

# Real body versus 3D avatar: the effects of different embodied learning types on EFL listening comprehension

Yu-Ju Lan<sup>1</sup> · Wei-Chieh Fang<sup>2</sup> · Indy Y. T. Hsiao<sup>1</sup> · Nian-Shing Chen<sup>3</sup>

Published online: 18 April 2018

© Association for Educational Communications and Technology 2018

**Abstract** The current study aimed at investigating how different types of embodied learnings influence elementary school students' English as a foreign language (EFL) listening performance. Two kinds of embodied learnings: real and physical body versus the 3D avatar, were compared with non-embodied learning. 69 fifth graders from two elementary schools participated in this study, and were randomly assigned into three groups (Kinect, Second Life, and paper). They learned the identical English phrases of doing sports by involving different types of embodied learnings. During the 11-week experiment, an identical EFL performance test was administered six times: before (once), during (3 times), and after (twice) the learning activities. The results depict that students learned better by watching their own 3D avatars doing motions than by moving their own bodies to produce the motions or doing nothing. Further analysis showed that the improvements made by those in the Second Life group were greater than those made by the participants in the other two groups when the performances of students with low achievement were compared. It was also found that learning by watching one self's avatar benefits both

---

✉ Nian-Shing Chen  
nianshing@gmail.com

Yu-Ju Lan  
yujulan@gmail.com

Wei-Chieh Fang  
wfjohnny@gmail.com

Indy Y. T. Hsiao  
indymajere@gmail.com

<sup>1</sup> Department of Chinese as a Second Language, National Taiwan Normal University, 162, Sec. 1, Ho-Ping E. Rd, Taipei, Taiwan

<sup>2</sup> Department of Education, Washington University in St Louis, Campus Box 1183, St Louis, MO, USA

<sup>3</sup> Department of Applied Foreign Languages, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou City, Yunlin 64002, Taiwan

students with high- and low-achievement in EFL performance. Some suggestions for pedagogical applications and future research are also provided.

**Keywords** Embodied cognition · Gesture-based learning · Virtual worlds · Avatar · English as a foreign language (EFL)

## Introduction

Mental processing is usually the focus when discussing how learning happens and takes effect. Based on the cognitive perspective, knowledge acquisition happens when internal coding and structuring by the learner are actively involved (Derry 1996). It is believed that the deeper in mental processing it is, the more elaborate, longer lasting, and stronger the memory traces in learning will be gained based on the cognitive perspective ( Craik and Lockhart 1972). Consequently, the efficiency of the input information processed in and retrieved from a learner's mind has been the main concern in the relevant research on technology-assisted second language teaching/learning from a cognitive perspective. With the increasing awareness of how a person learns a second language, it is known that a human's mind and body learn together and they mutually influence each other (Macedonia and Knösche 2011; Glenberg et al. 2011). According to the theories of embodied cognition, all the aspects of cognition are shaped by the aspects of the body (Goldin-Meadow and Wagner 2005; Goldin-Meadow and Beilock 2010). The way a person's body and his or her surroundings interact actually influences the abilities of the mind, just as the mind influences the body's movements (Barsalou 2008; Cowart 2005; Chen and Fang 2014; Gibbs 2005; Wilson 2002). Moreover, the perspective of embodied language processing also indicates that moving one's body in a certain way while learning will impact how a person comprehends certain concepts (Rueschemeyer et al. 2010). Take the gesture-based approach, commonly used in language teaching as an example, when the learners of Chinese as a foreign language (CFL) embody the tones of Chinese characters, their comprehension and memory will be enhanced (Morett and Chang 2015; Tsai 2011). The well-known teaching approach, total physical response (TPR) (Asher 1969; Asher and Price 1967), widely adopted in English as a foreign language (EFL) classes, is another application example of embodied cognition in EFL teaching. In short, the mind and the body will be mutually dependent on and influential to each other.

It is interesting that a human's motor system is not only activated by actually moving one's own body, but is also activated by one's observation alone. According to Wilson's (2002) argument, embodied cognition has two versions, weak and strong. Weak embodiment supports that semantic representations do not solely rely on sensory processes, but go beyond the concrete level, while strong embodiment claims that semantic representations entirely depend on sensory and motor information. The current study adopted the weak version of embodiment without taking the metaphor representation into account. Furthermore, Mahon and Caramazza (2008) argued that humans' motor systems can be activated under three conditions: when they (1) observe manipulable objects, (2) process action verbs, and (3) observe another individual's movements. The results of Zwaan and his colleagues' (Zwaan et al. 2002) research not only support but also expand Mahon and Caramazza's (2008) argument. Thus, the arguments of Mahon and Caramazza (2008) can be expanded as "watching 2D pictures of manipulable objects also triggers a human's motor system." They found that by watching the pictures of people doing actions match the sentences shown to the participants also makes the sentences more comprehensible

(Zwaan et al. 2002). This conclusion then raises other questions: What will happen if a person watches her/his 3D avatar doing some motions? Will it yield the similar effects as it does if s/he does the motions by her/his own physical body?

To this end, this study aims at investigating the learning effects of different types of embodied movements, real body versus 3D avatar, on elementary school EFL students' learning comprehension of phrases about sports.

