sessions, data from part C was collected but is not included in the analysis. Because we see both L2 spatial term comprehension and production as closely related skills, for data analysis scores from part A and B were combined into a general accuracy score (Novack et al., 2014). The testing session lasted 15–20 min. PsychoPy Experiment Builder (v1.84.2)
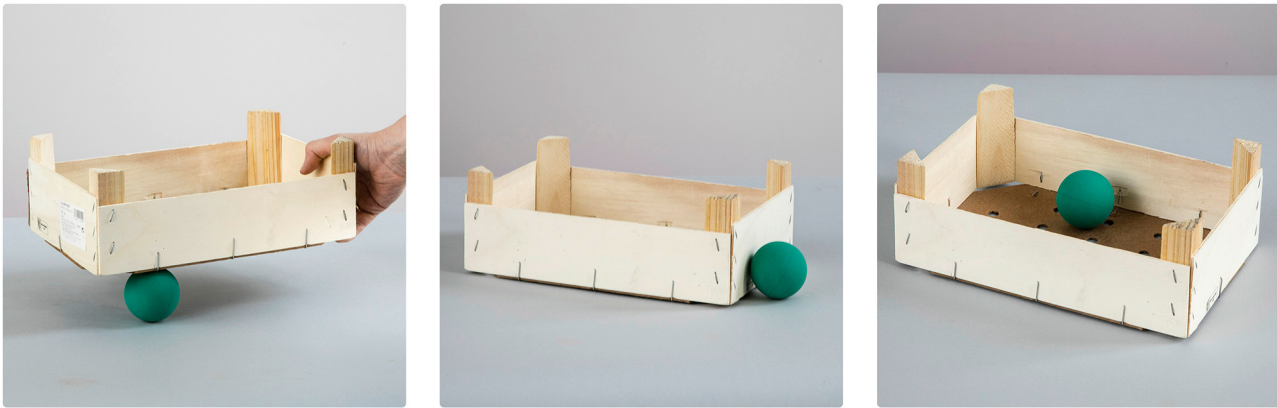
FIGURE 2 | Coding examples for test item put the ball under the box. The first picture is correct, the second and third are incorrect.

was used to create and run the test sessions (Peirce, 2009) meaning that children in Germany and Poland both heard the same instructions.

Testing Procedures

Warm Up
Children first completed a warm-up phase to familiarize themselves with the room and the test objects, as well as speaking with the experimenter. This warm-up phase was scripted and involved each child repeating the name of the test objects and physically touching them.

Part A
The first section of the test was about understanding and implementing action statements by moving or positioning objects in physical space (see **Figure 2**). Test items were only played once. Performance was measured in the following way:

- If a child complied with the action statement, they received one point.
- If a child did not comply with an action statement, and did not make a movement, but did make eye contact, the examiner said, "Just do the best you can."
- If a child did not comply with an action, make any movement, or make eye contact, after 10 s the examiner said, "Just try the next one." and the next recording was played.
- If a child made an action that was incorrect, they did not receive a point and the next recording was played.

Part B
The second section of the test was about recognizing actions and naming the position of objects in physical space. For the sentence *Put the ball under the box.* the instructor said, "Here is the ball. Here is the box." The instructor then did the action, put the ball under the box and asked, "Where is the ball?" and noted what the child said. For sentences using *around*, *out,* and *through* the experimenter asked, "Where did the [object] go?"

Performance was measured in the following way:

- If a child named the correct spatial term, they received one point.
- If the child demonstrated understanding in movement (e.g., through a spontaneous gesture or repeating a gesture from the training phase) or a language other than English, they did not receive a point.
- If a child named an incorrect spatial term, they did not receive a point and the next recording was played.

Children themselves were not given any feedback about whether or not an answer was correct, but were thanked for their participation at the end of the test. Exit interviews for all children established that in general children enjoyed the test. Even children who received no points for spatial term knowledge, reported feeling successful because they had recognized and spoken English words and in conclusion many said the test "wasn't hard."

