Educational Psychology Review (2020) 32:815–850
https://doi.org/10.1007/s10648-020-09527-z

INTERVENTION STUDY

Learning Foreign Language Vocabulary with Gestures and Pictures Enhances Vocabulary Memory for Several Months Post-Learning in Eight-Year-Old School Children

Christian Andrä[1,2] · Brian Mathias[3,4] · Anika Schwager[5] · Manuela Macedonia[4,6] · Katharina von Kriegstein[3,4]

Published online: 18 April 2020

Abstract
The integration of gestures and pictures into pedagogy has demonstrated potential for improving adults' learning of foreign language (L2) vocabulary. However, the relative benefits of gestures and pictures on children's L2 vocabulary learning have not been formally evaluated. In three experiments, we investigated the effects of gesture-based and picture-based learning on 8-year-old primary school children's acquisition of novel L2 vocabulary. In each experiment, German children were trained over 5 consecutive days on auditorily presented, concrete and abstract, English vocabulary. In Experiments 1 and 2, gesture enrichment (auditorily presented L2 words accompanied with self-performed gestures) was compared with a non-enriched baseline condition. In Experiment 3, gesture enrichment was compared with picture enrichment (auditorily presented words accompanied with pictures). Children performed vocabulary recall and translation tests at 3 days, 2 months, and 6 months post-learning. Both gesture and picture enrichment enhanced children's test performance compared with non-enriched learning. Benefits of gesture and picture enrichment persisted up to 6 months after training and occurred for both concrete and abstract words. Gesture-enriched learning was hypothesized to boost learning outcomes more than picture-enriched learning on the basis of previous findings in adults. Unexpectedly, however, we observed similar benefits of gesture and picture enrichment on children's L2 learning. These findings suggest that both gestures and pictures enhance children's L2 learning and that performance benefits are robust over long timescales.

Keywords Multisensory learning · Sensorimotor learning · Gesture · Enrichment · Vocabulary learning · Foreign language education

Christian Andrä and Brian Mathias are Joint first authors

✉ Brian Mathias
brian.mathias@tu–dresden.de

Extended author information available on the last page of the article

Introduction

Multisensory and Sensorimotor Enrichment

Learning in natural environments is multisensory: Information arising from sensory modalities is integrated when we acquire knowledge or skills. Learning to recognize the voice of a new acquaintance, for example, takes into account both the sight of the individual and their speech characteristics (von Kriegstein et al. 2008; Sheffert and Olson 2004). Interactions between sensory and motor modalities may also be essential for the learning of complex skills such as reading, writing, and arithmetic and are therefore highly relevant for many issues associated with education (Kiefer and Trumpp 2012).

The presence of complementary information across multiple sensory modalities during learning has been referred to as *multisensory enrichment* (Mayer et al. 2015; Repetto et al. 2017). Current pedagogical and neurocognitive theories propose that multisensory input, compared with unisensory input, is beneficial for learning outcomes (Mahmoudi et al. 2012; Sadoski and Paivio 2013; Shams and Seitz 2008; von Kriegstein and Giraud 2006).

In the classroom, multisensory enrichment can take many forms: flash cards (Wissman et al. 2012); videos (Tan and Pearce 2011; for review see Snelson 2011); video games (Annetta et al. 2009; Hsu 2011); songs and poetry (Foster and Freeman 2008; Millington 2011); and interactive activities that make use of mobile phones, computers, and tablets (Ehret and Hollett 2014; for a review see Herodotou 2018; Volk et al. 2017). These enrichment strategies have all been studied as means of enhancing learning outcomes. While some of these pedagogical tools are already widely in use in educational settings, others have scarcely been adopted.

There is mounting evidence that *sensorimotor enrichment* also increases learning efficiency and memory performance. We use the term sensorimotor enrichment to indicate the presence of movements such as gestures during learning that are semantically congruent with information presented in another sensory modality (for a review see Macedonia 2014). Note that sensorimotor enrichment involves information presented across at least one sensory modality, because movements induce somatosensory feedback. Sensorimotor enrichment can also involve the presence of further sensory modalities, e.g., due to seeing the self-generated movements. Like multisensory enrichment, sensorimotor enrichment benefits learning (MacLeod et al. 2010). The pairing of auditory stimuli such as novel spoken words with semantically related gestures, for example, enhances subsequent auditory stimulus recognition (Mayer et al. 2015) and retrieval (Macedonia et al. 2011). In most studies involving the performance of gestures, an individual who models the gestures for the participants is integrated into the participants' learning experience. Also, the performance of pantomimes during the learning of pseudowords and associated visual objects can enhance subsequent visual object recognition compared with the performance of simple pointing movements during learning (v. Soden-Fraunhofen et al. 2008). Memory enhancements following sensorimotor enrichment have variously been termed enactment effects (Engelkamp and Zimmer 1985), production effects (Dodson and Schacter 2001), and subject-performed task effects (Cohen 1981).

Sensorimotor enrichment is of interest to educators as principles of active learning have moved from the periphery of education to its center over the past decades (Lewis and Williams 1994; Michael 2006). During active learning, students engage in activity that encourages them to reflect upon the learning material (Collins and O'Brien 2003). Active learning techniques contrast with other pedagogical approaches that focus on only reading, watching, or listening

to learning material and may facilitate children's learning because they complement the use of more cognitive learning strategies (Cook et al. 2012; Goldin-Meadow 2003; Gullberg et al. 2008). Student engagement and active participation are considered to be among the most critical factors driving their persistence to learn (Braxton et al. 2008).

Cognitive and Neuroscientific Theories of Multisensory and Sensorimotor Enrichment

Benefits of multisensory and sensorimotor enrichment for learning are accounted for by several cognitive and neuroscientific theories. Theories of embodiment argue that concepts are mentally represented in terms of their perceptual features, motor features, and other aspects of one's personal experience (for a review see Barsalou 2008). According to an embodied perspective, multisensory and sensorimotor enrichment may enhance memory by grounding the remembered material in multisensory and sensorimotor experiences. For example, training studies have demonstrated that writing letters by hand enhances subsequent recognition of those letters compared with typing in preschool children (Kiefer et al. 2015). Such studies hint at why multisensory and sensorimotor experience may aid children's education: When meaningful interactions with to-be-learned material are lacking and are replaced by verbal descriptions, children may rely simply on verbal associations for learning, which may be less durable than sensorimotor associations (Kiefer and Trumpp 2012; see also DeLoache et al. 2010).

Nested within an embodiment framework are theories of dual coding (Engelkamp and Zimmer 1984; Hommel et al. 2001; Paivio 1991; Paivio and Csapo 1969) and simulation or imagery accounts (Jeannerod 1995; Kosslyn et al. 2006; Saltz and Dixon 1982). Dual coding theory proposes that stimuli presented in various sensory and sensorimotor modalities are coded either verbally or nonverbally. For example, vocabulary words that are heard and then pronounced are coded verbally, whereas related gestures that are seen and then performed are coded nonverbally. Benefits of multisensory- and sensorimotor-enriched learning can be attributed to the encoding of learned material both verbally and nonverbally in one or more sensory modalities, with the nonverbal code contributing more to memory than the verbal code (Sadoski and Paivio 2013).

How beneficial effects of enrichment are instantiated in the human brain are to date unknown. One overarching mechanistic account of brain function is the Bayesian brain hypothesis. It assumes that the brain represents information probabilistically and uses an internal generative model and predictive coding to most effectively process sensory input (Friston 2005; Friston and Kiebel 2009; Kiebel et al. 2008; Knill and Pouget 2004). In this view, simply listening to a stimulus that has been encoded both in terms of its auditory and visual features, for example, may trigger an internal dynamic generative model that reconstructs its stored visual features (implemented in visual cortices) and thereby helps to recognize the perceptual input (for a review see von Kriegstein 2012; Mayer et al. 2015; Yildirim and Jacobs 2012). These internal generative models of the enriched learning material could explain enhancing learning outcomes (von Kriegstein and Giraud 2006; Yildirim and Jacobs 2012).

Effects of Enrichment on Native Language (L1) Vocabulary Learning

A quintessential example of the benefits of *multisensory* and *sensorimotor* enrichment can be found in the acquisition of novel vocabulary. When acquiring their native language, individuals accumulate extensive multisensory and sensorimotor experience with caregivers and the environment (Kuhl 2010). Over time, specific sensory experiences and motor responses

become associated with each other and labeled with a sequence of phonemes, i.e., a word (Lupyan and Thompson-Schill 2012; Macedonia 2015).

Multisensory-enriched learning of native language words aids L1 acquisition (Hadley et al. 2016). Many parents begin reading picture books to their children shortly after birth; the reading of picture books prior to age 2 correlates with pre-literate children's oral L1 skills (DeBaryshe 1993). However, quantifying the benefits of picture book reading has been difficult due to the lexical diversity of picture books, as well as the lack of adequate control learning conditions (Montag et al. 2015; Sénéchal and Cornell 1993). In one study, 135 third- and fourth-grade children showed improved memory for L1 words accompanied by pictures compared with words with no visual enrichment (Acha 2009). The inclusion of sign language curriculum in preschool, in which words are presented visually, kinesthetically, and verbally, may also speed up children's acquisition of L1 vocabulary (Daniels 1997).

Sensorimotor-enriched learning of native language words also contributes to L1 vocabulary knowledge. For example, speaking native language words can aid their acquisition: Five-year-olds who are taught novel L1 words by speaking them aloud show greater memory for those words compared with words taught only by listening, a sensorimotor enrichment benefit (Icht and Mama 2015). Several studies have suggested that the performance of gestures and other movement-based interventions during learning facilitates the comprehension of novel L1 sentences (for a review see Sadoski 2018). For example, 6- and 8-year-olds remember action phrases in L1 such as "lift the bottle" better following enactment using gestures compared to verbal repetition (Mecklenbräuker et al. 2011). Gestures produced early in language development provide a way for children to communicate information that they cannot yet express verbally and predict the future acquisition of corresponding L1 vocabulary (Iverson and Goldin-Meadow 2005; Caselli et al. 2012). Cross-cultural studies (Bavin et al. 2008; Eriksson and Berglund 1999) suggest a common cultural and biological basis for the role of gestures in the development of spoken language.

Effects of Enrichment on Vocabulary Learning in the Foreign Language (L2) Classroom

Only a few studies have examined the effects of multisensory and sensorimotor enrichment on vocabulary learning in applied educational contexts. These studies have mostly investigated the learning of L2 vocabulary, as young children typically spend less than 15% of classroom time on direct L1 vocabulary instruction depending on the curricula (McGill-Franzen et al. 2006; Scott et al. 2003).

Multisensory enrichment for the learning of L2 vocabulary is commonly used by teachers of foreign languages. However, empirical investigations of the use of visual materials for L2 learning in young children are sparse. Silverman and Hines (2009) found that the viewing of short video clips that supplemented teachers' regular instruction increased kindergartners' through second graders' knowledge of L2 vocabulary words. This multimedia enrichment also narrowed the gap in vocabulary knowledge between L2 learners and native speakers. The use of images, animations, and videos in the teaching of English as an L2 has been suggested as a means of improving L2 learning outcomes in university-age students (for a review see Gilakjani 2012; Konomi 2014). Video, in particular, provides a rich semantic and pragmatic context to which lexical phrases can be linked for the teaching of foreign languages (Tschirner, 2001).

In the case of sensorimotor enrichment, one recent study conducted with 111 preschool children (5-year-olds) in an educational context found that the performance of physical exercise and iconic gestures while listening to foreign language vocabulary increased recall compared with exercising without gestures (Mavilidi et al. 2015). In another study, a teacher of English as a foreign language enacted gestures, repeated utterances, and checked children's comprehension while telling a simple story. Together, these modifications increased Spanish primary school children's (10-year-olds') comprehension of the story compared with non-modified storytelling (Cabrera and Martínez 2001). Finally, Macedonia et al. (2014); see also de Wit et al. 2018) demonstrated that 11-year-old children's L2 vocabulary learning outcomes were boosted more by performing related gestures themselves during learning than by viewing a pedagogical agent who performs the gestures.

Comparing the Effectiveness of Multisensory and Sensorimotor Enrichment on L2 Vocabulary Learning

The quest for optimal L2 teaching strategies generates a key question: Is the use of gesture-enriched learning more beneficial than the use of more commonly practiced picture-enriched learning? Studies on young adults have suggested that gesture-enriched learning can enhance cued memory recall of L2 vocabulary even more than picture-enriched learning (Mayer et al. 2015; Repetto et al. 2017): In one recent study, adults learned L2 vocabulary by reading an L2 word aloud while viewing its written L1 translation and performing a gesture (gesture-enriched learning) or viewing a picture (picture-enriched learning; Repetto et al. 2017). After 35 min of training, participants produced fewer translation errors for gesture-enriched L2 words compared with picture-enriched L2 words in a multiple choice task. In another study, gesture-enriched learning yielded more accurate L2 translation performance compared with picture-enriched learning 6 months following a week-long L2 training period (15 h of training; Mayer et al. 2015).

Whether these findings in adults (Mayer et al. 2015; Repetto et al. 2017) can be translated to children is an open question. Adults and children differ in many ways with regard to learning mechanisms, such as working memory abilities (Luna et al. 2004) and use of visual and motor imagery (Frick et al. 2009; Funk et al. 2005; for a review see Gabbard 2009). However, some studies show similar learning mechanisms in adults and children (Raviv and Arnon 2018; Saffran et al. 1999). It may also be the case for children's L2 learning in the classroom that gesture-enriched learning is more effective than picture-enriched learning. The potentially greater effectiveness of gesture-enriched compared with picture-enriched learning would have consequences for the development of evidence-based teaching strategies. Additionally, to our knowledge, no previous studies on children's L2 vocabulary learning have directly compared sensorimotor enrichment strategies with a unisensory baseline learning condition. This leaves an open possibility that facilitative effects of sensorimotor enrichment could merely be an artifact of impaired learning in one of the enriched control conditions, as enriched learning creates a larger cognitive load than unisensory learning (Mayer and Moreno 2003).