## Literature review

### Embodied cognition and language learning

According to the perspective of cognitive linguistics, meanings are not explored simply as grasped directly from the world but as conceptualized out of the way human bodies configure the reality (Holme 2012), which echoes the argument of embodied cognition. Furthermore, embodied cognition takes into account the interactions among perception, the body, and the environment. Cowart (2005) argued that all aspects of cognition are shaped by aspects of the body and are arisen via the interaction between a person's body and the surroundings which involve bodily sensation, perception, and action. In particular, an individual's experiences of visual, verbal, tactile, and kinesthetic aspects are deeply integrated within the system of knowledge presentation (Lan et al. 2015b; Hung et al. 2014). To put it simply, the motor system influences our cognition (Borghi and Cimatti 2010), and action and language are mutually dependent (Rueschemeyer et al. 2010).

In language learning, learners can use and retrieve perceptual and bodily details as a part of their conceptual representations of words, phrases, or sentences (Aziz-Zadeh and Damasio 2008; Willems and Casasanto 2011). Several theoretical explanations have been proposed to account for the rationale. Motor trace theory, for example, suggests that performing an action while learning a word creates a motor trace in memory accompanying the word. This motor trace can lead to the verbal information to be subsequently accessed faster, more accurately and decays more slowly as opposed to verbal information only heard or read (Macedonia and Knösche 2011). There is plenty of experimental evidence that supports embodied cognition in language teaching/learning (Atkinson 2010; Hung et al. 2015). For example, Rueschemeyer and her colleagues (Rueschemeyer et al. 2010) confirmed that action execution can affect lexical-semantic processing. They investigated the differences in processing words that denote function manipulable (FM) (e.g., cup and hammer) and function nonmanipulable (NM) (e.g., bookend and clock) objects by thirty-two right-handed native speakers of Dutch. They found that the participants performed better when they processed FM words by doing intentional actions than when processing NM words by doing arbitrary actions. Another support comes from study by Hung et al. (2014), which found that enhancing the connection between the body movements and the target language led to better learning retention as opposed to the conventional TPR approach. It was found that learners remembered more target vocabulary after being guided by a motion-sensing technology individually. Similar evidence obtained from neuroscience research also approves that listening to action-related sentences modulates the activity of the motor system in the brain (Buccino et al. 2005). Chinese character writing is another example of embodied cognition in language learning. Yang (2010) found that teaching Chinese as a second language (CSL) learners to write Chinese characters help them develop Chinese character perception.

In addition to the aforementioned experimental evidence obtained from the research on language learning, the findings of gestures in second language (L2) learning (e.g., vocabulary acquisition and reading comprehension) also echo that greater learning gains can be achieved while making or watching gestures, which is involved in a L2 learning process (Atkinson 2010; Chao et al. 2013; McCafferty 2002; Stevanoni and Salmon 2005; Tellier 2008). To account for the benefits of viewing gestures, Krauss' study (1998) about Lexical Gesture Process Model highlighted the function of gestures on activating lexical items and making them easier to be accessed. In the Indexical Hypothesis, Glenberg and Kaschak (2002) also argued that prior embodied experiences or representations can be reactivated when an object or action is perceived. With the activation of the embodied experiences, the abstract symbols of a language such as words can be mapped onto them (Glenberg et al. 2011). So that language can become meaningful to learners. For example, Morett and Chang (2015) found that viewing pitch gestures enhanced English speakers' ability to discriminate the meanings of Mandarin words in different tones. Additionally, the findings obtained from Gao's (2009) study proved that adopting gestures benefits Chinese learners' acquisition of lexical tones of Mandarin Chinese. Additionally, Chao et al. (2013) found that gesture-based learning benefited the learning efficiency of EFL. The arguments made by Chao et al. (2013) echo the evidence obtained from Tellier's (2008) research which stated that young children learned EFL words better if they did the corresponding actions with the pictures they were looking at. Furthermore, according to McNeill (1992), a language may even originate ontogenetically and phylogenetically in gestures because gestures not only support meaning making but also carry clues to the process of conceptualization from which meanings evolve (Negueruela and Lantolf 2004). Holme (2012) further argued that embodied approach makes language more memorable by reinvesting a new form in the movements, gestures, and physical imagery from which meaning was conceptualized. As argued by Asher (1977) that physical activity allows learners to construct the meanings of L2 words without translation, even this has never been proven beyond a doubt. In summary, learning with body involvement in general is more effective than learning by focusing on brain without body involvement. Learners' body movements or gestures actually can be viewed as an approach to helping strengthen the connection between learners' linguistic and motor systems, and therefore enhance learners' language learning. Additionally, the embodied approach can be applied to the learning of multiple linguistic skills, such as listening (e.g., the tones of Chinese characters), L2 vocabulary learning, and sentence meaning making and retrieving. The findings obtained from the studies described above show that the involvement of embodied movement by the learners enhanced the measured target linguistic abilities (e.g., Holme 2012; Morett and Chang 2015; Negueruela and Lantolf 2004; Tellier 2008).