Removing Outliers
Van den Broeck et al. (2005: p. 967) write that in research "error-prevention strategies can reduce many problems but cannot eliminate them" sometimes making data cleaning a necessity. During first inspection of the data from the first school, *between* and *through,* two of the nine initial words, were identified as unusually difficult, with baseline correct answers for *through* missing entirely from one experimental group in this school. The word *out* was also removed, but for other reasons. Unlike other spatial terms, enacting an *out* command (e.g., *Put the blanket out of the box.*) requires implicit knowledge of *in.* If the blanket happens to be in the box, the same test item becomes easier than if the blanket is not in the box, which introduced additional variability into the test procedure for this particular item. Data for *between, through,* and *out* were removed, meaning three of the nine original spatial terms. This same procedure was followed for both schools. This reduced the total number of test items from 18 to 12 per test and resulted in eight percent of the data for which participants would have received a point being cleaned during analysis. Cronbach's alpha is a summary measure of the correlations between items and can

be used as a measure of test reliability. The overall alpha was 0.76 with the mean correlation among the test items being 0.21. This is above 0.70, the level often considered satisfactory for exploratory research.

Data Analysis

We conducted multiple regression analyses on long-term comprehension and use of L2 spatial terms to test the long-term effects of learning a text using two English language teaching methods, one with teacher gestures at the level of morphology without access to the written text (CG), and one with gestures at the sentence level with access to the written text (SL). Our binary dependent variable (correct vs. incorrect responses on the spatial term test) was analyzed using a multilevel modeling approach. We used a hierarchical model including class and preposition as random effects with students nested within classes. Experimental group and session, meaning the time point when the tests were conducted, were included as fixed effects. All analyses were conducted with R Version 3.4.3 with the lme4 package (Bates et al., 2015). We compared each model with updated versions of the model that systematically excluded the main effect and interaction terms of interest.

RESULTS

Data Description

Our analysis of student outcomes includes 76 students who completed all assessments and for whom a questionnaire was received about their age, years of English language tuition, and whether the primary home language was the language of school instruction. Because of data privacy laws, while it was possible to ask if a child's L1 was or was not the language of instruction (i.e., German in Germany or Polish in Poland), it was not permitted to ask what a child's L1 was. As noted above, knowledge of English spatial terms was tested before the project began. For each participant, an accuracy score (i.e., number correct on test) was calculated. Preliminary analyses indicated that there were no significant effects or interactions found for gender or age, $p's >$ 0.05, so these variables were removed from further analyses. To test for a possible effect of location (Poland vs. Germany) on the gain in spatial terms, we ran an ANOVA with school and experimental group as a between groups factor, which showed no interaction between schools, $F_{(1,72)} = 0.47$, $p = .49$, so this variable was also removed. The primary analysis yielded the same pattern of results whether or not the language of instruction was a learner's L1 or L2, so all reported analyses include all participants. An independent-samples t-test (number of correct spatial terms by experimental group) compared the mean scores of the two experimental groups[3]. The initial mean number of correct spatial terms for the CG group was $M = 4.28$ ($SD = 2.80$) and for the SL group $M = 4.86$ (3.12), $t_{(73.19)} = -0.85$, $p = 0.39$ two-tailed, indicating that the groups are comparable.

---

[3] All analyses were conducted with R Version 3.4.3 with two-tailed tests using $p >$ 0.05 for null hypothesis rejection.

Long-Term Gain in Spatial Term Use

After the text-learning phases, experimental groups were combined (in Germany into one group and in Poland at the grade level) and the final 5 h of teaching time were used to focus on presenting the play on stage in an artistic way. Given that the children had learned and practiced an adventure story which contained spatial language, but that the focus of the performance had passed, it was unknown whether spatial term comprehension and production would improve on the test. The posttest took place in the week following the final presentation of gestures, followed by the retest 7 weeks after the initial test and 5 weeks after the theater project. Comparing the two experimental groups in **Figure 3**, our first analysis demonstrated successful learning across both conditions. Our first research question asks if a long-term gain in L2 spatial term ability can be measured and can be answered through visual inspection of **Figure 3**.

Differences Between Experimental Groups

Figure 3 shows children's mean spatial term ability organized by mean number correct, time, and learning condition and shows that both experimental groups improved over time, as demonstrated by an increase in mean accuracy in both conditions, but with a higher gain in spatial term ability for the CG condition. The mean gain in spatial term ability (post – pre) for the CG condition was $M = 3.52$ ($SD = 2.28$) and for the SL group $M = 1.86$ (2.00), $t_{(72.73)} = 3.36$, $p = 0.001$ two-tailed, $d = 0.77$, indicating that the experimental groups the children belonged to had a significantly different effect. Because this gain was calculated as a per-child variable, any gain in ability measures how children compare to themselves, so cultural or first-language differences among children cannot influence our results.