Two studies have attempted to clarify the question of whether gesture enrichment or picture enrichment may differ in terms of their effects on children's L2 vocabulary acquisition (Porter 2016; Tellier 2008). A study conducted by Tellier (2008) with 4- to 5-year-olds demonstrated that gesture enrichment may yield larger gains in children's L2 vocabulary learning compared with picture enrichment (viewing pictures while hearing novel words). Porter (2016)

additionally showed that combined gesture and picture enrichment may yield larger gains for 5- to 6-year-olds than picture enrichment alone. However, these studies possess some limitations. Besides the use of small sample sizes (10 children per condition in one study in a between-subjects design; Tellier 2008), one study did not control the number of L2 word repetitions that occurred within each enrichment condition or the classes of words that were taught (Porter 2016), so that learning effects could be attributed to participant's amounts of practice rather than enrichment types. Also the studies examined enrichment effects over limited time periods (i.e., up to 2 weeks after instruction had stopped; Porter 2016).

Aims of the Current Study

In adults, benefits of gesture enrichment are known to exceed those of picture enrichment (Mayer et al. 2015). It is, however, unclear whether similar outcomes would be shown by children in naturalistic classroom environments. Evaluating learning outcomes in children is important as it would be a first step toward answering the question of whether educators should integrate both forms of enrichment in classrooms for school children, or whether one type of enrichment should be preferred due to its greater effectiveness. The approach of the current study was to follow a design that was used previously in laboratory-based tests in adults (Mayer et al. 2015) and translate it into a primary school classroom setting.

Our primary aim was to compare the effects of gesture and picture enrichment on L2 vocabulary learning. In order to evaluate enrichment effects, we first compared benefits of gesture enrichment with a non-enriched baseline learning condition. Comparison of gesture-enriched learning and non-enriched learning is necessary for quantifying the extent to which gesture enrichment enhances non-enriched learning, as well as addressing the alternative explanation for differences between learning conditions that enrichment may under some circumstances impede learning (Mayer and Moreno 2003).

Gesture and picture enrichment have yielded similar effects for both concrete (e.g., *tent*) and abstract (e.g., *thought*) L2 vocabulary in adults (Macedonia and Knösche 2011; Mayer et al. 2017). However, abstract vocabulary learning typically lags behind concrete vocabulary learning during development (McFalls et al. 1996). Our second aim was to investigate whether gesture or picture enrichment has the potential to boost children's learning of both concrete and abstract word types.

An important aspect of developing L2 proficiency is the retention of new vocabulary over long time periods, even when words are not retrieved from memory on a regular basis. Therefore, our third aim was to investigate long-term influences of gesture and picture enrichment on school children's L2 vocabulary retention by comparing the effects of the two types of enrichment up to 6 months post-learning.

Summary of the Experimental Approach and Hypotheses

We conducted three experiments that all included school children enrolled in grade three of primary school. Three main hypotheses were tested. First, on the basis of studies performed in young adults (Mayer et al. 2015), we expected that pairing self-performed gestures with L2 vocabulary would enhance learning outcomes compared with non-enriched learning (Experiments 1 and 2). In adults, gesture enrichment has been shown to improve performance on both cued translation (Mayer et al. 2015) and free recall (Macedonia and Knösche 2011; Zimmer et al. 2000) tests. Cued translation refers to a learner's translation of a written or spoken L1 or

L2 word that serves as a memory cue, and free recall refers to a learners remembering an L1 word and its L2 translation in the absence of any written or spoken cue. Gesture-enriched learning was expected to enhance children's cued L1 and L2 translation, as well as free recall of L1-L2 translations, compared with non-enriched learning.

Second, though the learning of concrete words was expected to exceed the learning of abstract words, enrichment effects of a similar magnitude were expected for both concrete and abstract L2 words based on studies in adults (Macedonia and Knösche 2011; Mayer et al. 2017) (Experiments 1 and 2).

Third, as adults show greater memory for L2 words learned using gestures compared with pictures over the long term (6 months following learning; Mayer et al. 2015) on cued translation tasks, gesture enrichment was expected to benefit post-learning cued L1 and L2 translation more than picture enrichment at later time points following learning (Experiment 3). We also explored whether gesture-enriched learning benefitted children's free recall performance more than picture-enriched learning; this particular effect was not observed in adults (Mayer et al. 2015).

Experiment 1

Experiment 1 served to test the hypotheses that pairing self-performed gestures with auditorily presented L2 vocabulary would enhance learning outcomes compared with non-enriched learning and that the gesture-based learning benefits would occur for both concrete and abstract nouns.

Methods

Participants Participants were school children enrolled in grade three at a primary school in Leipzig, Germany. The school is classified within the German education system as a *bewegte Schule* ("movement school"; http://www.bewegte-schule-und-kita.de/konzept/bewegteSchule/english/html/konzept.html). This type of school emphasizes the role of movement in student learning and development (Breithecker and Dordel 2003). Seventy-one children participated in Experiment 1. The investigators briefed all children and teachers on the study procedures in an introductory session that took place prior to the experiment. Children who were absent from at least one training or test session were excluded from the analyses. Therefore, 54 children were included in the analyses. Written informed consent was obtained from the legal guardians of all individual school children who participated. None of the children were reported by the school principal or teachers to possess learning disabilities. All of the children possessed normal or corrected-to-normal vision. Demographic information can be found in Table 1. Experiment 1 was reviewed and approved by the Education Department of the state of Saxony, Germany.

Stimulus Materials Forty English words were used in Experiment 1 (Table 2). Half of the words were concrete nouns, and the other half were abstract nouns. The concrete and abstract nouns significantly differed in terms of concreteness on the basis of ratings from a corpus of 350,000 German lemmas, $t\ (38) = 13.06$, $p < 0.001$ (concrete nouns, $M = 6.58$, $SD = 0.98$, range, 4.07–7.99; abstract nouns, $M = 2.95$, $SD = 0.76$, range, 1.80–4.61) (Köper and

Table 1 Participant demographic information for Experiments 1, 2, and 3

	Exp. 1	Exp. 2	Exp. 3
Sample size	54	43	51
Age (years)	M = 8.7 SD = 0.5 Range = 8.5–10.2	M = 8.3 SD = 0.4 Range = 8.0–10.0	M = 8.2 SD = 0.5 Range = 8.0–10.0
Females	23	22	32
Children who spoke English as a second language	0	0	0
Children who spoke at least one other language besides German	4	3	4
Children reportedly currently or previously learning to play a musical instrument	19	13	26
Children reportedly currently participating regularly in extracurricular sports	41	33	38

Schulte Im Walde 2016). Words were selected from the "Vimmi" language corpus (Macedonia et al. 2010, 2011). Word selection was based on recommendations of the children's school-teachers and was constrained by two factors. First, children had not yet encountered the words in lessons, and the words were not anticipated to be included in teaching curriculum for the 6-month duration of the investigation. Second, the words were considered by the teacher to be relevant for future use by the children. Word frequencies in written German (http://wortschatz.uni-leipzig.de/en), as well as the numbers of syllables and letters contained in the English translations, were counterbalanced between learning conditions. Word frequencies and lengths were also counterbalanced between sets of concrete and abstract words included in the two learning conditions.

Experiment 1 made use of two stimulus types: audio recordings of English words and their German translations and videos of an actress performing gestures that were semantically related to word meanings (iconic gestures). Videos and pictures were adopted from the "Vimmi" corpus (Macedonia et al. 2010, 2011; Mayer et al. 2015).

We recorded German stimuli with a bilingual Italian-German speaker (female, age 44) and English translations with a native speaker of British English (female, age 21). Recordings were made using a RØDE NT55 microphone (RØDE Microphones, Silverwater, Australia) in a sound-dampened room. The 40 German word recordings used in Experiment 1 were M = 0.86 s (SD = 0.18 s) in length, and English word recordings were M = 0.84 s (SD = 0.15 s) in length. The 24 German word recordings used in Experiments 2 and 3 were M = 0.81 s (SD = 0.16 s) in length, and English word recordings were M = 0.83 s (SD = 0.15 s) in length.

Videos were recorded using a Canon Legria HF S10 camcorder (Canon Inc., Tokyo, Japan). Each video was 4 s long and shot in color. The actress shown in the videos began and ended each video by standing motionless with her arms by her sides. During the videos, she used head movements, movements of one or both arms or legs, fingers, or combinations of these body parts to convey the meaning of the foreign language word through the movement. For example, the word *tent* was conveyed by moving the arms and fingers together to form an upside-down "V" shape, and the word *thought* was conveyed by touching the head with one hand and subsequently pointing upward with the same hand above the head (Fig. 1). The actress always maintained a neutral facial expression. Gestures selected for abstract nouns were previously agreed upon by three independent raters (Macedonia et al. 2011; Mayer et al. 2015).

Table 2 German and English words used in Experiments 1, 2, and 3

Experiment 1			
Concrete nouns		Abstract nouns	
German	English	German	English
Ampel	traffic light	Absage	Cancelation
Anhänger	Trailer	Anforderung	Requirement
Bildschirm	Screen	Aufwand	Effort
Briefkasten	Mailbox	Befehl	Order
Decke	Ceiling	Besitz	Property
Denkmal	Monument	Bitte	Request
Faden	Thread	Empfehlung	Recommendation
Fernbedienung	Remote	Gedanke	Thought
Flasche	Bottle	Geduld	Patience
Handtasche	Bag	Langeweile	Boredom
Kasse	Till	Mut	Courage
Koffer	Suitcase	Tatsache	Fact
Reifen	Tire	Teilnahme	Participation
Rucksack	Backpack	Übung	Exercise
Sammlung	Collection	Unschuld	Innocence
Schlüssel	Key	Veränderung	Change
Schublade	Drawer	Verständnis	Sympathy
Spiegel	Mirror	Vorgehen	Procedure
Tageszeitung	Newspaper	Zulassung	Permit
Zelt	Tent	Zweck	Purpose
Experiments 2 and 3			
Concrete nouns		Abstract nouns	
German	English	German	English
Ampel	traffic light	Absage	Cancelation
Anhänger	Trailer	Anforderung	Requirement
Decke	Ceiling	Aufwand	Effort
Faden	Thread	Besitz	Property
Fernbedienung	Remote	Bitte	Request
Kasse	Till	Gedanke	Thought
Koffer	Suitcase	Geduld	Patience
Reifen	Tire	Langeweile	Boredom
Sammlung	Collection	Teilnahme	Participation
Schublade	Drawer	Unschuld	Innocence
Spiegel	Mirror	Zulassung	Permit
Zelt	Tent	Zweck	Purpose

Design and Procedure Experiment 1 had a 2 × 2 × 3 factorial within-subjects design with the factors learning condition (gesture enrichment, no enrichment), word type (concrete, abstract), and testing time point (3 days, 2 months, and 6 months post-learning).

Learning Phase Children completed 5 consecutive days of L2 vocabulary training (Fig. 2). Training occurred over four 5-min blocks per day. Rest activities occurred for 5 min between each of the blocks. Thus, each daily training session had a total duration of approximately 35 min.

L2 words and their L1 translations were presented in non-enriched and gesture-enriched trials (Fig. 3). In both trial types, children first heard an English word, which was followed by its auditorily presented German translation and then by a repetition of the English word. The children's teacher then cued the children to recite the English word aloud with the words "all together." The teacher stood at the front of the classroom during the entire training period. In the gesture enrichment condition, recorded English words were

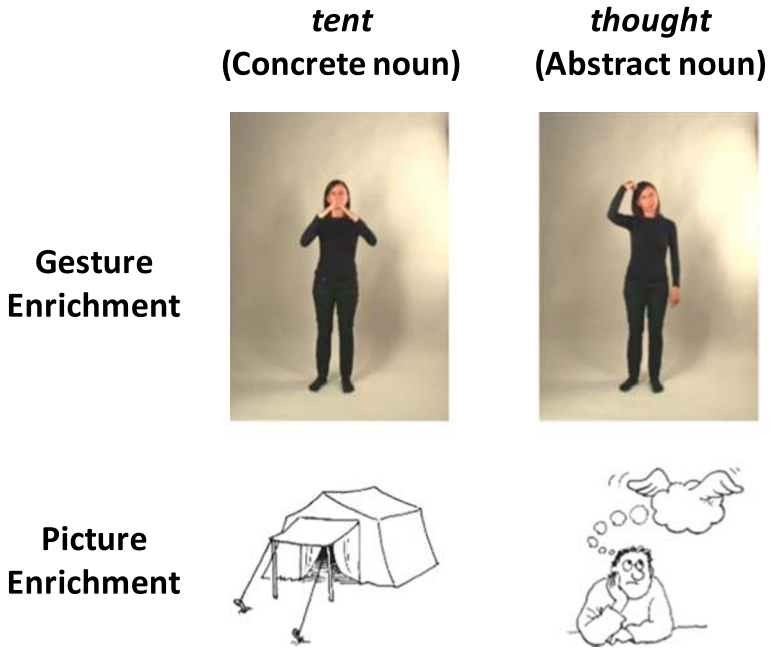

Fig. 1 Gesture and picture stimuli. Top: Screen captures from the videos of the actress performing gestures for one of the concrete nouns (*tent*) and one of the abstract nouns (*thought*) used in the gesture enrichment condition. Bottom: Pictures used in the picture enrichment condition corresponding to the same words