However, among the works on embodied cognition, many of them focused on the effects of embodied actions on learning action-related words (e.g., Tellier 2008) or sentences (e.g., Chao et al. 2013), and the differences between learners physically doing the corresponding actions or not. Few investigated the potential effects of 3D avatar-based actions on L2 learning. In fact, to our best knowledge, the current study is among the few studies that examined the differences in L2 learning between two different embodied approaches by either viewing 3D avatars doing actions or using one's own physical body to do the actions. When learning an L2 in virtual worlds, learners experienced a sense of "presence," an objective sense of being in the virtual place (Regenbrecht et al. 1998). The cognitive representation in virtual worlds thus captures the relation between learners' avatars and the objects in the virtual environments, forming the meaning of the situation (Schubert et al. 1999). Will the avatar-based actions also benefit L2 learners' second

language acquisition (SLA)? It is without doubt an interesting research issue and worthy of more researchers' attention and efforts as empirical findings on embodied learning provide implications for the curriculum design in long distance education as well as in special education for learners with physical challenge.

### Language learning in 3D virtual worlds

Innovative technologies such as gesture recognition and virtual reality enable learners to engage in immersed and virtual activities with either physical or virtual movements. Both technologies have the potential for brining learners to deeper levels of cognition as they interact with virtual objects, which meets the evolving expectations of learners (Johnson et al. 2016; Becker et al. 2017). In learning a second language (L2), getting immersed in an authentic context is particularly important (Lan 2014; Krashen 1982). Since language is arbitrary and consists of abstract symbols, learners at the initial stage of language acquisition cannot acquire it merely via context-reduced practicing by rote learning. L2 learning which emphasizes that learners using the target language in an authentically immersive environment benefits L2 learners' oral performance and forms accuracy (Deutschmann et al. 2009; Lan 2014; Lan et al. 2013). The evidence obtained from neuroscience research also supports context-immersive learning for L2 acquisition (Linck et al. 2009; Zinszer and Li 2012).

Remarkable advances have been made in 3D virtual technology, allowing learning L2 in an authentically immersive environment in recent years to be much easier than learning L2 by having to create an environment in conventional classrooms (Lan 2014). A multi-user virtual environment (MUVE), such as second life (SL), provides L2 learners with an authentically immersive context via integrating virtual reality with network (Lan et al. 2016) in which L2 learners use their avatars to interact with both others' avatars and the objects in the environment without any spatial or temporal barriers (Lan et al. 2013). Obviously, the benefits of learning in 3D virtual worlds to learners' L2 acquisition have gained increasing attention from L2 researchers and educators (Lan 2015; Lin and Lan 2015). Consistent with the perspective of sociocultural SLA, the learning effects in virtual worlds involve immersively avatar-based role-play and socially authentic interaction. Based on the existing limited volume of literature on virtual worlds used in language learning, some recent research and application trends could be identified (Lin and Lan 2015): (1) the obtained evidence gradually shifts from self-reported-based perception to practical experimental evidence (Lan 2014; Lan et al. 2016), (2) the application of virtual worlds in education moves by degrees from blending in formal learning to non-formal learning (Lan et al. 2013; Lan 2014, 2015b; Lan et al. 2015b; Ryu 2013), and (3) the involving activities progressively transfer from teacher-guided or loose-structured learning to student-centered and more-structured task-based learning (Deutschmann et al. 2009; Lan 2014, 2016).

Among the literature trends mentioned above, the amount of research focusing on comparing the effects in different types of embodied movement is insufficient. Although the study of Lan et al. (2015b) investigated the embodied effects on CFL learners' vocabulary learning in a 3D environment, physical movements were not investigated in their study. Based on the results obtained from Lan et al. (2015b) research mentioned above, CFL learners learned better in a 3D virtual environment than those learning in a 2D web-based environment. Will similar or different results be obtained if 3D avatar movements versus real physical movements are involved in learning identical foreign language (FL) materials?

Without a doubt, it is an interesting topic and worthy of researchers' attentions. If practical evidence on embodied learnings supports that learning by watching learners' 3D avatars' movements also benefits L2 learners' achievement, not only the knowledge of embodied cognition can be expanded, its pedagogical application on L2 teaching and learning can be also more flexible than only focusing on humans' physical body movements. In sum, to deal with the abovementioned issues and to reach the study aim stated in the end of the Introduction section, two research questions will be addressed:

- (1) What are the differences in the effects of different types of embodied movements (real body vs. 3D avatar) on elementary school EFL students' learning of phrases about sports?
- (2) What are the differences in the effects of different types of embodied movements (real body vs. 3D avatar) on the learning of phrases about sports by elementary school students of different EFL levels?