To further investigate these differences, children's spatial term ability (correct vs. incorrect responses) was entered in a hierarchical model including class and preposition as random effects, with students nested within classes. Experimental group and session were included as fixed effects.

```
m = glmer(result ~ exp_group*session
+ (1|preposition) + (1|class/code), bb,
family=binomial)
```
The next model excluded the interaction of group and session.
```
m0 = glmer(result
~ exp_group+session + (1|preposition) +
(1|class/code), family=binomial, data=bb)
```
Comparing the results of the two models summarized in **Table 3** allow us to see that the fit of the model with the interaction between experimental group and session is slightly favored.

As can be seen from the output of the first model (see **Supplementary File**), the interaction between the experimental group and session appears to be specific to the second time point or posttest in session 2. Based on **Figure 3**, this interaction is to be expected. Learners in the CG condition improve more between the first two testing sessions ($p = 0.013$ *), but then between session two and three students in both conditions appear to have similar knowledge at the final test ($p = 0.491$ ns).

Summary

These results in L2 spatial term learning show that while there are enhancements for both experimental groups and
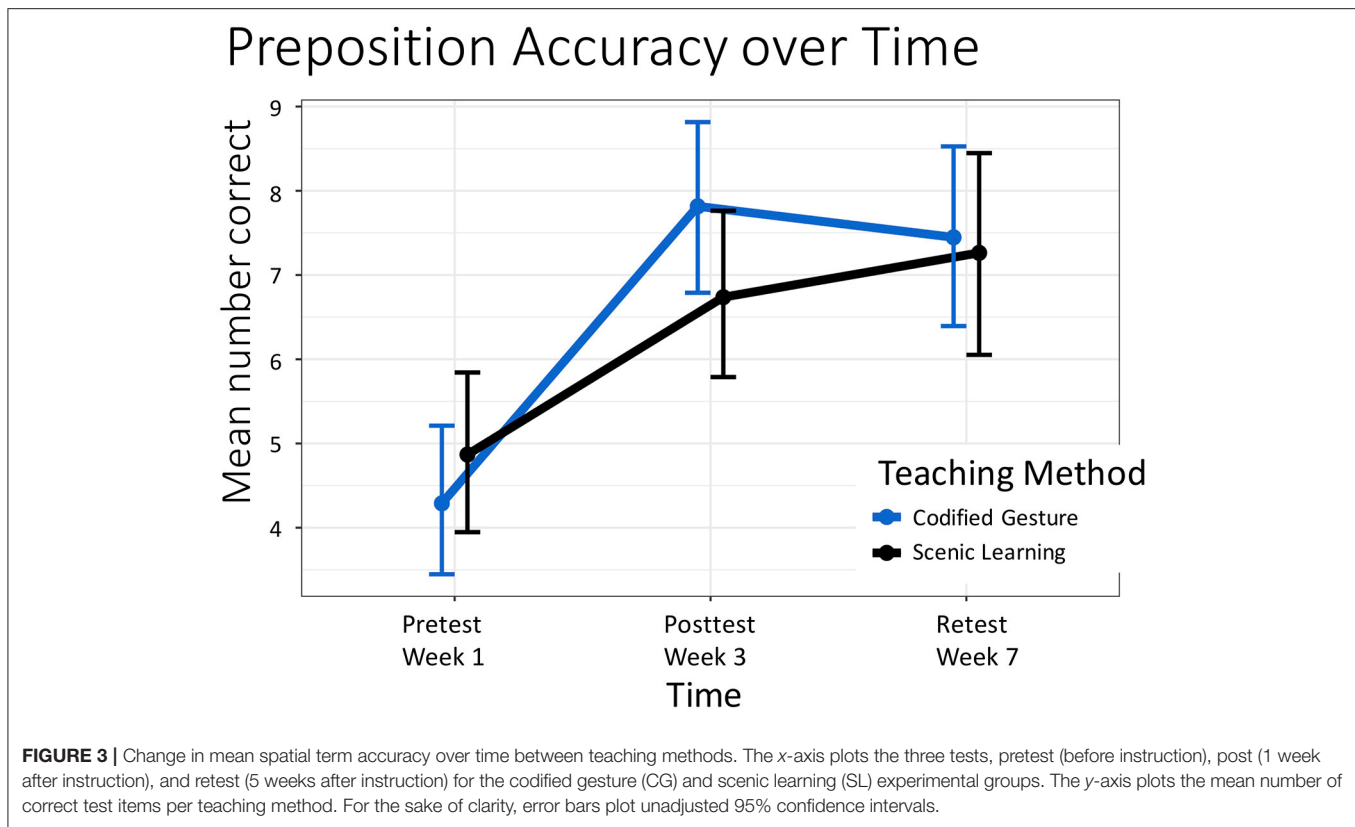
FIGURE 3 | Change in mean spatial term accuracy over time between teaching methods. The *x*-axis plots the three tests, pretest (before instruction), post (1 week after instruction), and retest (5 weeks after instruction) for the codified gesture (CG) and scenic learning (SL) experimental groups. The *y*-axis plots the mean number of correct test items per teaching method. For the sake of clarity, error bars plot unadjusted 95% confidence intervals.