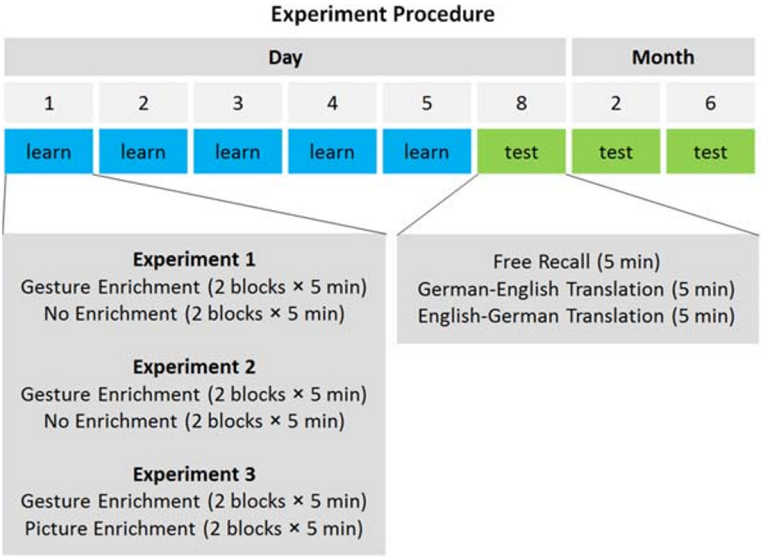

Fig. 2 Experimental procedure and design. The learning phase of each experiment occurred over 5 days ("learn"). Free recall and translation tests ("test") were administered 3 days, 2 months, and 6 months following the end of the learning phase. Primary school children learned foreign language words in gesture, picture, and no enrichment conditions. Experiments 1 and 2 included gesture and no enrichment learning conditions, and Experiment 3 included gesture and picture enrichment conditions

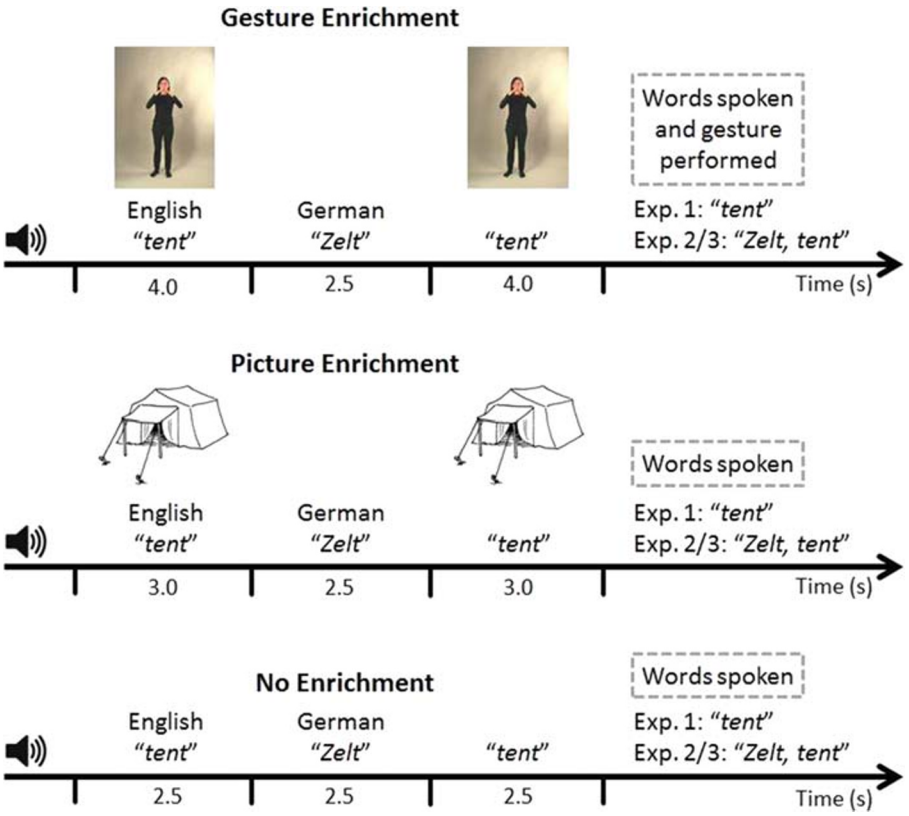

Fig. 3 Learning phase trial design. In each trial, auditorily presented English words were accompanied either by a video of an actress performing a gesture (gesture enrichment), a picture (picture enrichment), or no complimentary stimulus (no enrichment). English words were followed by the auditorily presented German translation and a repetition of the English word accompanied again by the enrichment stimulus. The students then spoke the foreign word (Experiment 1) or both the foreign and native words (Experiments 2 and 3) together with their teacher. In the gesture enrichment condition, the children performed gestures with their teacher while speaking. The children's task was to learn the correct association between the foreign words and their German translations

accompanied by videos of an actress performing an iconic gesture. At the end of the trial, children performed the gesture along with the teacher. The time interval between English and German word onsets in the non-enriched learning condition was 2.5 s. A shorter time interval was used for the presentation of the English words in the non-enriched learning condition, compared with the videos in the gesture condition, in order to avoid a large difference in inter-stimulus intervals between the two conditions. Long time periods during which no sensory information is presented could potentially decrease attention, motivation, or stimulus-driven arousal during the non-enriched trials compared with the gesture-enriched trials, which would favor higher performance outcomes for the gesture learning condition compared with the picture learning condition. The time interval between the onset of the German word's presentation and the onset of the English word's repetition was 2.5 s in both conditions. Children's locations in the classroom were randomly assigned for each training block. Children sat at desks during no enrichment trials and stood next to desks during the gesture trials. One of the investigators monitored the testing equipment and initiated each trial as soon as the children were ready.

Learning phase trials were blocked by learning condition. In Experiment 1, children completed 4 blocks containing 10 trials (5 concrete word trials and 5 abstract word trials) per day. Half of the blocks comprised gesture enrichment, and the other half comprised non-enriched learning. Each German word and its English translation were presented in only a single trial each day. Word orders within each block and orders of enrichment condition blocks were counterbalanced across learning days.

Five-minute rest activities occurred between learning blocks. The activities were intended to promote cognitive recovery and relaxation and consisted of skipping rope, ball-throwing, and partner massage in groups of 2 to 4 individuals. These activities were familiar to the children as they were sometimes integrated into the School in Motion curriculum outside of the context of the study. The order of rest activities was counterbalanced across learning days.

Children completed training sessions in groups of up to 13 to ensure adequate space to perform the gestures and minimization of distraction. Stimuli were counterbalanced across groups of children: Some children learned one set of words in the gesture condition and another set in the non-enriched learning condition, and the remaining children received the reverse assignment of words to the enrichment conditions. Children's positions within the classroom were counterbalanced across 5 days of training.

Test Phase Children completed vocabulary tests at three time points: 3 days, 2 months, and 6 months following the completion of the learning phase. Free recall, German-English, and English-German translation tests were conducted orally at each time point, since none of the children possessed adequate writing skills in English as a foreign language.

Native German-speaking examiners conducted the test sessions individually at the same school where the learning phase took place. The examiners were university students enrolled in teaching certification programs at the University of Leipzig. Examiners were blind with respect to which words had been learned in enrichment condition. Further, they had no knowledge of gestures or pictures paired with individual words in the experiment.

During each test session, one of the school children sat at a desk opposite one of the examiners. In the free recall test, children were asked to verbalize as many German-English or English-German translations, individual German words, or individual English words as they could remember from the training. A time limit of 5 min was imposed; children were not instructed about this time limit, and no child in any experiment exceeded 5 min. Following the free recall test, the children completed the two translation tests. The free recall test was always administered prior to the translation tests to eliminate influences of memory cues present in the translation tests.

During the German-English translation test, the examiner spoke the German words one at a time, and the children were asked each time to give the correct English translation. During the English-German translation test, the examiner spoke the English words one at a time, and the children were asked each time to give the correct German translation. The German-English translation test was always administered prior to the English-German test, as translation from one's native to a foreign language has been shown to be a more difficult task than the translation from a foreign language into one's native language (Kroll and Stewart 1994). Examiners were instructed to speak the English words with a pronunciation based on the recordings used in the experiments. Children were given 5 s to state their answers before moving to the next word. Test word orders in the two translation tests were randomized for each testing time point (3 days, 2 months, and 6 months post-learning).

Examiners recorded test sessions as an audio file for subsequent analysis using a personal recording device such as a mobile phone. The children did not receive any feedback regarding the correctness of their answers. Children were instructed not to discuss the tests with their classmates. Each test session lasted approximately 10–15 min.

Data Analysis Audio files from individual student test sessions were independently scored for accuracy by two raters. The two raters had not conducted any of the test sessions and were also blind with respect to which words had been learned in each enrichment condition. In cases of disagreement, a third independent rater was employed, and the majority decision was adopted.

German-English translation and English-German translation tests were scored in terms of the total number of correct translations recalled in each test (one point for each correct translation). One point was given for each correct translation (German-English or English-German word pair) provided during the free recall test. No points were given for a German word that was missing a corresponding English translation or vice versa.

In order to evaluate effects of enrichment and vocabulary type on learning, three-way analyses of variance (ANOVAs) with the factors enrichment condition (gesture enrichment, no enrichment), word type (concrete and abstract), and time point (3 days post-learning, 2 months post-learning, and 6 months post-learning) were conducted for each learning outcome (L1-L2 translation, L2-L1 translation, and free recall). All post hoc comparisons were evaluated with Tukey's HSD tests. The significance threshold was set to $\alpha = 0.05$, and partial eta-squared and Hedge's g were computed as measures of effect size (Greenland et al. 2016).

Table 3 Mean scores on German-English translation, English-German translation, and free recall tests in Experiments 1 and 2

		3 days post-learning		2 months post-learning		6 months post-learning	
		Non-enriched M (SE)	Gesture M (SE)	Non-enriched M (SE)	Gesture M (SE)	Non-enriched M (SE)	Gesture M (SE)
Exp. 1	German-English translation						
	Concrete	2.70 (0.24)	3.19 (0.33)	2.39 (0.30)	3.11 (0.36)	4.02 (0.36)	2.93 (0.22)
	Abstract	0.69 (0.15)	3.02 (0.38)	0.15 (0.07)	1.35 (0.23)	0.41 (0.11)	0.98 (0.20)
	English-German translation						
	Concrete	4.70 (0.31)	4.61 (0.32)	3.89 (0.30)	3.93 (0.30)	4.06 (0.30)	3.96 (0.32)
	Abstract	1.24 (0.22)	2.81 (0.27)	0.70 (0.16)	1.89 (0.26)	0.54 (0.15)	1.39 (0.20)
	Free recall						
	Concrete	0.76 (0.13)	0.72 (0.15)	0.78 (0.16)	0.93 (0.18)	1.33 (0.19)	1.44 (0.20)
	Abstract	0.15 (0.06)	0.57 (0.11)	0.02 (0.02)	0.41 (0.11)	0.13 (0.05)	0.22 (0.08)
Exp. 2	German-English translation						
	Concrete	2.00 (0.24)	2.30 (0.25)	0.91 (0.17)	1.58 (0.25)	0.88 (0.17)	1.21 (0.20)
	Abstract	1.23 (0.22)	1.86 (0.27)	0.53 (0.11)	0.91 (0.18)	0.21 (0.07)	0.37 (0.11)
	English-German translation						
	Concrete	3.00 (0.28)	3.72 (0.27)	2.23 (0.26)	3.12 (0.26)	2.12 (0.27)	2.77 (0.25)
	Abstract	1.51 (0.21)	2.63 (0.23)	1.05 (0.19)	2.05 (0.25)	0.58 (0.13)	1.07 (0.20)
	Free recall						
	Concrete	0.49 (0.11)	0.60 (0.13)	0.26 (0.08)	0.51 (0.11)	0.44 (0.13)	0.60 (0.12)
	Abstract	0.16 (0.06)	0.37 (0.09)	0.12 (0.06)	0.16 (0.06)	0.09 (0.04)	0.12 (0.05)

M mean test score, *SE* standard error

Results

Table 3 shows children's mean test scores at 3 days, 2 months, and 6 months post-learning for the L1-L2 translation, L2-L1 translation, and free recall tests.

L1-L2 Translation We first tested the hypothesis that gesture enrichment would benefit children's L1-L2 translation test scores compared with the non-enriched learning condition. A three-way ANOVA with the factors learning condition, word type, and time point yielded a significant main effect of learning condition [F (1, 53) = 6.29, $p < 0.05$, $\eta_p^2 = 0.11$]. Children's L1-L2 translation test scores were significantly higher for words learned with gesture enrichment compared with words learned without enrichment (Fig. 4a).