## Methodology

### Participants

The participants were 69 fifth graders from two elementary schools in Taipei City. All of them were EFL beginners, as according to the curricular and instructional reforms announced by the Taipei City Government's Department of Education (2000), elementary school students need to acquire 320 words and 99 basic sentences of basic daily conversation (e.g., Do you like apples? Yes, I do/No, I don't.) and classroom English upon graduation (e.g., Take out your book/Put away your book.). A quasi-experimental design was adopted in which participants were randomly assigned into three treatment groups: Kinect ( $N = 25$ ), Second Life (SL) ( $N = 22$ ), and paper ( $N = 22$ ). After the pretest which will be described below in the Instruments section, the participants in each treatment group were almost equally sorted and divided into three levels according to their scores at the pretest that represented their prior knowledge of the L2 phrases which were the learning materials during the experiment (high: above 11 points; medium: 9–11 points; and low: below 9 points). As a result, the Kinect group was composed of 9 high-, 7 medium-, and 9 low-achievement students in terms of their prior knowledge of the intended learned materials; both the SL and paper groups consisted 8 high-, 6 medium-, and 8 low-achievement students, respectively.

### Research design

A quasi-experimental research design was adopted in this study. All the students learned identical English phrases of sports within identical learning periods; only the learning approaches were different from group to group as Table 1 shows. While learning the English phrases, the students in the Kinect group used their physical bodies to act out the corresponding motions, and those in the SL group used mouse to control their 3D avatars to act out the motions; those in the paper group did nothing with their physical bodies but looked at the 2D line drawings on the paper. Additionally, all groups received identical audios and 2D visuals accompanying the English phrases.

**Table 1** Learning approaches by the three treatment groups

	Paper group	Second life group	Kinect group
Visual learning	View 2D line drawings	Watch 3D self-avatar acting out phrases after viewing the 2D line drawings	View 2D line drawings
Audio learning	Listen to audios	Listen to audios	Listen to audios
Kinesthetic learning	None	None	Act out phrases after viewing the 2D drawings

## Instruments

### *Learning materials*

The learning materials of this study were eighteen 4-syllable “do a \_\_\_” English phrases about six types of sports (soccer, Kungfu, Tae Kwon Do, swimming, gymnastics, and volleyball). All the 18 phrases were chosen because they could be easily acted out by the kids. Additionally, all the chosen phrases were not included in the regular EFL syllabus of Taiwanese elementary schools; therefore, all the participants would not learn the materials used in this study in their regular EFL classes. The following are some examples of the motions, each based on one sport.

- Do a header (soccer)
- Do a lunge step (kungfu)
- Do a front kick (Tae Kwon Do)
- Do a backstroke (swimming)
- Do a deep squat (gymnastics)
- Do a bump pass (volleyball)

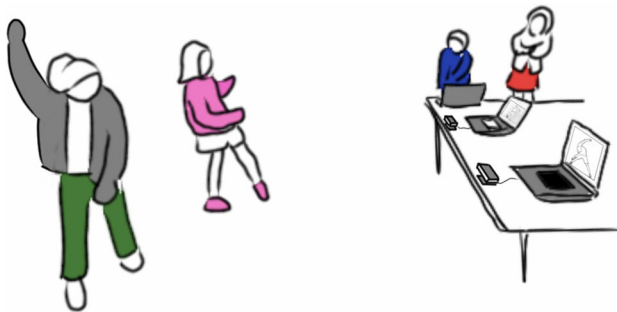
In addition to the audio records of each phrase, 2D line drawings of each motion were also shown to the students. For the Kinect and SL groups, a video composed of a successive presence of these 2D line drawings with audio were used as the learning materials. The learners in both groups can see the line drawings while listening to the phrases. Therefore, they could imitate the motions either via their own physical bodies or 3D avatars. On the other hand, the identical 2D line drawings which were printed on a mini book like a comic book were used by the paper group. Students in the paper group watched the motion book and did not use their own bodies to do any motions shown on the mini books while listening to the audio version of the phrases. Figure 1 shows the 2D line drawing of the motion of “do a header.”

### *Kinect*

Kinect is a low-cost consumer device applied to the Xbox 360 and was developed by Microsoft in 2010. Kinect can track the movements of 25 distinct skeleton joints on a human body and track as many as six complete skeletons at the same time and it has voice recognition capabilities (<https://dev.windows.com/en-us/kinect/hardware>). By using Kinect, users can interact with the Xbox 360 interface via gestures instead of any handheld devices or stampede. Figure 2 is the schematic picture showing how students learned the