TABLE 3 | Summary of model fit statistics.

	Df	AIC	BIC	log Lik	Deviance	Chisq	Chi Df	*P*(>Chisq)
m	9	3060.217	3113.266	−1521.109	3042.217	6.59	2	0.04
m0	7	3062.809	3104.069	−1524.404	3048.809	NA	NA	NA

both lead to long-term learning processes as indicated by the retest measurement, the CG condition appears to be the initially more efficient learning procedure. The error bars for the retest, especially for the SL group, indicate more variation in learning, meaning differences between experimental groups become much less clear over time. Especially for learning which is new, this suggests that teaching over time is important in order to consolidate what has been learned (Kelley et al., 2018).

DISCUSSION

Through work on cross-linguistic categories of spatial relations, Brala (2002: p. 135) concludes that categories of functional configurations are formed and organized into meaning clusters "on a combinatorial basis, out of universal, primitive, bodily-based semantic features … [which are] shared between the human language faculty and other sub-systems of human cognition." This means that while different languages may treat spatial categories differently, there is an underlying implicit

"logic" to how these categories are formed. These categories have been found to influence compatibility effects between language processing and action or perception and provide behavioral evidence that how spatial terms are used in different languages not only "matter" in terms of correct usage, but "matter" in terms of how space is mentally represented, which can be very different across languages (Bowerman, 1996). It has been previously established that spontaneous gestures schematize information in language-specific ways (Kita and Özyürek, 2003). Thus, attention to embodied teaching methods relevant to these language specific categories could potentially benefit learning, because, as Bowerman (1996) suggests, successful L1 and L2 acquisition depends on learning to attend to these topological relationships. This experiment compares two different teaching methods. Because of the naturalistic nature of this experiment (interaction effects), there are limitations to the direct conclusions one can make based on certain teaching elements. Because English was presented in two modalities (reading and gestures in the scenic learning condition and gestures alone in the codified gesture condition), no direct claims about gesture or writing based on these results can be made. Additional studies with different paradigms are required to investigate whether different gesture types independent of reading can also facilitate L2 spatial term learning. Nonetheless, the differences in spatial term learning over time raise certain questions worth investigating. Before addressing two additional questions, we would like to return to our original research questions:

1. In the context of learning and performing a play, can a long-term gain in L2 spatial term ability be measured?
2. If the same text is learned in different ways, using a gesture for every word without the written text (CG) or using a gesture for the most important sentences with access to the written text (SL), are there measurable differences between experimental groups?

Regarding question one, visual inspection (**Figure 3**) and the main effect of test on spatial term ability described in the results section suggest that in both groups the benefits of learning and performing a play featuring L2 spatial terms can be measured. Note that the results shown here cannot be separated from any possible benefit (or detriment) of performing the test itself. This transfer of concept learning from one context (learning and performing in a group setting) to another (speaking and moving objects as an individual during the test) is in line with research which shows that neglecting movement as a learning strategy leaves a particularly important source of support for learning under-utilized (Sambanis and Walter, 2019: p. 8). Moving on to question two, the difference in spatial term gain between the pre and posttest demonstrates that within the children in these schools, there was a measurable difference between teaching methods with an effect size of $d = 0.77$, which, when rounded to 0.80, is considered a strong effect (Cohen, 1988).

The two additional questions we would like to address are:

1. Why is the CG condition more efficient?
2. What else is learned in the SL condition?

Because gestures on the level of morphology were the only input form in the CG condition, children in this condition saw more gestures. On the part of the teachers, producing more gestures meant more practice, possibly leading to more gesture consistency. In support of this viewpoint, observers also remarked on an increase in the gesture quality over time. Gesture practice also improved in the SL condition, but here, because there were simply fewer gestures, this effect would be expected to be less.