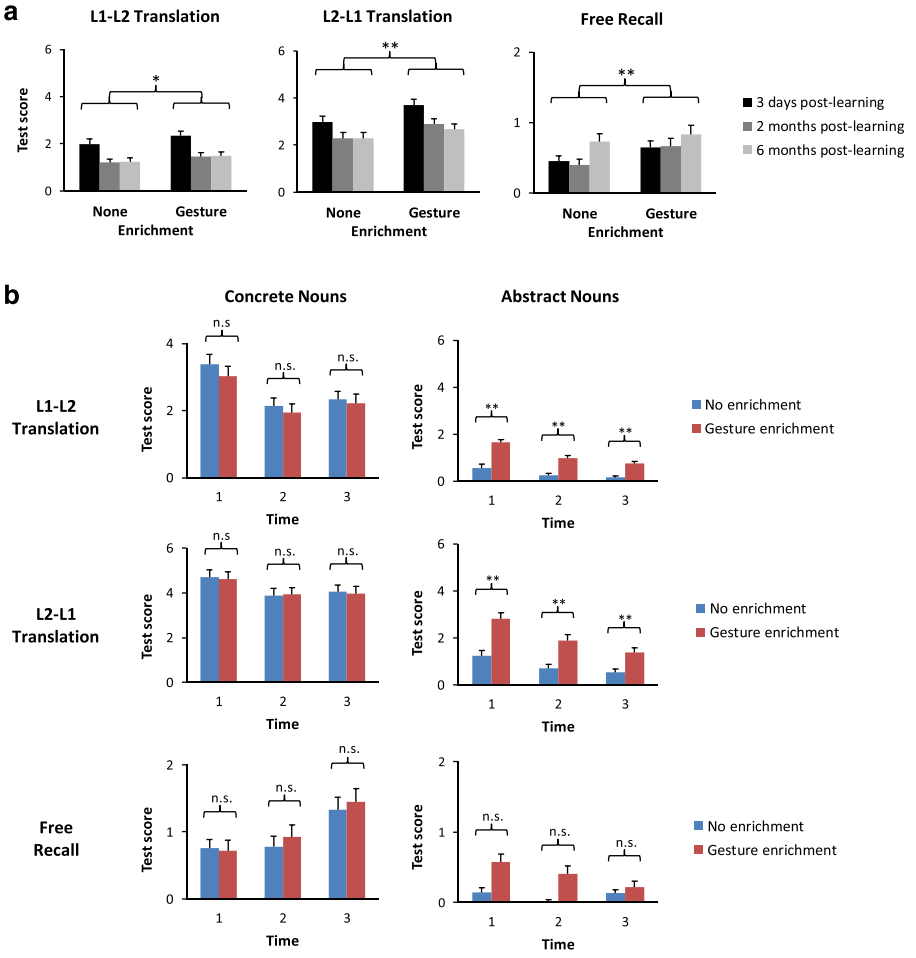

Fig. 4 Experiment 1 results. **a** Higher overall test scores were observed for foreign language nouns that had been learned with gesture enrichment compared to no enrichment for all three test types (L1-L2 translation, L2-L1 translation, and free recall). **b** Higher scores were observed for abstract nouns that had been learned with gesture enrichment compared with no enrichment for both translation tests. Error bars represent one standard error of the means. L1, native language; L2, foreign language. Time 1 , 3 days post-learning; Time 2, 2 months post-learning; and Time 3, 6 months post-learning. *$p < 0.05$, **$p < 0.01$

We next examined whether gesture enrichment benefits were influenced by the class of vocabulary that children had learned. The ANOVA yielded a significant interaction between learning condition and word type factors [F (1, 53) = 22.57, $p < 0.001$, $\eta_p^2 = 0.30$], as well as a significant three-way learning condition × word type × time point interaction [F (2, 106) = 3.51, $p < 0.05$, η_p^2 =0.06]. Contrary to our hypothesis (see hypothesis two in the introduction), significantly higher L1-L2 translation test scores were observed across time points for abstract nouns learned with gesture enrichment compared with abstract nouns learned without enrichment ($p < 0.01$, Hedge's $g = 0.69$). The three-way interaction revealed higher L1-L2 translation scores for abstract nouns learned with gesture enrichment compared with no enrichment at 3 days post-learning ($p < 0.01$, Hedge's $g =$ 0.80), 2 months post-learning ($p < 0.01$, Hedge's $g = 0.65$), and 6 months post-learning ($p < 0.01$, Hedge's $g = 0.68$) (Fig. 4b).

The ANOVA additionally revealed a significant main effect of time point [F (2, 106) = 43.89, $p < 0.001$, $\eta_p^2 = 0.45$]. Tukey's post hoc comparisons revealed that L1-L2 translation scores were significantly higher at 3 days post-learning compared with both 2 months post-learning ($p < 0.01$, Hedge's $g = 0.43$) and 6 months post-learning ($p < 0.01$, Hedge's $g = 0.41$). There was also a significant main effect of word type [F (1, 53) = 163.70, $p < 0.001$, $\eta_p^2 = 0.76$]: Children's L1-L2 translation scores were higher for concrete nouns compared with abstract nouns (Fig. 4b). These main effects were expected due to memory decay over time (Caramelli et al. 2004; Howe and Brainerd 1989) and previous reports of children's greater performance for concrete than abstract nouns (Schwanenflugel 1991). These main effects were qualified by a significant word type × time point interaction [F (2, 106) = 10.00, $p < 0.001$, $\eta_p^2 = 0.16$], which revealed that L1-L2 translation scores were significantly higher for concrete words—but not abstract words—at 6 months post-learning compared with 2 months post-learning ($p < 0.05$, Hedge's $g = 0.13$). There were no other significant main effects or interactions.

L2-L1 Translation We next tested whether gesture enrichment benefitted children's L2-L1 translation test scores compared with the non-enriched learning condition. A three-way ANOVA with the factors learning condition, word type, and time point yielded a significant main effect of learning condition [F (1, 53) = 12.81, $p < 0.01$, $\eta_p^2 = 0.19$]. Children's L1-L2 translation test scores were significantly higher for words learned with gesture enrichment compared with words learned without enrichment (Fig. 4a).

We examined whether gesture enrichment benefits on the L2-L1 translation test were influenced by the class of vocabulary that children had learned. The ANOVA yielded a significant interaction between learning condition and word type factors [F (1, 53) = 22.57, $p < 0.001$, $\eta_p^2 = 0.30$] and a significant three-way learning condition × word type × time point interaction [F (2, 106) = 3.51, $p < 0.05$, $\eta_p^2 = 0.06$]. Contrary to our hypothesis, significantly higher test scores were observed across time points for abstract nouns learned with gesture enrichment compared with abstract nouns learned without enrichment ($p < 0.01$, Hedge's $g =$ 0.73). The three-way interaction revealed higher L2-L1 translation scores for abstract nouns learned with gesture enrichment compared with no enrichment at 3 days post-learning ($p < 0.01$, Hedge's $g = 0.86$), 2 months post-learning ($p < 0.01$, Hedge's $g = 0.75$), and 6 months post-learning ($p < 0.01$, Hedge's $g = 0.66$) (Fig. 4b).

The ANOVA additionally revealed a significant main effect of time point [F (2, 106) = 41.19, $p < 0.001$, $\eta_p^2 = 0.44$]. Tukey's post hoc comparisons revealed that L2-L1 translation scores were significantly higher at 3 days post-learning compared with both 2 months post-

learning ($p < 0.01$, Hedge's $g = 0.30$) and 6 months post-learning ($p < 0.01$, Hedge's $g = 0.34$). There was also a significant main effect of word type [$F(1, 53) = 235.58$, $p < 0.001$, $\eta_p^2 = 0.82$]: Children's L2-L1 translation scores were higher for concrete nouns compared with abstract nouns. These main effects were qualified by a significant word type × time point interaction [$F(2, 106) = 5.64$, $p < 0.01$, $\eta_p^2 = 0.10$], which revealed that L2-L1 translation scores were significantly higher for abstract words—but not concrete words—at 3 days post-learning compared with 6 months post-learning ($p < 0.01$, Hedge's $g = 0.63$). There were no other significant main effects or interactions.

Free Recall We also tested whether gesture enrichment benefitted children's free recall test scores compared with the non-enriched learning condition. A three-way ANOVA with the factors learning condition, word type, and time point yielded a significant main effect of learning condition [$F(1, 53) = 8.19$, $p < 0.01$, $\eta_p^2 = 0.13$]. Similar to both of the translation test scores, children's free recall test scores were significantly higher for words learned with gesture enrichment compared with words learned without enrichment (Fig. 4a).

Lastly, we examined whether gesture enrichment benefits on the free recall test were influenced by the class of vocabulary that children had learned. The learning condition × word type interaction did not reach significance ($p = 0.07$).

The ANOVA additionally revealed a significant main effect of time point [$F(2, 106) = 5.07$, $p < 0.01$, $\eta_p^2 = 0.09$]. Tukey's post hoc comparisons revealed that free recall scores were significantly higher at 6 months post-learning compared with 3 days post-learning ($p < 0.01$, Hedge's $g = 0.21$) and 2 months post-learning ($p < 0.01$, Hedge's $g = 0.22$). This main effect was qualified by a significant word type × time point interaction [$F(2, 106) = 19.15$, $p < 0.001$, $\eta_p^2 = 0.27$], which revealed that free recall scores were significantly higher for concrete words—but not abstract words—at 6 months post-learning compared with 3 days post-learning ($p < 0.01$, Hedge's $g = 0.52$). The increase in free recall performance at 6 months post-learning compared with 3 days post-learning is difficult to explain, as we expected memory for the translations to decay over time.

There was also a significant main effect of word type [$F(1, 53) = 60.42$, $p < 0.001$, $\eta_p^2 = 0.53$]: Children's free recall scores were higher for concrete nouns compared with abstract nouns (Fig. 4b). There were no other significant main effects or interactions.

Analysis of Moderating Effects of Age on Observed Gesture Enrichment Benefits In an exploratory analysis, we tested whether children's ages moderated the observed learning effects. Children in Experiment 1 ranged in age from 8.5 to 10.2 years (M = 8.7 years, SD = 0.5 years). For each dependent variable (free recall scores, L1-L2 translation scores, and L2-L1 translation scores), we used multiple linear regression to generated both a null model and an alternative model of learning outcomes. The null model predicted test scores from children's ages and the learning condition factor (dummy-coded). The alternative model predicted test scores from children's ages, learning condition (dummy-coded), and a moderator term (age multiplied by learning condition). Variance accounted for by the two models was compared using chi-square tests, which revealed no significant differences between the null and alternative models (free recall, $p(\chi 2) = 0.46$; L1-L2 translation, $p(\chi 2) = 0.82$; L2-L1 translation, $p(\chi 2) = 0.92$). These nonsignificant comparisons suggest that children's ages did not moderate the effects of gesture enrichment on vocabulary test performance.

Discussion

The results of Experiment 1 confirmed our first hypothesis that children demonstrated a significant benefit of gesture enrichment on vocabulary learning in comparison to non-enriched learning. This occurred for both the L1-L2 translation and L2-L1 translation tests, as well as the free recall test. Children also demonstrated higher overall performance for concrete vocabulary compared with abstract vocabulary for all three test types, consistent with previous reports (Schwanenflugel 1991). Unexpectedly in relation to the previous findings in adults (Mayer et al. 2015), enrichment benefits were not equivalent across classes of foreign vocabulary: The performance of gestures during learning significantly aided the subsequent L1-L2 and L2-L1 translation of abstract words, but not concrete words, at all time points. The greater effectiveness of gesture enrichment for abstract words compared with concrete words occurred for both the L1-L2 and L2-L1 translation tests, but not for the free recall test. We performed Experiment 2 to test the robustness of these effects.

Experiment 2

The relative superiority of the benefits of gesture enrichment for the learning of abstract words as opposed to concrete words observed in Experiment 1 was not predicted based on the previous results in adults (Mayer et al. 2015), and overall test accuracy in Experiment 1 was low considering the total number of stimuli included in the learning phase (Table 2). We therefore sought to replicate the findings of Experiment 1 in Experiment 2 using a subset of the stimuli used in Experiment 1 (24 German words and 24 English translations). We assumed that the presentation of fewer words overall would lead to higher performance (Dennis et al. 2008).

Methods

Participants Fifty children participated in Experiment 2. Although the children were from the same school as the children who participated in Experiment 1, no child who participated in Experiment 1 also participated in Experiment 2. The investigators briefed all children and teachers on the study procedures in an introductory session that took place prior to the experiment. Children who were absent from at least one training or test session were excluded from the analyses. Therefore, 43 children were included in the analyses. Written informed consent was obtained from the legal guardians of all individual school children who participated. None of the children were reported by the school principal or teachers to possess learning disabilities. All children possessed normal or corrected-to-normal vision. Demographic information can be found in Table 1. Experiment 2 was reviewed and approved by the Education Department of the state of Saxony, Germany.

Stimulus Materials Experiment 2 included a subset of 24 of the English words used in Experiment 1 (Table 2). Half of the words were concrete nouns and the other half were abstract nouns. The concrete and abstract nouns significantly differed in terms of concreteness on the basis of ratings from a corpus of German lemmas (Köper and Schulte Im Walde 2016), t

(22) = 8.98, $p < 0.001$. The selection of the subset of words included in Experiment 2 was based on the recommendations of the children's schoolteachers. None of the words included in Experiment 1 had been previously encountered during school lessons and were not anticipated to be included in teaching curriculum for the 6-month duration of the investigation. Therefore, the teachers identified words from Experiment 1 such as *backpack* or *screen* that they believed the children were most likely to encounter outside of the school environment in broadcast or Internet media or in multilingual environments such as airports. These words were removed in Experiment 2. As in Experiment 1, word frequencies in written German (http://wortschatz.uni-leipzig.de/en), as well as the numbers of syllables and letters contained in the English translations, were counterbalanced between learning conditions in all experiments. Word frequencies and lengths were also counterbalanced between sets of concrete and abstract words included in the two learning conditions.