**Fig. 1** The line drawing of the motion of “do a header”



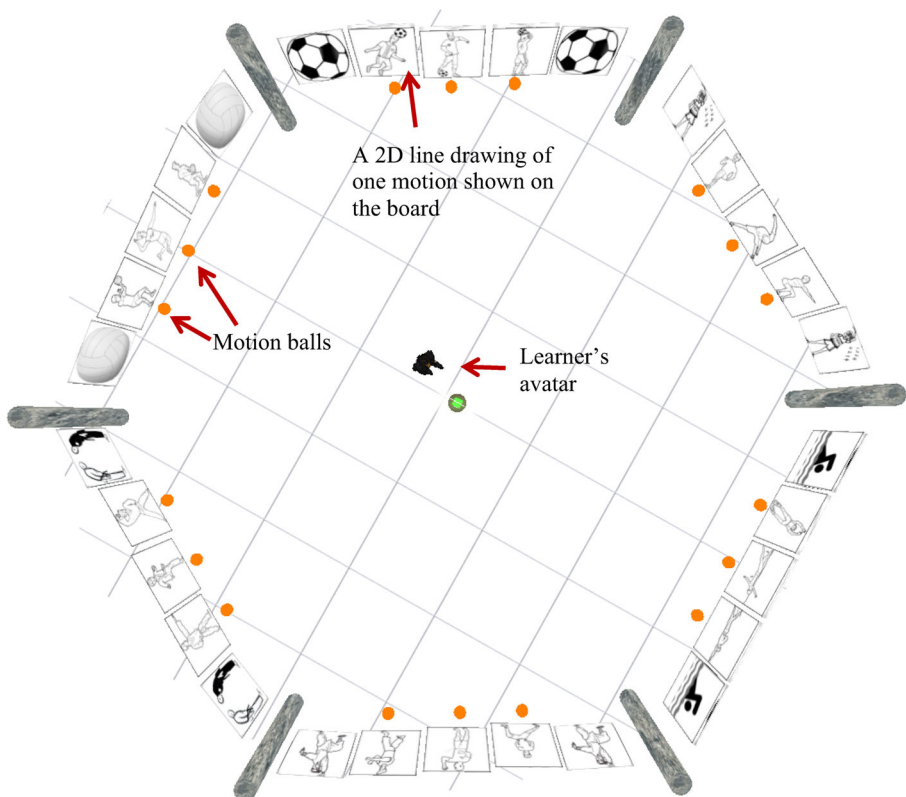
**Fig. 2** The learning circumstances of the Kinect group

phrases in the Kinect group. As shown in Fig. 2, there was a laptop in front of each user. The 2D line drawings of the motions were shown on the screen of the laptop to guide the users in acting out the target motion.

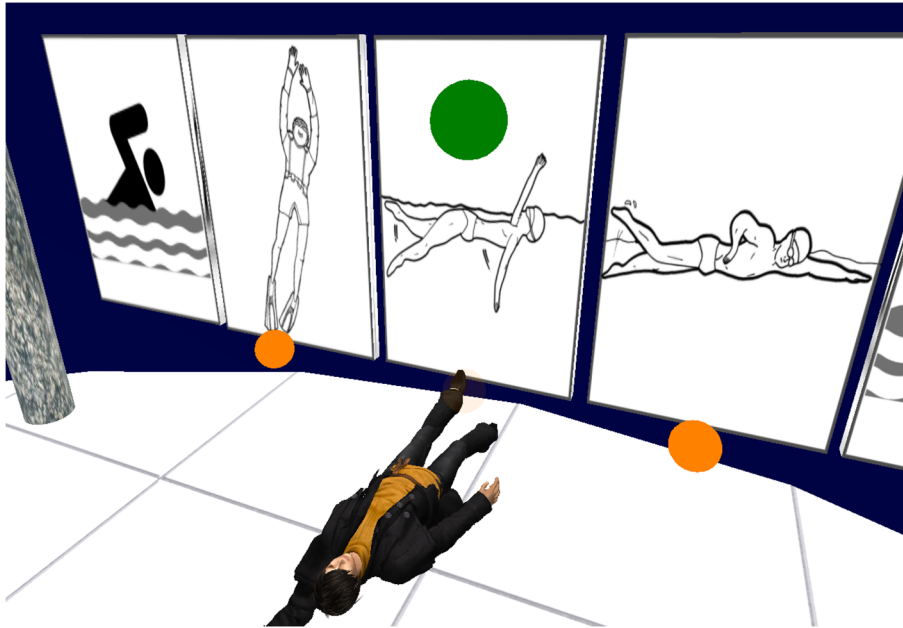


### *The virtual gym and the motions*

The 3D virtual environment used in this study is Second Life. A virtual round gym was developed with 18 boards on which were the identical 2D line drawings of the motions from different types of sports used in the other two groups. In front of the board was a motion ball which allowed a user's avatar to do the same motion as that shown on the board once the ball was clicked. Each motion ball had different Second Life animations and LSL (Linden Script Language) scripts which would set the avatars and the angles of the SL camera to optimize the presentation of the motions. Additionally, to create all the avatar motions mentioned above, we used Kinect as a motion capture tool in this study. First, we captured the movements of six sports items with the Brekel Kinect application and saved the motions as BVH motion files (Yoon and Park 2013). Next, we modified the files with the BVHacker application before uploading them to Second Life as Second Life animations. Figure 3 shows the virtual gym in which the learners' avatars could select any motion balls and make their avatars perform the identical motions shown on the boards. Additionally, Fig. 4 shows the motion of "do a backstroke" done by a learner's avatar.



**Fig. 3** The virtual gym



**Fig. 4** The motion of “do a backstroke” done by a learner’s avatar

### *EFL performance test*

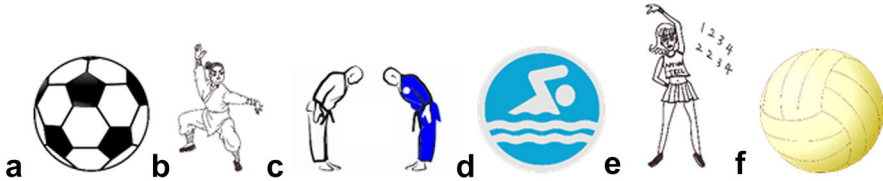
The performance test includes three categories of test items: (1) listen and recognize the corresponding type of sport; (2) listen and recognize the corresponding motion; and (3) say out the name of the motions. Each category consists of 18 items embedded in 18 phrases, i.e., all the 18 phrases randomly appear once in each category. In terms of scoring, for the first two categories (multiple choice), each correct response worth 1 point, while each incorrect one scores 0. For the third category, because each sentence contains four syllables (3 or 4 words). The first two syllables (i.e., “do a”) together worth 1 point, while the latter two syllables or two words (e.g., “header” or “front kick”) worth 1.5 each. Figure 5 shows examples from each category of the test items.