Although the Retrieval-Integration account of language processing is largely based on language data, Gunter et al. (2015) extended it to gesture processing making this model more widely applicable. Their experiment showed that incongruent abstract pointing leads to higher retrieval and integration effort as reflected in increased N400 and P600 amplitudes. Although only indirect support, these results suggest that the reliable teaching gestures present in both teaching conditions could directly influencing sentence comprehension and possibly learning. When presented at the same time, speech, and gesture appear to encourage learners to simultaneously attend to and integrate ideas conveyed in the two modalities and thus create long-lasting and more flexible new concepts (Novack et al., 2014). Perhaps a "cleaner" gesture signal in the CG condition or one gesture per spatial term allowed for more consolidation in a shorter time.

The question about what else was learned in the SL condition is difficult to answer. Other experiments using SL have shown positive long-term effects, but in these experiments the teaching time was considerably longer and was compared to teaching methods which were not embodied (Hille et al., 2010). Teaching in the SL condition involved reading and gestures on the sentence level for memorizing the text. The SL teaching method also has certain advantages in terms of planning, because outside of an experimental setting, gestures can be spontaneous. It is also conceivable that being a part of a scene and "being in the moment" has emotional advantages that the CG condition, which is more closely tied to the actual text might not have. Actually moving in the scene could support learning not measured by the test. In addition, reading supports learning and is a familiar activity.

In previous experiments when measuring fluency (Janzen Ulbricht, 2018), practice with SL using sentence-level gestures has been cited as being better for higher-level learners, suggesting that when a text alone can provide a clear MRC, gestures at the morphological level may not be helpful. Combined (more and higher quality gestures), these results suggest that for L2 spatial term learning, the more consistent speech-gesture input in the CG condition may more efficiently support learning, resulting in an increased ability to generalize to new situations. Hebbian mechanisms for synaptic modification explain why consolidation of learning is an important concept. Insufficient consolidation could explain why learning from second to third measurement (post to retest) in the CG condition did not increase. A follow-up experiment could space teaching over several weeks, as opposed to just one. In addition to spaced teaching, an experiment which addresses the interaction effects between gesture type and access to the written text would be of interest[4]. Much research has shown that gesture, language, and thought are closely linked. The present study exploits this relationship by investigating stable gesture meaning pairs as a teaching tool for young learners.

CONCLUSION

This naturalistic study with a diverse group of learners examined the affects of teacher gestures on long-term spatial term learning. It is widely known that gestures can embody speech and facilitate L2 learning, but gesture research from the classroom on spatial term learning is rare. Although both teaching conditions led to an increase in spatial term ability, in this study children who received gestures at the level of morphology were sooner able to retain and generalize learning than children who received gestures at the sentence level with access to the written text. Children in the CG condition learned their text through interpreting their teachers' gestures, so learners who

---

[4]A follow-up experiment could have the following four groups: (1) + gestures for every morpheme – access to the written text; (2) + gestures for every morpheme + access to the written text; (3) + gestures at the sentence level – access to the written text; and finally (4) + gestures at the sentence level + access to the written text. Because of statistical power such an experiment would require more resources (in terms of participant numbers and teacher time etc.) but could shed light on the interaction between gesture type and access to the written text inherent in the present experiment. Given that gesture and text are readily available in classrooms, an experiment focusing on these different forms could be a worthwhile investment.

struggle with reading and writing in an additional language may especially benefit from the opportunity to learn texts through multimodal means. Further more focused research is needed to isolate whether other factors, such the learning modalities themselves (reading vs. not reading or gesture type) are relevant. Because both teaching methods described here may be applicable to the teaching of other languages, the results of this study should be of interest to researchers seeking effective methods for teaching spatial terms in languages other than English.

There are, of course, many limitations to this study. The careful reader may have noticed that the *dark* gesture in **Figure 1** does not have a direct semantic relationship to any spatial term. This can be explained by the task given to the teachers while creating the gestures. Teachers were asked to embody the most significant sentences of the play in movement and not given any restrictions on what should be important. To shed more light on this aspect, further studies should be conducted in order to more directly ask teachers to act out the locative words, instead of leaving this up to chance. Another justifiable point of criticism could be that the children were not more explicitly instructed to gesture[2]. At the same time, there is also evidence that learners benefit from observing gestures and that "more gestures" are not necessarily better for learning (Huang et al., 2019). Because languages differ in how spatial thought is expressed, it is also plausible that taking the learner's L1 into consideration when designing gestures could have resulted in more specific and more effective learning gestures especially for the refugee children who did not share their teacher's L1. While using complete teaching methods can establish how instructional elements work in combination, results from this comparison cannot readily be extended to other combinations (such as gestures on the basis of morphology plus access to the text). For this reason in future research on the long-term effect of gestures on learning it would be interesting to consider including another condition for which instruction is entirely text-based and doesn't include any gestures in order to further investigate how groups differ over time.