Using concreteness ratings from Köper and Schulte im Walde (2016), we tested whether the concreteness of concrete and abstract words used in Experiment 1 (concrete nouns, M = 6.58, SD = 0.98, range, 4.07–7.99; abstract nouns, M = 2.95, SD = 0.76, range, 1.80–4.61) differed from the concreteness of the subset of concrete and abstract words used in Experiment 2 (concrete nouns, M = 6.36, SD = 1.04, range, 4.07–7.75; abstract nouns, M = 2.95, SD = 0.80, range, 1.84–4.61). There was no significant difference in concreteness between the concrete words used in Experiment 1 and the subset of concrete words used in Experiment 2, t (30) = 0.61, $p = 0.55$. There was also no significant difference in concreteness between the abstract words used in Experiment 1 and the subset of abstract words used in Experiments 2, t (30) = 0.01, $p = 0.99$.

Experiment 2 made use of the same audio recordings and videos that were used in Experiment 1 corresponding to the reduced set of 24 German words and 24 English translations.

Design and Procedure Experiment 2 had a 2 × 2 × 3 factorial within-subjects design with the factors learning condition (gesture enrichment, no enrichment), word type (concrete, abstract), and testing time point (3 days, 2 months, and 6 months post-learning).

Learning Phase The learning phase followed the same procedure as Experiment 1 with one minor difference. Instead of reciting only the English word aloud with their teacher at the end of each trial, children recited both the German and English words aloud for both learning conditions. This change in comparison to Experiment 1 was made as we assumed that it would benefit learning outcomes.

Learning phase trials were again blocked by learning condition. Children completed 4 blocks containing 12 trials (6 concrete word trials and 6 abstract word trials) per day. Half of the blocks comprised gesture enrichment, and the other half comprised non-enriched learning. Unlike Experiment 1, each German word and its English translation were presented in two trials each day. Thus, words were heard twice as often in Experiment 2 compared with Experiment 1. Word orders within each block and orders of enrichment condition blocks were counterbalanced across learning days.

Test phase The test phase followed the same procedure as Experiment 1.

Data analysis See Experiment 1.

Results

Table 3 shows children's mean test scores for the L1-L2 translation, L2-L1 translation, and free recall tests by experimental condition in Experiment 2.

L1-L2 Translation We first tested the hypothesis that gesture enrichment would aid L1-L2 translation performance compared with non-enriched learning. A three-way ANOVA with the factors learning condition, word type, and time point on L1-L2 translation test scores yielded a significant main effect of learning condition [F (1, 42) = 8.98, $p < 0.01$, $\eta_p^2 = 0.18$]. As observed in Experiment 1, and in agreement with our first hypothesis (see Introduction), L1-L2 translation scores in Experiment 2 were significantly higher for the gesture enrichment compared with the no enrichment condition (Fig. 5a).

We examined whether the specificity of gesture benefits for the L1-L2 translation of abstract words that occurred in Experiment 1 also occurred in Experiment 2. Unlike Experiment 1, the three-way ANOVA did not yield a significant interaction between learning condition and word type variables ($p = 0.24$) or significant three-way interaction ($p = 0.50$), suggesting that benefits of enrichment on the L1-L2 translation test did not differ across the two word types (Fig. 5b).

As in Experiment 1, the ANOVA revealed a significant main effect of time point [F (2, 84) = 56.16, $p < 0.001$, $\eta_p^2 = 0.57$]. Tukey's post hoc comparisons revealed that L1-L2 translation scores were significantly higher at 3 days post-learning than at 2 months post-learning ($p < 0.01$, Hedge's $g = 0.58$) and 6 months post-learning ($p < 0.01$, Hedge's $g = 0.85$). There was also a significant main effect of word type [F (1, 42) = 23.44, $p < 0.001$, $\eta_p^2 = 0.36$]: L1-L2 translation test scores were significantly higher for concrete nouns compared with abstract nouns. There were no other significant main effects or interactions.

L2-L1 Translation We next tested the hypothesis that gesture enrichment aided L2-L1 translation compared with non-enriched learning. A three-way ANOVA with the factors learning condition, word type, and time point on L2-L1 translation test scores yielded a significant main effect of learning condition [F (1, 42) = 26.42, $p < 0.001$, $\eta_p^2 = 0.39$]. As observed in Experiment 1, L2-L1 translation scores in Experiment 2 were significantly higher for the gesture enrichment compared with the no enrichment condition (Fig. 5a).

We next examined whether the specificity of gesture benefits for the L2-L1 translation of abstract words that occurred in Experiment 1 also occurred in Experiment 2. Unlike Experiment 1, the three-way ANOVA did not yield a significant interaction between learning condition and word type variables ($p = 0.61$) or significant three-way interaction ($p = 0.31$), suggesting that benefits of enrichment on the L2-L1 translation test did not differ across the two word types (Fig. 5b).

As in Experiment 1, the ANOVA revealed a significant main effect of time point [F (2, 84) = 81.15, $p < 0.001$, $\eta_p^2 = 0.09$]. Tukey's post hoc comparisons revealed that L2-L1 translation scores were significantly higher at 3 days post-learning than at 2 months post-learning ($p < 0.01$, Hedge's $g = 0.34$) and significantly higher at 2 months post-learning than at 6 months post-learning ($p < 0.01$, Hedge's $g = 0.28$).

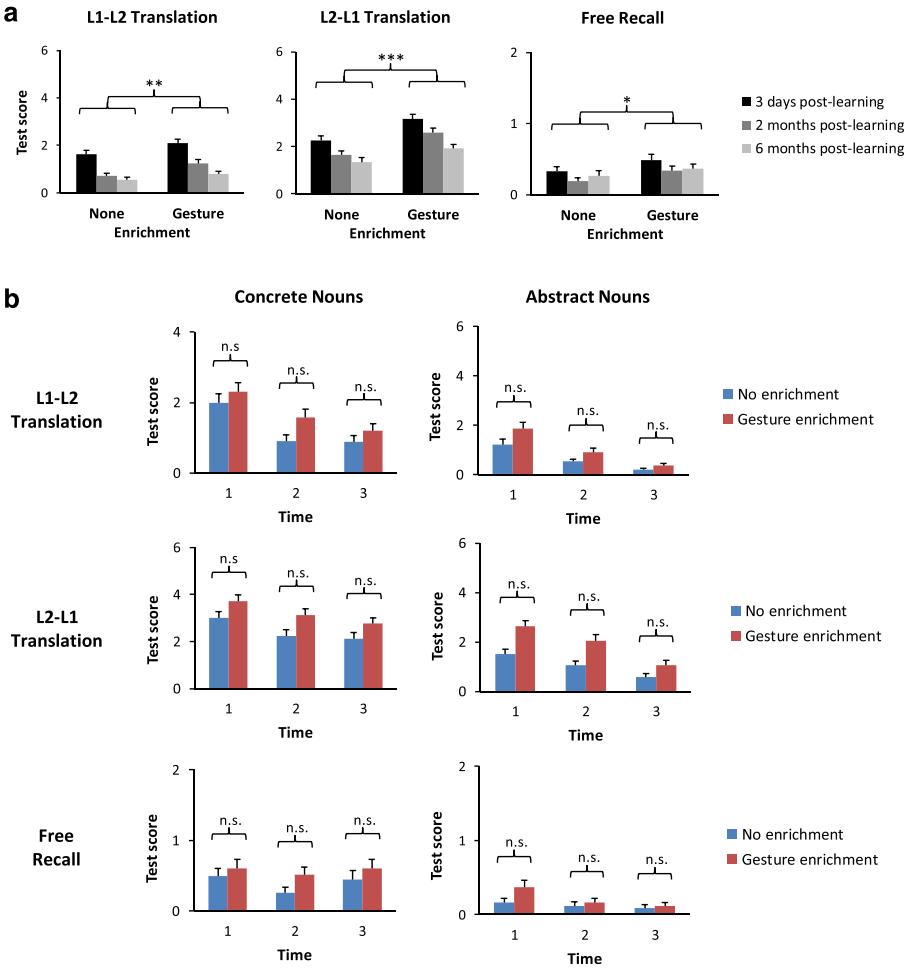

Fig. 5 Experiment 2 results. **a** Higher overall test scores were observed for foreign language nouns that had been learned with gesture enrichment compared to no enrichment for all three test types (L1-L2 translation, L2-L1 translation, and free recall). **b** The enrichment condition factor did not significantly interact with the word type or time point factors for any of the three test types. Error bars represent one standard error of the means. L1, native language; L2, foreign language. Time 1 , 3 days post-learning; Time 2 , 2 months post-learning; and Time 3 , 6 months post-learning. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

There was also a significant main effect of word type [F (1, 42) = 96.50, $p < 0.001$, $\eta_p^2 = 0.70$]: L2-L1 translation test scores were significantly higher for concrete nouns compared with abstract nouns. There were no other significant main effects or interactions.

Free Recall We also tested whether gesture enrichment aided free recall performance compared with non-enriched learning, as was the case in Experiment 1. A three-way ANOVA with the factors learning condition, word type, and time point on free recall test scores yielded a significant main effect of learning condition [F (1, 42) = 5.02, $p < 0.05$, $\eta_p^2 = 0.11$]. As observed in Experiment 1, free recall scores in Experiment 2 were significantly higher for the gesture enrichment compared with the no enrichment condition (Fig. 5a).

We next examined whether the overall gesture benefit differed for concrete and abstract words. The three-way ANOVA did not yield a significant interaction between learning condition and word type variables ($p = 0.34$), suggesting that benefits of enrichment on the free recall test did not differ across the two word types (Fig. 5b).

The main effect of time point did not reach significance ($p = 0.06$). There was, however, a significant main effect of word type [$F(1, 42) = 17.88$, $p < 0.001$, $\eta_p^2 = 0.30$]: Free recall scores were significantly higher for concrete nouns compared with abstract nouns. There were no other significant main effects or interactions.

Analysis of Moderating Effects of Age on Observed Gesture Enrichment Benefits We again explored whether children's ages moderated the observed learning effects by generating null and alternative multiple linear regression models of free recall, L1-L2 translation, and L2-L1 translation outcomes. Children in Experiment 2 ranged in age from 8.0 to 10.0 years (M = 8.3 years, SD = 0.4 years). The null model predicted test scores from children's ages and the learning condition factor (dummy-coded). The alternative model predicted test scores from children's ages, learning condition (dummy-coded), and a moderator term (age multiplied by learning condition). Variance accounted for by the two models was compared using chi-square tests, which revealed no significant differences between the null and alternative models (free recall, $p(\chi 2) = 0.98$; L1-L2 translation, $p(\chi 2) = 0.38$; L2-L1 translation, $p(\chi 2) = 0.56$). These nonsignificant comparisons suggest that children's ages did not moderate the effects of gesture enrichment on vocabulary test performance.

Discussion

In sum, Experiment 2 replicated Experiment 1 in that children benefitted from gesture enrichment in vocabulary learning in contrast to non-enriched learning. Additionally, as in Experiment 1, children's overall test performance was higher for concrete vocabulary compared with abstract vocabulary. In Experiment 2, there were no interactions of the learning condition factor with time point or word type factors, suggesting that gesture enrichment benefitted both word types across the three testing time points. This was the case for all three test types: L1-L2 translation, L2-L1 translation, and free recall. This differed from the beneficial effects of gesture enrichment observed in Experiment 1, which occurred only for abstract words at all testing time points in the case of the L1-L2 and L2-L1 translation tests.

One contributor to the variable effects of gesture enrichment on concrete and abstract words in the two experiments may be the reduction in statistical power caused by separating the data by time point. Additionally, we speculate based on children's higher overall test scores for concrete words in Experiment 1 compared with Experiment 2 that the inclusion of L2 words such as *backpack* and *screen* in Experiment 1 (but not Experiment 2) may have contributed to the lack of an effect of gesture enrichment on the learning of concrete words in Experiment 1. Although these words had not yet appeared in L2 lessons in the classroom, children may have been exposed to them outside of the classroom before or during the timeframe of Experiment 1. The exclusion of these words from Experiment 2 led to a nonsignificant reduction in concreteness of the concrete words used in Experiment 2 compared with Experiment 1, which may also have contributed to differences in effects of concreteness across the two experiments. Concrete words included in Experiment 1 received a rating of M = 6.58 (SD = 0.98) in terms of

concreteness, and concrete words included in Experiment 2 received a rating of M = 6.36 (SD = 1.04), a difference of 0.22 points (Köper and Schulte Im Walde 2016; see Experiment 2 methods for more information).

Experiment 3

After establishing that gesture enrichment benefitted school children's learning of L2 vocabulary in Experiments 1 and 2, we used the same training paradigm in Experiment 3 to test our main hypothesis that gesture enrichment would benefit children's L2 learning more than picture enrichment, particularly over the long term. This hypothesis was based on a previous study in adults that observed a greater long-term (6 months post-learning) benefit of gesture-enriched learning than picture-enriched learning (Mayer et al. 2015).

Methods

Participants Sixty-two children participated in Experiment 3. Although they were from the same school as in the other two experiments, no children who participated in Experiments 1 or 2 also participated in Experiment 3. The investigators briefed all children and teachers on the study procedures in an introductory session that took place prior to the experiment. Children who were absent from at least one training or test session were excluded from the analyses. Therefore, 51 children were included in the analyses. Written informed consent was obtained from the legal guardians of all individual school children who participated. None of the children were reported by the school principal or teachers to possess learning disabilities. All children possessed normal or corrected-to-normal vision. Demographic information can be found in Table 1. Experiment 3 was reviewed and approved by the Education Department of the state of Saxony, Germany.

Stimulus Materials Experiment 3 made use of the same 24 German words and 24 English translations, audio recordings, and videos that were used in Experiment 2 (Table 2).