### *Procedure*

The study lasted 11 weeks in the fourth season of 2014. Before the treatment, all the participants were individually and randomly assigned into different learning groups. In the first week, all the participants performed the EFL performance test as the pretest. Next, from the 2nd week, the participants learned the identical materials once a week, lasting 3 weeks (weeks 2–4). While learning, all the participants in both the Kinect and SL groups watched the 2D line drawing video of the 18 motions shown on the laptop screens and computer screens, respectively. Meanwhile, the participants in the Kinect group acted out the motions (as shown in Fig. 2); while those in the SL group, controlled their avatars and watched the avatars do the motions shown on the screen. In contrast, the participants in the paper group sat on the chairs, watching the comic books and listening to the audio without doing any embodied motions. All the three groups received identical audios and visuals.

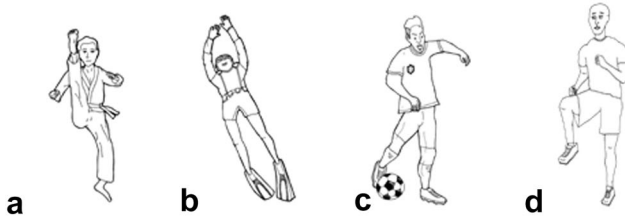
I. Listen and recognize the sport: Please listen carefully and choose the sport to which the motion belongs.

Audio: (Do a header)



II. Listen and recognize the motion: Please listen carefully and choose the correct motion.

Audio: (Do a front kick)



III. Look at the picture and say out the motion in English.



Fig. 5 Examples of test items from each category

After each learning treatment, the identical EFL performance test was administered to receive their instant learning outcome. After the three learning treatments, in order to investigate how long the learning effects would retain, two delay tests of the identical performance test were administered in weeks 6 and 11, respectively.

## Results

To evaluate how the different types of embodied learnings affected EFL students' English performance, a total of 6 test scores were collected and analyzed during the experiment. In addition to the scores of the items of multiple choices (1 point for each correct answer and 0 for each wrong one), the scores of the oral items were scored by three raters. The Pearson product-moment correlations of scores from the three raters were all above 0.99. They were 0.991, 0.994, and 0.995 respectively.

In addition to comparing all participants' learning outcomes among the three learning groups (Kinect, SL, and paper), the learning gains of the students with different EFL achievement levels (high- versus low-achievement) were further compared to understand how different embodied learning types benefited students with different levels. The analysis results are elaborated below.

### Comparison of learning gains on EFL performance of all the participants

The overall learning gains of all the participants of the three groups were first analyzed. Table 2 lists the descriptive statistics of the scores of all the participants in the three groups at the six tests. First of all, an  $F$ -test was conducted to test for differences in pretest among the three groups. The results revealed that there was no significant difference among the students receiving different treatments [ $F(2,66) = 0.8, p = 0.923$ ]. Additionally, by performing the Bonferroni test, two-way (test  $\times$  group) repeated measure ANOVA revealed that the interaction between group and test was non-significant [ $F(10,330) = 1.17, p = 0.308$ ]. Thus, there is no interaction among the variables, test scores and groups.

The main-effect analysis was then used to compare the differences among the test results of the six tests done by three groups (Kinect, SL, and paper). With respect to the EFL performance test scores, the results reveal that there were no statistically significant differences between subject variable (groups) [ $F(2,66) = 0.35, p = 0.71$ ], but significant differences were found within subject variable (test interval) [ $F(5, 330) = 18.66, p < 0.001, \text{partial } \eta^2 = 0.22, \text{power} = 0.92$ ]. The analysis results described above indicate that although the three groups did not perform significantly differently, they made significant improvements during the experiment. A repeated measure performing the Bonferroni test for each group was therefore conducted to identify the improvements made by each group among the 6 tests. The analysis results reveal that significant differences among the 6 test scores exist in both the Kinect [ $F(5, 120) = 8.10, p < 0.001, \text{partial}$

**Table 2** The descriptive statistics of the scores of all the participants in the three groups

Test	Kinect ( $N = 25$ )		SL ( $N = 22$ )		Paper ( $N = 22$ )	
	Mean	SD	Mean	SD	Mean	SD
Pretest (week 1)	11.08	4.36	10.57	4.47	10.78	4.36
Instant test 1 (week 2)	18.02	2.73	15.21	2.91	16.43	2.91
Instant test 2 (week 3)	20.70	3.10	18.84	3.31	18.64	3.31
Instant test 3 (week 4)	23.70	3.91	24.45	4.17	19.09	4.17
Delay test 1 (week 6)	23.22	3.75	27.73	4.00	19.41	4.00
Delay test 2 (week 11)	22.56	3.86	24.36	4.12	18.09	4.12

$\eta^2 = 0.25$ , power = 0.99] and the SL groups [ $F(5, 105) = 13.52$ ,  $p < 0.001$ , partial  $\eta^2 = 0.39$ , power = 0.99], but not in the paper group. Therefore, a post hoc analysis was also conducted to identify the differences and the results for the Kinect and the SL groups are listed in Table 3, respectively. Based on the data listed in Table 3, it can be found that the significant improvements exist in all the tests compared with the pretest as well as the three tests (instant test 3 and both delay tests compared with the instant test 1).