Gestures are an integral part of classroom situations and offer teachers a powerful tool for helping learners to acquire, retain and apply knowledge to new situations. In addition to exploring instructional gestures in experimental settings, research from the classroom is necessary since conditions in the classroom have a complexity that cannot be reduced while doing justice to how education is really practiced.

DATA AVAILABILITY STATEMENT

Because participants in this study are children, while reviewers were granted access to the raw data, for reasons of privacy, permission for a wider audience was not. Further inquiries can be made directed to the corresponding author.

AUTHOR'S NOTE

A reviewer pointed out the uneven distribution of refugee, German and Polish participants. Children from these groups were randomly assigned to the experimental learning conditions specifically for this reason; however, since the unit of inclusion was entire classes, it was beyond the author's control to manipulate the exact number of participants in the study.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2020.00109/full#supplementary-material

The complete spatial term testing materials and procedures can be found at https://osf.io/y8z6a/.

REFERENCES

Ahlberg, D. K., Bischoff, H., Strozyk, J. V., Bryant, D., and Kaup, B. (2018). How do German bilingual schoolchildren process German prepositions? – A study on language-motor interactions. *PLoS ONE* 13:e0193349. doi: 10.1371/journal.pone.0193349

Alibali, M. W., Young, A. G., Crooks, N. M., Yeo, A., Wolfgram, M. S., Ledesma, I. M., et al. (2013). Students learn more when their teacher has learned to gesture effectively. *Gesture* 13, 210–233. doi: 10.1075/gest.13.2.05ali

Arndt, P. A., and Sambanis, M. (2017). *Didaktik und Neurowissenschaften: Dialog zwischen Wissenschaft und Praxis*. Tübingen: Narr Francke Attempto.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 67. doi: 10.18637/jss.v067.i01

Böttger, H., and Sambanis, M. (2017). *Sprachen Lernen in der Pubertät*. Tübingen: Narr Francke Attempto.

Bowerman, M. (1996). "The origins of children's spatial semantic categories: cognitive versus linguistic determinants," in *Rethinking Linguistic Relativity*, eds J. J. Gumperz, and S. C. Levinson (Cambridge: Cambridge University Press), 145–176.

Brala, M. M. (2002). Prepositions in UK monolingual learners' dictionaries: expanding on Lindstromberg's problems and solutions. *Appl. Linguist.* 23, 134–140. doi: 10.1093/applin/23.1.134

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055

Bruner, J. (1983). *Child's Talk: Learning to Use Language.* New York, NY: Norton.

Cichy, R. M., and Teng, S. (2017). Resolving the neural dynamics of visual and auditory scene processing in the human brain: a methodological approach. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160108. doi: 10.1098/rstb.2016.0108

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence.

Cook, S. W. (2018). "Enhancing learning with hand gestures: potential mechanisms," in *Psychology of Learning and Motivation,* Vol. 69, ed K. D. Federmeier (Cambridge, MA: Elsevier), 107–133. doi: 10.1016/bs.plm.2018.10.001

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2010). Gesturing makes memories that last. *J. Mem. Lang.* 63, 465–475. doi: 10.1016/j.jml.2010.07.002

Crystal, D. (2007). *How Language Works: How Babies Babble, Words Change Meaning, and Languages Live or Die 1st Trade pbk.* New York, NY: Avery.

Dargue, N., Sweller, N., and Jones, M. P. (2019). When our hands help us understand: a meta-analysis into the effects of gesture on comprehension. *Psychol. Bull.* 145, 765–784. doi: 10.1037/bul0000202

de Koning, B. B., Bos, L. T., Wassenburg, S. I., and van der Schoot, M. (2017). Effects of a reading strategy training aimed at improving mental simulation in primary school children. *Educ. Psychol. Rev.* 29, 869–889. doi: 10.1007/s10648-016-9380-4

Eskildsen, S. W., and Wagner, J. (2015). Embodied L2 construction learning. *Lang. Learn.* 65, 268–297. doi: 10.1111/lang.12106

Foster, P. (2001). "Rules and routines: a consideration of their role in the task-based language production of native and non-native speakers," in *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing,* eds M. Bygate, P. Skehan, and M. Swain (London: Longman), 75–93.