Pictures for the picture enrichment condition consisted of black and white line drawings created by a professional cartoon artist. The drawings iconically communicated word meanings by depicting objects, humans, or scenes. Abstract nouns were conveyed using scenes. Pictures representing one of the concrete nouns and one of the abstract nouns are shown in Fig. 1. The complexity of line drawings was not matched for concrete and abstract nouns, as differences in complexity are also expected to occur in naturalistic teaching settings.

Design and Procedure Experiment 3 had a 2 × 2 × 3 factorial within-subjects design with the factors learning condition (gesture enrichment, picture enrichment), word type (concrete, abstract), and testing time point (3 days, 2 months, and 6 months post-learning).

Learning Phase As in Experiments 1 and 2, children completed 5 consecutive days of foreign language training (Fig. 2). Training occurred over four 5-min blocks per day. Rest activities took place for 5 min between each of the blocks. Thus, each daily training session had a total duration of approximately 35 min.

L2 words and their L1 translations were presented in gesture enrichment and picture enrichment trials (Fig. 3). Gesture enrichment trials followed the procedure used in

Experiments 1 and 2. In picture enrichment trials, recorded English words were accompanied by iconic line drawings. As in the gesture enrichment trials, children first heard an English word, which was followed by its auditorily presented German translation and a repetition of the English word. Pictures were presented for 3 s. As in Experiment 2, the children and their teacher recited both the German and English words aloud for both gesture and picture conditions. Children were seated at their desks during picture enrichment trials. An additional motor task was not included in the picture condition as we were interested in testing the effects of gesture performance against picture viewing. Pictures used in enrichment materials such as flash cards are typically only viewed without also performing motor tasks during L2 lessons (Cook 2008). Further, the combination of picture viewing with a motor task (tracing an outline of presented pictures) has been shown to be less beneficial for learning than simply viewing the pictures without performing a motor task in adults (Mayer et al. 2015).

Learning phase trials were again blocked by learning condition. Children completed 4 blocks containing 12 trials (6 concrete word trials and 6 abstract word trials) per day. Half of the blocks comprised gesture enrichment, and the other half comprised picture enrichment. As in Experiment 2, each German word and its English translation were presented in two trials each day. Thus, words were heard twice as often in Experiments 3 compared with Experiment 1. Word orders within each block and orders of enrichment condition blocks were counterbalanced across learning days.

Test phase The test phase followed the same procedure as Experiments 1 and 2.

Data Analysis Data were analyzed following the same procedures used in Experiments 1 and 2. In order to evaluate effects of enrichment and vocabulary type on learning, three-way analyses of variance (ANOVAs) on test scores with the factors enrichment condition (gesture, picture), word type (concrete and abstract), and time point (3 days post-learning, 2 months post-learning, and 6 months post-learning) were conducted.

Results

Table 4 shows children's mean test scores for the L1-L2 translation, L2-L1 translation, and free recall tests by experimental condition in Experiment 3.

L1-L2 Translation We first tested the hypothesis that gesture enrichment benefitted L1-L2 translation performance compared with picture enrichment. An ANOVA with factors learning condition, word type, and time point on L1-L2 translation test scores yielded a nonsignificant main effect of learning condition ($p = 0.12$) (Fig. 6a). This result did not support the hypothesis that gesture enrichment would enhance learning outcomes.

To evaluate the hypothesis that L1-L2 translation scores for gesture-enriched words exceeded those of picture-enriched words at later time points, we next examined interactions between the learning condition and time point factors. The two-way interaction of learning condition and time point factors was not significant ($p = 0.31$). The three-way learning condition × word type × time point interaction, however, was significant [$F(2, 100) = 3.97$, $p < 0.05$, $\eta_p^2 = 0.07$]. Tukey's post hoc comparisons revealed no significant differences between gesture enrichment and picture enrichment learning outcomes for either vocabulary type at any time point. The three-way interaction was driven

Table 4 Mean scores on German-English translation, English-German translation, and free recall tests in Experiment 3

		3 days post-learning		2 months post-learning		6 months post-learning	
		Picture M (SE)	Gesture M (SE)	Picture M (SE)	Gesture M (SE)	Picture M (SE)	Gesture M (SE)
Exp. 3	German-English translation						
	Concrete	2.51 (0.22)	2.20 (0.22)	1.18 (0.14)	1.18 (0.14)	0.86 (0.11)	0.65 (0.14)
	Abstract	1.96 (0.21)	2.16 (0.22)	1.02 (0.14)	0.78 (0.13)	0.86 (0.11)	0.41 (0.08)
	English-German translation						
	Concrete	4.29 (0.18)	3.84 (0.20)	3.27 (0.21)	3.00 (0.22)	2.49 (0.18)	2.20 (0.21)
	Abstract	3.08 (0.20)	2.98 (0.19)	2.02 (0.20)	1.75 (0.19)	1.10 (0.16)	1.16 (0.17)
	Free recall						
	Concrete	0.51 (0.11)	0.57 (0.11)	0.33 (0.09)	0.31 (0.09)	0.39 (0.08)	0.29 (0.10)
	Abstract	0.59 (0.14)	0.41 (0.11)	0.35 (0.09)	0.27 (0.08)	0.24 (0.08)	0.22 (0.06)

M mean test score, *SE* standard error

by significantly higher scores for concrete picture-enriched words at 3 days post-learning compared with abstract picture-enriched words at 3 days post-learning ($p < 0.05$, Hedge's $g = 0.36$), which did not occur for the gesture-enriched words.

As observed in Experiments 1 and 2, there was a significant main effect of time point [F (2, 100) = 108.66, $p < 0.001$, $\eta_p^2 = 0.68$]. Tukey's post hoc comparisons revealed that L1-L2 translation scores were significantly higher at 3 days post-learning than at 2 months post-learning ($p < 0.01$, Hedge's $g = 0.89$) and at 2 months post-learning compared with 6 months post-learning ($p < 0.01$; Hedge's $g = 0.37$). There was also a significant main effect of vocabulary type [F (1, 50) = 8.90, $p < 0.01$, $\eta_p^2 = 0.15$]: L1-L2 translation test scores were significantly higher for concrete words compared with abstract words. There were no other significant main effects or interactions.

L2-L1 Translation We next tested the hypothesis that gesture enrichment benefitted L2-L1 translation compared with picture enrichment. An ANOVA with factors learning condition, word type, and time point on L2-L1 translation test scores yielded a nonsignificant main effect of learning condition ($p = 0.12$) (Fig. 6a). This result did not support the hypothesis that gesture enrichment would enhance learning outcomes.

To evaluate the hypothesis that L2-L1 translation scores for gesture-enriched words exceeded those of picture-enriched words at later time points, we next examined interactions between the learning condition and time point factors. The two-way interaction of learning condition and time point factors was not significant ($p = 0.61$), and the three-way learning condition × word type × time point interaction was also not significant ($p = 0.44$) (Fig. 6b).

As observed in Experiments 1 and 2, there was a significant main effect of time point [F (2, 100) = 231.21, $p < 0.001$, $\eta_p^2 = 0.82$]. Tukey's post hoc comparisons revealed that L2-L1 translation scores were significantly higher at 3 days post-learning than at 2 months post-learning ($p < 0.01$, Hedge's $g = 0.67$) and at 2 months post-learning compared with 6 months post-learning ($p < 0.01$; Hedge's $g = 0.51$). There was also a significant main effect of vocabulary type [F (1, 50) = 88.93, $p < 0.001$, $\eta_p^2 = 0.64$]: L2-L1 translation test scores were significantly higher for concrete words compared with abstract words. There were no other significant main effects or interactions.

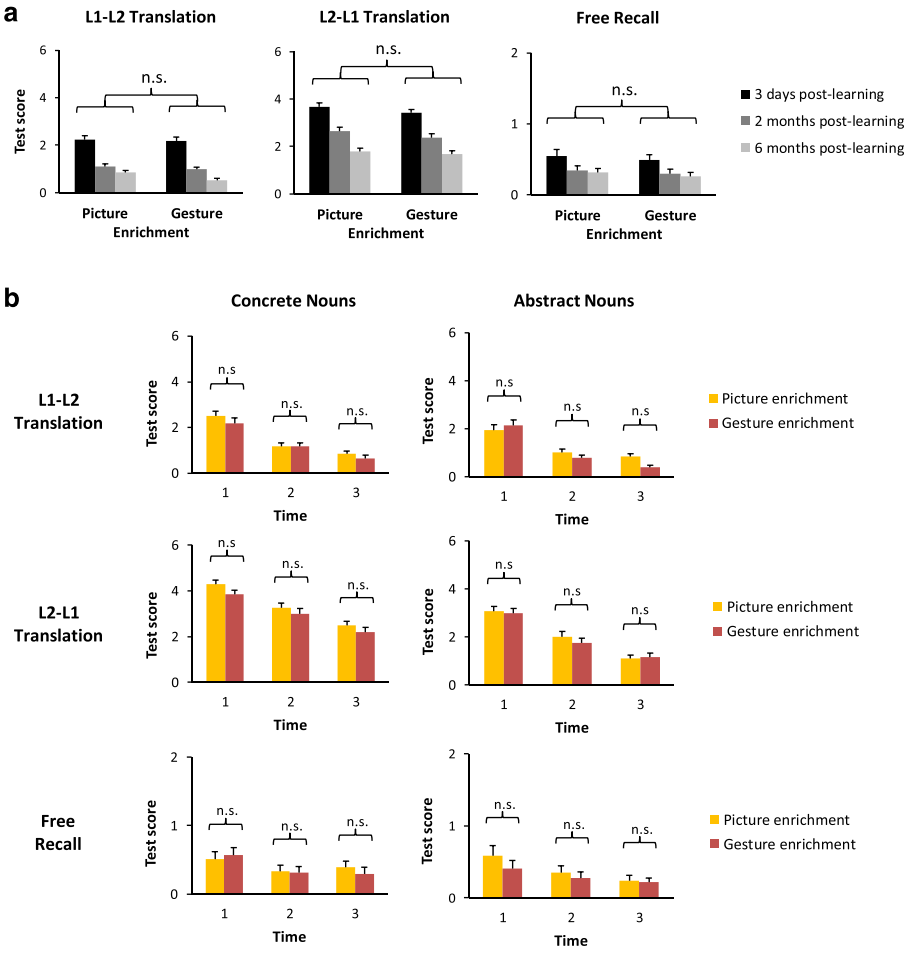

Fig. 6 Experiment 3 results. **a** No significant differences in test scores were observed for foreign language nouns that had been learned with picture enrichment compared with gesture enrichment. **b** The enrichment condition factor did not interact with word type or time point factors for any of the three test types. L1, native language; L2, foreign language. Time 1, 3 days post-learning; Time 2, 2 months post-learning; and Time 3, 6 months post-learning

Free Recall We also tested the hypothesis that gesture enrichment would benefit free recall performance compared with picture enrichment. An ANOVA with factors learning condition, word type, and time point on free recall test scores yielded a nonsignificant main effect of learning condition ($p = 0.47$) (Fig. 6a). This result did not support the hypothesis that gesture enrichment would enhance learning outcomes.

To evaluate the hypothesis that free recall scores for gesture-enriched words exceeded those of picture-enriched words at later time points, we next examined interactions between the learning condition and time point factors. The two-way interaction of learning condition and time point factors was not significant ($p = 0.99$), and the three-way learning condition × word type × time point interaction was also not significant ($p = 0.35$) (Fig. 6b).

As observed in Experiments 1 and 2, there was a significant main effect of time point, [F (2, 100) = 10.84, $p < 0.001$, $\eta_p^2 = 0.18$]. Tukey's post hoc comparisons revealed that free recall scores were significantly higher at 3 days post-learning than at 2 months post-learning ($p < 0.01$, Hedge's $g = 0.27$) and 6 months post-learning ($p < 0.01$; Hedge's $g = 0.32$). There were no other significant main effects or interactions.

Discussion

Overall, gesture enrichment did not significantly benefit learning outcomes compared with picture enrichment for any of the three test types. This result did not support our hypothesis that gesture enrichment would facilitate the learning of L2 vocabulary more than picture enrichment, particularly over the long term. L1-L2 translation performance following picture-enriched learning was sensitive to the word type variable only at the earliest time point (3 days post-learning). In contrast, free recall and L2-L1 translation scores were not influenced by word type following gesture-enriched learning or picture-enriched learning.

General Discussion

We examined the effects of gesture and picture enrichment on 8-year-old school children's L2 vocabulary learning in three experiments. There were four main findings. First, as predicted, gesture enrichment enhanced L2 vocabulary learning compared with non-enriched learning. This occurred for cued L1 and L2 translation as well as free recall of L1-L2 translations. Second, compared with non-enriched learning, gesture enrichment benefitted the learning of both concrete nouns (e.g., *tent*) and abstract nouns (e.g., *thought*). The benefit for concrete nouns was variable; it was present for children who participated in Experiment 2 but not for children who participated in Experiment 1. Third, gesture enrichment enhanced learning outcomes compared with non-enriched learning in both the short term (3 days post-learning) and long term (6 months post-learning). Fourth, contrary to our hypotheses, gesture-enriched learning did not yield stronger translation outcomes compared with picture-enriched learning. Taken together, these findings suggest that the viewing of iconic pictures by 8-year-old children during L2 vocabulary learning is just as beneficial for learning outcomes as the performance of iconic gestures. This contrasts with previous findings in adults, for whom gestures were more beneficial than pictures for cued L1 and L2 translation, particularly over the long term (Mayer et al. 2015).