In contrast, more significant improvements made by the SL group compared with the Kinect group were found. By comparing the results of the pretest, instant tests 1 or 2, all the scores of the latter ones are significantly higher than those of the former ones. The test trend is shown in Fig. 6, conveying the same results as described above.

According to the results of post hoc analysis, the improvements made by the students learning by watching their virtual avatars doing the motions while listening to the English phrases of the six sports clearly benefited the most. The next group which is also benefited is learning by using their own bodies to do the motions. Obviously, those learning by watching the paper books without doing any motions did not make any significant improvements after the treatments.

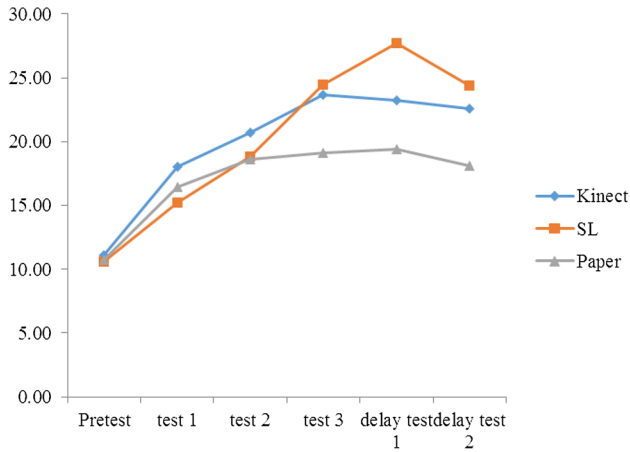
### Comparison of learning gains on EFL performance between the students with high- and low-achievement

To further investigate how students with different EFL levels were benefited by learning with different embodied motions, the participants in the three groups were further grouped into different EFL achievement levels based on their scores at pretest as described in the Participants section. Additionally, only the scores of EFL performance tests of the students with both high- and low-achievement in each group (Kinect, SL, and paper) were analyzed here. Tables 4 and 5 list both high- and low-achievement students' scores at the EFL performance tests, respectively.

**Table 3** The results of post hoc analysis of the six test scores of the Kinect and SL group

Tests (Sig.)					
	Instant test 1	Instant test 2	Instant test 3	Delay test 1	Delay test 2
Kinect group					
Pretest	0.005**	0.001**	0.002**	0.001**	0.003**
Instant test 1		0.070	0.004**	0.019*	0.046*
Instant test 2			0.161	0.205	0.428
Instant test 3				0.715	0.573
Delay test 1					0.777
SL group					
Pretest	0.018*	0.007**	0.001**	0.000***	0.002**
Instant test 1		0.035*	0.001**	0.000***	0.001**
Instant test 2			0.003**	0.001**	0.036*
Instant test 3				0.075	0.970
Delay test 1					0.103

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



**Fig. 6** The means of the six test scores of the three groups

**Table 4** The descriptive statistics of high-achievement students' scores at the EFL performance tests

Test	Kinect ( $N = 9$ )		SL ( $N = 8$ )		Paper ( $N = 8$ )	
	Mean	SD	Mean	SD	Mean	SD
Pretest (week 1)	15.72	3.13	14.25	5.20	14.31	4.57
Instant test 1 (week 2)	28.06	16.22	21.19	14.43	26.81	20.21
Instant test 2 (week 3)	29.72	16.85	26.00	20.48	28.19	19.07
Instant test 3 (week 4)	35.89	22.56	32.38	20.31	28.13	23.32
Delay test 1 (week 6)	32.50	19.25	34.75	18.02	29.56	27.39
Delay test 2 (week 11)	30.89	22.30	36.19	23.30	28.13	19.49

**Table 5** The descriptive statistics of low-achievement students' scores at the EFL performance tests

Test	Kinect ( $N = 9$ )		SL ( $N = 8$ )		Paper ( $N = 8$ )	
	Mean	SD	Mean	SD	Mean	SD
Pretest (week 1)	7.00	1.73	7.13	1.55	6.75	1.58
Instant test 1 (week 2)	10.00	3.68	9.50	4.50	9.00	3.89
Instant test 2 (week 3)	9.72	3.31	13.94	10.12	9.00	3.85
Instant test 3 (week 4)	10.94	5.80	18.19	15.79	10.63	8.37
Delay test 1 (week 6)	10.50	6.52	20.06	10.22	12.38	6.72
Delay test 2 (week 11)	10.44	6.68	13.38	10.19	9.00	7.45