Glenberg, A. M. (2011). How reading comprehension is embodied and why that matters. *Int. Electron. J. Elementary Educ.* 4, 5–8.

Göksun, T., Lehet, M., Malykhina, K., and Chatterjee, A. (2013). Naming and gesturing spatial relations: Evidence from focal brain-injured individuals. *Neuropsychologia.* 51, 1518–27. doi: 10.1016/j.neuropsychologia.2013.05.006

Goldin-Meadow, S. (2000). Beyond words: the importance of gesture to researchers and learners. *Child Dev.* 71, 231–239. doi: 10.1111/1467-8624.00138

Goldin-Meadow, S., Kim, S., and Singer, M. (1999). What the teacher's hands tell the student's mind about math. *J. Educ. Psychol.* 91, 720–730. doi: 10.1037/0022-0663.91.4.720

Gullberg, M. (2013). "Gestures and second language acquisition," in *Body - Language - Communication: An International Handbook on Multimodality in Human Interaction,* eds C. Müller, A. J. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 1868–1875.

Gunter, T. C., Weinbrenner, J. E., and Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Front. Psychol.* 6:80. doi: 10.3389/fpsyg.2015.00080

Hattie, J., and Yates, G. C. (2013). *Visible Learning and the Science of How We Learn.* London: Routledge. doi: 10.4324/9781315885025

Hilger, N., and Beauducel, A. (2017). "Parallel-forms reliability," in *Encyclopedia of Personality and Individual Differences*, eds V. Zeigler-Hill, T. K. Shackelford (Cham: Springer), 1–3.

Hille, K., Vogt, K., Fritz, M., and Sambanis, M. (2010). Szenisches Lernen im Fremdsprachenunterricht: die evaluation eines schulversuchs. *Diskurs Kindheits- Und Jugendforschung* 5, 337–350.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., and Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Front. Psychol.* 3:74. doi: 10.3389/fpsyg.2012.00074

Huang, X., Kim, N., and Christianson, K. (2019). Gesture and vocabulary learning in a second language. *Lang. Learn* 69, 177–197. doi: 10.1111/lang.12326

Hupp, J. M., and Gingras, M. C. (2016). The role of gesture meaningfulness in word learning. *Gesture* 15, 340–356. doi: 10.1075/gest.15.3.04hup

Janzen Ulbricht, N. (2018). An experiment on gesture and fluency in two German schools. *ELT J.* 72, 309–318. doi: 10.1093/elt/ccx059

Johansson Falck, M. (2018). "Embodied experience and the teaching and learning of L2 prepositions: a case study of abstract in and on," in *What is Applied Cognitive Linguistics? Answers From Current SLA Research*, eds A. Tyler, L. Huang, and H. Jan (Berlin: De Gruyter Mouton), 287–304.

Kang, S., Tversky, B., and Black, J. B. (2015). Coordinating gesture, word, and diagram: explanations for experts and novices. *Spatial Cogn. Comput.* 15, 1–26. doi: 10.1080/13875868.2014.958837

Kelley, P., Evans, M. D., and Kelley, J. (2018). Making memories: why time matters. *Front. Hum. Neurosci.* 12:400. doi: 10.3389/fnhum.2018.00400

Kelly, S. D. (2017). "Chapter 11. Exploring the boundaries of gesture-speech integration during language comprehension," in *Why Gesture?: How the Hands Function in Speaking, Thinking and Communicating*, Vol. 7, eds R. Breckinridge Church, M. W. Alibali, and S. D. Kelly (Amsterdam: John Benjamins Publishing Company), 243–265.

Kelly, S. D., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Lang. Cogn. Process* 24, 313–334. doi: 10.1080/01690960802365567

Kiefer, M., Sim, E. J., Liebich, S., Hauk, O., and Tanaka, J. (2007). Experience-dependent plasticity of conceptual representations in human sensory-motor areas. *J. Cogn. Neurosci.* 19, 525–542. doi: 10.1162/jocn.2007.19.3.525

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

Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *J. Memory. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3

Knoeferle, P., Crocker, M. W., and Pulvermüller, F. (2010). Sentence processing and embodiment. *Brain Lang.* 112, 137–142. doi: 10.1016/j.bandl.2009.11.004

Krifka, M., Błaszczak, J., Leßmöllmann, A., Meinunger, A., Stiebels, B., Tracy, R., (eds.)., et al. (2014). *Das Mehrsprachige Klassenzimmer.* Berlin: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-34315-5

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and the Challenge to Western Thought.* New York, NY: Basic Books.