Gesture Enrichment Enhanced Learning Outcomes

Compared with non-enriched learning of L2 vocabulary, gesture-enriched learning has been shown to enhance subsequent translation performance by about 10% in adults (Mayer et al. 2015). The benefits observed for 8-year-old school children in the current study exceed the benefits observed previously in adults: Across all tests, word types, and time points in Experiments 1 and 2, gesture-enriched learning increased children's test scores by about 14% compared with non-enriched learning. Children received minimal word exposure (only a single exposure per day in Experiment 1 and two exposures per day in Experiments 2 and 3). They also never viewed written words, which makes the current learning procedure

comparable to L1 acquisition apart from the children's ages. However, the two primary factors that are known to drive L1 vocabulary acquisition in pre-literate children—parent speech (Hart and Risley 2003) and verbal context (Weizman and Snow 2001)—were entirely absent from the current learning environment. Despite these hurdles, children were able to make use of the enrichment in a way that supported subsequent vocabulary knowledge.

Positive effects of gesture-enriched learning persisted up to 6 months following the 5-day learning period, suggesting that enrichment effects were robust enough to support translation after long delays in L2 usage. Note that children's schoolteachers omitted the L2 stimuli from English lessons for 6 months following the learning phase in each of the experiments. Thus, children did not receive any intermittent vocabulary training leading up to the tests that took place 2 and 6 months post-learning. Such long-lasting effects of enrichment have not previously been investigated in children and suggest that gesture-enriched learning contributes to robust L2 vocabulary representations in memory.

Beneficial effects of gesture enrichment on L2 vocabulary learning are consistent with a variety of psychological accounts described in the introduction. According to an embodied or grounded cognition perspective, gesture enrichment could have enhanced children's memory for L2 words by grounding the newly acquired sequences of sounds in sensorimotor experiences (cf. Barsalou 2008; Kiefer and Trumpp 2012; Macedonia & von Kriegstein 2012). From a dual coding perspective (Engelkamp and Zimmer 1984; Hommel et al. 2001; Paivio 1991; Paivio and Csapo 1969), both L1 and L2 words were likely encoded verbally, while the viewing of pictures generated nonverbal visual encoding, and the performance of gestures generated both visual and haptic encodings. The findings can also be explained by simulation or imagery mechanisms (Jeannerod 1995; Kosslyn et al. 2006; Saltz and Dixon 1982): Simply hearing an L1 or L2 word at test may have triggered reconstructions of gesture or picture enrichment material, which could have enhanced children's performance by establishing more routes to retrieving the correct translation.

Despite the proportionally large overall improvement in learning outcomes following gesture learning compared with non-enriched learning, the overall amount of translations remembered by the children was low considering the total number of L2 stimuli. Test scores tended to be particularly low for freely recalled abstract word translations. These scores are on par with similar previous studies. For example, Mavilidi et al. (2015) report that 5-year-olds taught 14 Italian action words in four 15-min training sessions spread across 2 weeks recalled a mean of 0.98 words (SD = 0.81 words) when the words were taught by combining the performance of gestures with physical exercise. Low test scores in these training studies are likely attributable to the relatively short time span of the training periods; further studies are needed to better understand how test scores may scale up with longer training time spans, such as enrichment interventions lasting a year or longer.

Gesture Enrichment Did Not Benefit 8-Year-Old School Children's L2 Vocabulary Learning More Than Picture Enrichment

Contrary to our hypotheses, gesture-enriched learning did not benefit L1-L2 translation or L2-L1 translation compared with picture-enriched learning. Tellier (2008) observed greater memory for gesture-enriched compared with picture-enriched L2 words in 4- and 5-year-olds. Children in Tellier's study were exposed to 6 repetitions of each L2 word during three training sessions spread over 3 weeks. The children showed a gesture advantage when they were asked to remember words by viewing the learned pictures or gestures. However, when the children

heard an L2 word at test and were asked to produce its associated gesture or identify its associated picture, performance was equivalent for gesture- and picture-enriched vocabulary. This set of results suggests that observed benefits of gesture- and picture-enriched learning may partially depend on how vocabulary knowledge is tested. By using three different test types, the present study lessened the influence of any individual test on overall outcomes and tested vocabulary knowledge in a similar way to previous studies on L2 learning in adults (Mayer et al. 2015; Repetto et al. 2017).

Gesture-enriched learning is fundamentally more active than picture-enriched learning; this could have led to differences between these conditions in terms of children's motivation, cognitive effort, attention, and arousal, as well as the enrichment's long-term effects (Appleton et al. 2008). However, any potential differences between gesture and picture enrichment conditions were not large enough to distinguish the two types of enrichment at test. Previously, the combination of a motor task with picture viewing (tracing an outline of presented pictures), while hearing foreign language words, yielded weaker effects than viewing pictures without performing any movements (Mayer et al. 2015). In another study, the performance of semantically related gestures during L2 vocabulary learning enhanced memory for the vocabulary, but the performance of meaningless gestures did not (Macedonia et al. 2011). These outcomes suggest that gesture enrichment benefits cannot be explained simply by the presence of movement during learning. Additionally, simply viewing gestures during learning without performing them did not enhance performance compared to non-enriched learning, suggesting that picture enrichment benefits cannot be explained simply by the presence of visual information during learning (Mayer et al. 2015).

One factor that remains to be explored is whether object-directed gestures may further enhance enrichment benefits compared to other types of gestures. In previous studies, tracing pictures with one's index finger in the air (Mayer et al. 2015) and performing meaningless gestures (Macedonia et al. 2011) did not benefit L2 vocabulary learning. The gestures performed by children in the current study were iconic gestures, indicating that they displayed some aspect of the associated word's semantics (McNeill 1992). Object-directed gestures, which entail meaningful actions toward a referent, may be even more powerful an aid for learning than iconic gestures (c.f., v. Soden-Fraunhofen et al. 2008). However, this type of gesture is more readily applied to the learning of concrete rather than abstract vocabulary, as abstract words possess intangible referents.

Potential Effects of Stimulus Timing

The lengths of gesture-, picture-, and non-enriched trials in the current study varied: Gestures were presented for 4.0 s, pictures for 3.5 s, and spoken English words for 2.5 s. A shorter time interval was used for the presentation of the spoken English words in the non-enriched learning condition, compared with videos in the gesture-enriched learning condition, in order to avoid a large difference in inter-stimulus intervals between the two conditions. Long time periods during which no sensory information is presented could have the effect of decreasing attention, motivation, or stimulus-driven arousal in the non-enriched task compared with the other tasks.

If trial duration was the primary factor driving enrichment benefits, we would have expected greater benefits of gesture enrichment compared with picture enrichment, as the time intervals during which gestures were presented (4.0 s) were 60% longer than the time intervals during which spoken English words were presented in non-enriched trials (2.5 s), and the time intervals during which pictures were presented (3.0 s) were only 20% longer than time

intervals allotted for spoken English words in non-enriched trials (2.5 s). However, we observed similar levels of performance following both gesture and picture enrichment. If trial duration was the primary factor driving learning outcomes, we would also have expected enhanced performance following gesture-enriched learning compared with picture-enriched learning in Experiment 3, as the time intervals during which gestures were presented (4.0 s) were 33% longer than the intervals during which pictures were presented (3.0 s). However, we observed higher test scores following picture-enriched learning compared with gesture-enriched learning in Experiment 3.

Mayer et al. (2015) presented young adults with the same learning conditions (gesture-, picture-, and non-enriched learning) and maintained a constant trial length across conditions. Even when trial lengths were identical across conditions, participants showed enhanced L2 vocabulary learning for the enriched compared with non-enriched conditions.

Role of L2 Vocabulary Type and Test Type

In Experiment 1, gesture enrichment benefitted the learning of abstract nouns and did not benefit to the same extent the learning of concrete nouns. In Experiment 2, such a difference could not be replicated. Instead, gesture enrichment benefitted the learning of both concrete and abstract nouns at 3 days and 2 months post-learning. Abstract concepts pose a challenge for theories of grounded or embodied cognition because they do not possess a perceivable referent (for a review see Borghi et al. 2017). The finding that gesture enrichment benefitted the learning of abstract words in two experiments is consistent with recent evidence that, like concrete concepts, abstract concepts may be grounded in perception and action (Harpaintner et al. 2020; Harpaintner et al. 2018). Performance on concrete words was also consistently higher than performance on abstract words throughout the three experiments (except for the free recall test in Experiment 3). This was expected as nouns are more difficult for children to comprehend and produce if they possess low conceptual perceptibility (e.g., tangibility, visibility), as is the case for abstract nouns (McFalls et al. 1996). Adult learners also demonstrate a concrete L2 word learning advantage (Macedonia and Knösche 2011).

Effects of vocabulary type on learning outcomes can also be phrased in terms of dual coding theory (Hommel et al. 2001; Paivio 1991). In Experiments 1 and 2, only the gesture-enriched learning condition provided children with dual-coded learning. In Experiment 3, both enrichment conditions involved dual-coded learning, i.e., learning that linked verbal stimuli (spoken words) and nonverbal stimuli (gestures and pictures). The advantage of gesture-enriched compared with non-enriched learning in Experiments 1 and 2 suggests that dual-coded learning concretized the abstract words. However, scores for concrete words tended to be higher than scores for abstract words even in cases of dual-coded learning. This could be the case because concrete words may evoke internal mental simulations, such as the image of a traffic light, even without the provision of an associated gesture or picture (Richardson 2003; Schwanenflugel 1991). In sum, even though dual-coded learning facilitated the concretization of abstract words, concrete words maintained an advantage compared with abstract words at test.

Experiment 1 and 2 results also differed with respect to children's pattern of performance across cued translation and free recall tests. In Experiment 1, gesture-enriched learning benefitted children's L1-L2 and L2-L1 translation compared with non-enriched learning, but did not benefit children's free recall performance. In Experiment 2, however, benefits of gesture-enriched learning were observed for all three test types. The absence of a gesture benefit on free recall outcomes in

Experiment 1 is inconsistent with previous findings in adults (Macedonia and Knösche 2011; Zimmer et al. 2000). It could be the case that children's free recall scores for abstract words in Experiment 1 were not sufficiently high enough to detect a gesture advantage. Free recall tasks tend to be more difficult than cued memory tasks for both primary school children (Karpicke et al. 2016) and adults (for review see Cleary 2018), in line with the particularly low scores for freely recalled abstract word translations in the current study.

Differences in Effects of Enrichment for Children and Adults

The finding that picture-enriched and gesture-enriched learning yielded similar performance outcomes departs from the pattern of performance observed in adults. Adults showed a greater gesture benefit than picture benefit 6 months after a week-long training period (Mayer et al. 2015), suggesting that picture-enriched L2 words decay more quickly from memory than gesture-enriched L2 words. The current findings therefore imply that teaching strategies derived from studies on adults may not directly translate into teaching strategies for children. Differences in enrichment effects have immediate implications for evidence-based teaching techniques, as gestures and other sensorimotor-based interventions may be more challenging for educators to integrate into pedagogy than picture-based interventions. Enrichment outcomes for 8-year-old children in the present study differed from adult outcomes. Specifically, gesture enrichment was not more effective than picture enrichment for 8-year-old children; in contrast, in adults gesture enrichment was more beneficial than picture enrichment, particularly over the long term (Mayer et al. 2015; Repetto et al. 2017). Differences in enrichment benefits between children and adults are likely not attributable to differences in gesture or picture stimuli across studies. This is because identical or similar enrichment materials have been used in experiment with adults (Macedonia and Knösche 2011; Macedonia et al. 2010; Mayer et al. 2015; Repetto et al. 2017). Differences between children and adults also cannot be attributed to perceptual stimulus characteristics, as L2 words in the current and prior studies were counterbalanced between enrichment conditions. Finally, differences between children and adults are unlikely to be due to differences in learning procedures, as the training procedure was highly similar across studies. Adults were, however, trained on a larger number of L2 words (about 90 words) than the children here but also received greater amounts of training (3 h per day for a total of about 15 h) and were as a result exposed more often to each L2 word (Macedonia and Knösche 2011; Macedonia et al. 2010; Mayer et al. 2015). Whether this difference in overall L2 word exposure can explain the difference in effectiveness of gesture versus picture enrichment for adults compared with children is an open question.

Potential Neural Mechanisms Underlying Enhanced Learning Outcomes

Several studies investigating learning enrichment in adults have also included neural measurements revealing that visual sensory and motor cortices are involved in translating sensorimotor and multisensory-enriched auditory L2 words (Macedonia et al. 2011; Macedonia and Mueller 2016; Mayer et al. 2015). Interfering with processes in these brain areas via neurostimulation can negate the beneficial effect of gesture-enriched learning on L2 translation performance (Mathias et al. 2019). These findings indicate that translating L2 vocabulary may trigger an internal dynamic generative model that reconstructs enrichment-related information, which in turn aids in L2 word recognition (for a review see von Kriegstein 2012; von Kriegstein and Giraud 2006; Mayer et al. 2015; Yildirim and Jacobs 2012). Studies including children are

needed to learn whether children's recognition of L2 words in the current study may have similarly involved visual and motor cortical responses to auditory cues during vocabulary translation.

Conclusion

We identified long-term benefits of gesture and picture enrichment on school children's learning of novel L2 vocabulary. Gesture and picture enrichment strategies were tested systematically using large sample sizes of children in an applied, naturalistic school environment. Gesture enrichment improved children's ability to translate novel vocabulary, compared with non-enriched learning. These benefits were present immediately following L2 training and persisted up to 2 to 6 months after training had ended. Gesture enrichment was, however, not more beneficial than picture enrichment in terms of L2 retention. We conclude that self-performed gestures and pictorial forms of enrichment may serve as equally beneficial strategies for the learning of foreign language vocabulary in primary school contexts. Conventional primary school teaching strategies such as listening to verbal material may therefore be improved by implementing pictures or gestures into vocabulary learning tasks.

Acknowledgments Open access funding provided by Projekt DEAL. We thank Julia Schwerin for assistance with planning and preparing the study, as well as serving as a teacher to the children in two experiments.

Author Contributions C.A. and K.v.K. conceived the study. C.A., A.S., and K.v.K. designed and performed the experiments. B.M. analyzed the data. M.M. contributed stimulus material. B.M., M.M., and K.v.K. wrote the manuscript. C.A. contributed to the editing of the manuscript.

Funding Information This work was funded by the German Research Foundation grant KR 3735/3-1, a Schulbezogene Forschung grant from the Saxony Zentrum für Lehrerbildung und Schulforschung (ZLS), and an Erasmus Mundus Postdoctoral Fellowship in Auditory Cognitive Neuroscience. B.M. is also supported by the European Research Council Consolidator Grant SENSOCOM 647051 to K.v.K.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40*(1), 23–31.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74–85.
Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*(5), 369–386.
Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.
Bavin, E. L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., ... & Ukoumunne, O. C. (2008). The early language in Victoria study: Predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language, 35*(3), 687–701.
Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin, 143*(3), 263–292.
Braxton, J. M., Jones, W. A., Hirschy, A. S., & Hartley III, H. V. (2008). The role of active learning in college student persistence. *New Directions for Teaching and Learning, 2008*(115), 71–83.
Breithecker, D., & Dordel, S. (2003). Bewegte Schule als Chance einer Förderung der Lern-und Leistungsfähigkeit. *Haltung und Bewegung, 23*(2), 5–15.
Cabrera, M. P., & Martínez, P. B. (2001). The effects of repetition, comprehension checks, and gestures, on primary school children in an EFL situation. *ELT Journal, 55*(3), 281–288.
Caramelli, N., Setti, A., & Maurizzi, D. D. (2004). Concrete and abstract concepts in school age children. *Psychology of Language and Communication, 8*(2), 19–34.
Caselli, M. C., Rinaldi, P., Stefanini, S., & Volterra, V. (2012). Early action and gesture "vocabulary" and its relation with word comprehension and production. *Child Development, 83*(2), 526–542.
Cleary, A. M. (2018). From free recall to recognition. In H. Otani & B. L. Schwartz (Eds.), *Handbook of research methods in human memory*. New York: Routledge.
Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22*(1), 267–281.
Collins, J. W., & O'Brien, N. P. (Eds.). (2003). *The Greenwood dictionary of education*. Westport, CT: Greenwood.
Cook, V. (2008). *Second language learning and language teaching*. London: Hodder Education.
Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language & Cognitive Processes, 27*(4), 594–610.
Daniels, M. (1997). Teacher enrichment of prekindergarten curriculum with sign language. *Journal of Research in Childhood Education, 12*(1), 27–33.
DeBaryshe, B. D. (1993). Joint picture-book reading correlates of early oral language skill. *Journal of Child Language, 20*(2), 455–461.
DeLoache, J. S., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G. L., ... & O'Doherty, K. (2010). Do babies learn from baby media?. *Psychological Science, 21*(11), 1570–1574.
Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language, 59*(3), 361–376.
de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E., & Vogt, P. (2018). The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 50-58). ACM.
Dodson, C. S., & Schacter, D. L. (2001). If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*(1), 155–161.
Ehret, C., & Hollett, T. (2014). Embodied composition in real virtualities: Adolescents' literacy practices and felt experiences moving with digital, mobile devices in school. *Research in the Teaching of English, 48*(4), 428–452.
Engelkamp, J., & Zimmer, H. D. (1984). Motor programme information as a separable memory unit. *Psychological Research, 46*(3), 283–299.
Engelkamp, J., & Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *German Journal of Psychology, 28*, 239–254.
Eriksson, M., & Berglund, E. (1999). Swedish early communicative development inventories: Words and gestures. *First Language, 19*(55), 55–90.
Foster, W., & Freeman, E. (2008). Poetry in general practice education: Perceptions of learners. *Family Practice, 25*(4), 294–303.
Frick, A., Daum, M. M., Wilson, M., & Wilkening, F. (2009). Effects of action on children's and adults' mental imagery. *Journal of Experimental Child Psychology, 104*(1), 34–51.
Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 360*(1456), 815–836.
Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 364*(1521), 1211–1221.

Funk, M., Brugger, P., & Wilkening, F. (2005). Motor processes in children's imagery: The case of mental rotation of hands. *Developmental Science, 8*(5), 402–408.
Gabbard, C. (2009). Studying action representation in children via motor imagery. *Brain and Cognition, 71*(3), 234–239.
Gilakjani, A. P. (2012). Visual, auditory, kinaesthetic learning styles and their impacts on English language teaching. *Journal of Studies in Education, 2*(1), 104–113.
Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge: The Belknap Press.
Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337–350.
Gullberg, M., De Bot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture, 8*(2), 149–179.
Harpaintner, M., Sim, E. J., Trumpp, N. M., Ulrich, M., & Kiefer, M. (2020). The grounding of abstract concepts in the motor and visual system: An fMRI study. *Cortex, 124*, 1–22.
Harpaintner, M., Trumpp, N. M., & Kiefer, M. (2018). The semantic content of abstract concepts: A property listing study of 296 abstract words. *Frontiers in Psychology, 9*, 1748.
Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., & Nesbitt, K. T. (2016). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly, 51*(2), 181–198.
Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27*(1), 4–9.
Herodotou, C. (2018). Young children and tablets: A systematic review of effects on learning and development. *Journal of Computer Assisted Learning, 34*(1), 1–9.
Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). Codes and their vicissitudes. *Behavioral and Brain Sciences, 24*(5), 910–926.
Howe, M. L., & Brainerd, C. J. (1989). Development of children's long-term retention. *Developmental Review, 9*(4), 301–340.
Hsu, H. M. J. (2011). The potential of Kinect in education. *International Journal of Information and Education Technology, 1*(5), 365.
Icht, M., & Mama, Y. (2015). The production effect in memory: A prominent mnemonic in children. *Journal of Child Language, 42*(5), 1102–1124.
Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367–371.
Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33*(11), 1419–1432.
Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology, 7*, 350.
Kiebel, S. J., Garrido, M. I., Moran, R. J., & Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. *Cognitive Neurodynamics, 2*(2), 121–136.
Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., & Sachse, S. (2015). Handwriting or typewriting? The influence of pen-or keyboard-based writing training on reading and writing performance in preschool children. *Advances in Cognitive Psychology, 11*(4), 136–146.
Kiefer, M., & Trumpp, N. M. (2012). Embodiment theory and education: The foundations of cognition in perception and action. *Trends in Neuroscience and Education, 1*(1), 15–20.
Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719.
Konomi, D. K. (2014). Using visual materials in teaching vocabulary in English as a foreign language classrooms with young learners. In *Proceedings of the international conference on new perspectives in science education* (pp. 256–260). Padova: Webster.
Köper, M., & Schulte Im Walde, S. (2016). Automatically generated affective norms of abstractness, arousal, imageability and valence for 350,000 German lemmas. In *LREC*.
Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford: Oxford University Press.
Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174.
Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713–727.
Lewis, L. H., & Williams, C. J. (1994). Experiential learning: Past and present. *New Directions for Adult and Continuing Education, 1994*(62), 5–16.
Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development, 75*(5), 1357–1372.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General, 141*(1), 170–186.
Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. *Frontiers in Psychology, 5*, 1467.
Macedonia, M. (2015). Learning styles and vocabulary acquisition in second language: How the brain learns. *Frontiers in Psychology, 6*, 1800.
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*(5), 162–169.
Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education, 5*(4), 196–211.
Macedonia, M., Müller, K., & Friederici, A. D. (2010). Neural correlates of high performance in foreign language vocabulary learning. *Mind, Brain, and Education, 4*(3), 125–134.
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*(6), 982–998.
Macedonia, M., & Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Frontiers in Psychology, 7*, 953.
MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 671–685.
Mahmoudi, S., Jafari, E., Nasrabadi, H. A., & Liaghatdar, M. J. (2012). Holistic education: An approach for 21 century. *International Education Studies, 5*(2), 178–186.
Mathias, B., Sureth, L., Hartwigsen, G., Macedonia, M., Mayer, K. M., & von Kriegstein, K. (2019). A causal role of sensory cortices in behavioral benefits of 'learning by doing'. *arXiv*, 1903.04201.
Mavilidi, M. F., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review, 27*(3), 413–426.
Mayer, K. M., Macedonia, M., & von Kriegstein, K. (2017). Recently learned foreign abstract and concrete nouns are represented in distinct cortical networks similar to the native language. *Human Brain Mapping, 38*(9), 4398–4412.
Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology, 25*(4), 530–535.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.
McFalls, E. L., Schwanenflugel, P. J., & Stahl, S. A. (1996). Influence of word meaning on the acquisition of a reading vocabulary in second-grade children. *Reading and Writing, 8*(3), 235–250.
McGill-Franzen, A., Zmach, C., Solic, K., & Zeig, J. L. (2006). The confluence of two policy mandates: Core reading programs and third-grade retention in Florida. *The Elementary School Journal, 107*(1), 67–91.
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought* (11th ed.). Chicago: University of Chicago Press.
Mecklenbräuker, S., Steffens, M. C., Jelenec, P., & Goergens, N. K. (2011). Interactive context integration in children? Evidence from an action memory study. *Journal of Experimental Child Psychology, 108*(4), 747–761.
Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education, 30*(4), 159–167.
Millington, N. T. (2011). Using songs effectively to teach English to young learners. *Language Education in Asia, 2*(1), 134–141.
Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science, 26*(9), 1489–1496.
Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*, 255–287.
Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology, 80*(2p1), 279.
Porter, A. (2016). A helping hand with language learning: Teaching French vocabulary with gesture. *The Language Learning Journal, 44*(2), 236–256.
Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science, 21*(4), e12593.
Repetto, C., Pedroli, E., & Macedonia, M. (2017). Enrichment effects of gestures and pictures on abstract words in a second language. *Frontiers in Psychology, 8*, 2136.
Richardson, J. (2003). Dual coding versus relational processing in memory for concrete and abstract words. *European Journal of Cognitive Psychology, 15*(4), 481–509.

Sadoski, M. (2018). Reading comprehension is embodied: Theoretical and practical considerations. *Educational Psychology Review, 30*(2), 331–349.
Sadoski, M., & Paivio, A. (2013). *Imagery and text: A dual coding theory of reading and writing*. New York: Routledge.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.
Saltz, E., & Dixon, D. (1982). Let's pretend: The role of motoric imagery in memory for sentences and words. *Journal of Experimental Child Psychology, 34*(1), 77–92.
Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 223–250). Hillsdale: LEA.
Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology, 101*(2), 305.
Scott, J. A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal, 103*(3), 269–286.
Sénéchal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28*(4), 360–374.
Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417.
Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception & Psychophysics, 66*(2), 352–362.
Snelson, C. (2011). YouTube across the disciplines: A review of the literature. *MERLOT Journal of Online Learning and Teaching, 7*(1), 159–169.
Tan, E., & Pearce, N. (2011). Open education videos in the classroom: Exploring the opportunities and barriers to the use of YouTube in teaching introductory sociology. *Research in Learning Technology, 19*, 1–9.
Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture, 8*(2), 219–235.
Tschirner, E. (2001). Language acquisition in the classroom: The role of digital video. *Computer Assisted Language Learning, 14*(3-4), 305–319.
Volk, M., Cotič, M., Zajc, M., & Starcic, A. I. (2017). Tablet-based cross-curricular maths vs. traditional maths classroom practice for higher-order learning outcomes. *Computers & Education, 114*, 1–23.
von Kriegstein, K. (2012). A multisensory perspective on human auditory communication. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 683–700). Boca Raton: Taylor & Francis.
von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*(10), e326.
von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A. L., Kell, C. A., Grüter, T., ... & Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences, 105*(18), 6747–6752.
v. Soden-Fraunhofen, R., Sim, E.-J., Liebich, S., Frank, K., & Kiefer, M. (2008). Die Rolle der motorischen Interaktion beim Erwerb begrifflichen Wissens: Eine Trainingsstudie mit künstlichen Objekten. *Zeitschrift für Pädagogische Psychologie, 22*, 47–58.
Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*(2), 265–279.
Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory, 20*(6), 568–579.
Yildirim, I., & Jacobs, R. A. (2012). A rational analysis of the acquisition of multisensory representations. *Cognitive Science, 36*(2), 305–332.
Zimmer, H. D., Helstrup, T., & Engelkamp, J. (2000). Pop-out into memory: A retrieval mechanism that is enhanced with the recall of subject-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 658.

Affiliations

Christian Andrä[1,2] · Brian Mathias[3,4] · Anika Schwager[5] · Manuela Macedonia[4,6] · Katharina von Kriegstein[3,4]

[1] Department of Teacher Education and School Research, University of Leipzig, Leipzig, Germany

[2] Department of School Sport, Institute of Sports Psychology and Physical Education, Faculty of Sports Science, University of Leipzig, Leipzig, Germany

[3] Chair of Cognitive and Clinical Neuroscience, Faculty of Psychology, Technical University Dresden, Dresden, Germany

[4] Research Group Neural Mechanisms of Human Communication, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[5] Didactics of Physical Education, Institute of Primary and Pre-Primary Education, Faculty of Education, University of Leipzig, Leipzig, Germany

[6] Linz Center of Mechatronics, Johannes Kepler University Linz, Linz, Austria