Lütke, B. (2011). *Deutsch als Zweitsprache in der Grundschule: ein Untersuchung zum Erlernen lokaler Präpositionen.* Berlin: De Gruyter Mouton. doi: 10.1515/9783110267808

Macedonia, M. (2020). "Voice movement icons in der Praxis des Fremdsprachenunterrichts (am Beispiel des Italienisch)," in *Bewegtes Lernen—Ein Handbuch für Forschung und Praxis*, Vol 9, eds C. Andrä and M. Macedonia (Berlin: Lehmanns Media Berlin), 326–334.

Macedonia, M., and Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind Brain Education* 8, 74–88. doi: 10.1111/mbe.12047

Macedonia, M., Repetto, C., Ischebeck, A., and Mueller, K. (2019). Depth of encoding through observed gestures in foreign language word learning. *Front. Psychol.* 10:33. doi: 10.3389/fpsyg.2019.00033

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

McGregor, K. K., Rohlfing, K. J., Bean, A., and Marschner, E. (2008). Gesture as a support for word learning: the case of under. *J. Child Lang.* 36, 807–828. doi: 10.1017/S0305000908009173

McNeil, N. M., Alibali, M. W., and Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: now they need it, now they don't. *J. Nonverbal Behav.* 24, 133–150. doi: 10.1023/A:1006657929803

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought.* Chicago: University of Chicago Press.

Murphy, K., and Davidshofer, C. (2005). *Psychological Testing: Principles and Applications.* (6th edition). Upper Saddle River: Pearson.

Nakatsukasa, K. (2016). Efficacy of recasts and gestures on the acquisition of locative prepositions. *Stud. Second Lang. Acquis.* 38, 771–799. doi: 10.1017/S0272263115000467

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., and Goldin-Meadow, S. (2014). From action to abstraction: using the hands to learn math. *Psychol. Sci.* 25, 903–910. doi: 10.1177/0956797613518351

Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Front. Neuroinformatics* 2:10. doi: 10.3389/neuro.11.010.2008

Perniss, P., and Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130300. doi: 10.1098/rstb.2013.0300

Poggi, I. (2013). "Semantics and pragmatics of symbolic gestures," in *Body - Language - Communication: An International Handbook on Multimodality in*

Human Interaction, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 1481–1496.

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

Rohlfing, K. J., Wrede, B., Vollmer, A. L., and Oudeyer, P. Y. (2016). An alternative to mapping a word onto a concept in language acquisition: pragmatic frames. *Front. Psychol.* 7:470. doi: 10.3389/fpsyg.2016.00470

Roth, W. M. (2001). Gestures: their role in teaching and learning. *Rev. Educ. Res.* 71, 365–392. doi: 10.3102/00346543071003365

Rumme, P., Saito, H., Ito, H., Oi, M., and Lepe, A. (2008). *Gestures as Effective Teaching Tools: Are Students Getting the Point? - A Study in Pointing Gesture in the English as a Second Language Classroom.* Nagoya: Graduate School of Information Science, Nagoya University.

Sambanis, M., and Walter, M. (2019). *In Motion - Theaterimpulse zum Sprachenlernen Von neuesten Befunden der Neurowissenschaft zu konkreten Unterrichtsimpulsen.* Berlin: Cornelsen.

Skipper, J. I. (2014). Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130297. doi: 10.1098/rstb.2013.0297

Stokoe, W. C. (2000). "Gesture to sign (language)," in *Language and Gesture*, Vol. 2, ed D. McNeill (Cambridge: Cambridge University Press), 409.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., Herrmann, E., Gilby, I. C., et al. (2012). Two key stepsin the evolution of human cooperation: the interdependence hypothesis. *Curr. Anthropol.* 53, 673–692. doi: 10.1086/668207

Tyler, A., and Evans, V. (2003). *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning, and Cognition.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486517

Van den Broeck, J., Cunningham, S. A., Eeckels, R., and Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med.* 2:e267. doi: 10.1371/journal.pmed.0020267

Wu, Y. C., and Coulson, S. (2007). How iconic gestures enhance communication: an ERP study. *Brain Lang.* 101, 234–245. doi: 10.1016/j.bandl.2006.12.003

Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162