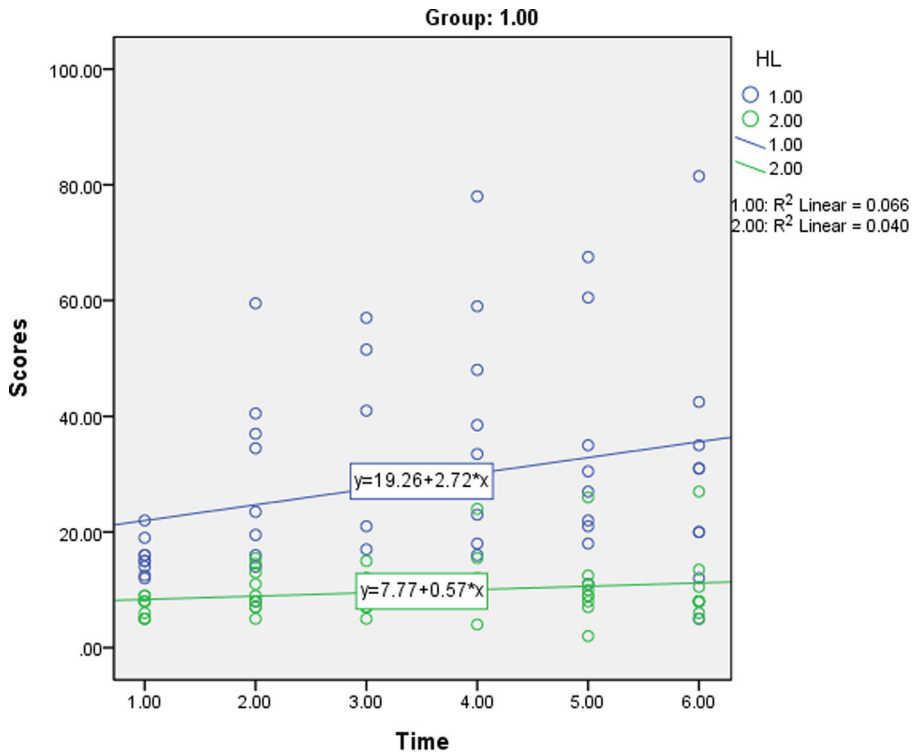
To analyze the benefits from doing embodied motions in different modes (by producing motions, Kinect group; by viewing avatar's motions, SL group; or without any of the former two kinds, paper group) to students with different EFL levels (both high and low) in

each group (Kinect, SL, and Paper), three linear mixed-effects models were performed respectively to model a linear relationship between test time and scores for each level group (see Table 6 and Figs. 7, 8, 9). As can be seen in the upper part of Table 6, the results of the first linear mixed effects model for the Kinect group show that the intercept for the high level is  $6.31 + 6.01 = 12.32$  and is significantly higher than for the low-achievement level at the pretest ( $t = 3.84, p = 0.002$ ). Additionally, the time coefficient of 0.95 is non-significant, suggesting that although the low-achievement students made an average gain by 0.95 points in the test score for each subsequent learning phase as well as the two delay tests, the improvement speed (slope) did not reach a significant level. However, the interaction estimates indicate a significant difference in slope for the high-achievement group compared to the low-achievement group. It indicates that the improvement speed made by the high-achievement group reached significant level ( $t = 2.38, p = 0.029$ ) compared to the low-achievement group, suggesting that the high-achievement group learned quicker than the low-achievement group. The learning gain over time is 0.42–6.89 points for the high- compared to the low-achievement students.

**Table 6** Results of linear mixed effects model

Group	Parameter	Estimate	Std. error	df	t	Sig.	95% confidence interval	
							Lower bound	Upper bound
Kinect	Intercept	6.312843	1.106499	15.140	5.705	0.000	3.956295	8.669392
	[HL = 1.00]	6.013291	1.564825	15.140	3.843	.002**	2.680628	9.345954
	[HL = 2.00]	0	0	–	–	–	–	–
	Time	0.944540	1.085087	17.386	0.870	0.396	– 1.340926	3.230005
	[HL = 1.00]* Time	3.655713	1.534545	17.386	2.382	0.029*	0.423576	6.887850
	[HL = 2.00]* Time	0	0	–	–	–	–	–
Second life	Intercept	4.677779	1.598735	13.017	2.926	0.012	1.224374	8.131184
	[HL = 1.00]	5.191284	2.260953	13.017	2.296	0.039*	0.307432	10.075136
	[HL = 2.00]	0	0	–	–	–	–	–
	Time	2.688758	1.181099	14.781	2.276	0.038*	0.168048	5.209467
	[HL = 1.00]* Time	2.286189	1.670326	14.781	1.369	0.192	– 1.278633	5.851011
	[HL = 2.00]* Time	0	0	–	–	–	–	–
Paper	Intercept	6.806188	2.053413	17.046	3.315	0.004	2.474764	11.137612
	[HL = 1.00]	7.261589	2.903964	17.046	2.501	0.023*	1.136030	13.387147
	[HL = 2.00]	0	0	–	–	–	–	–
	Time	0.476660	0.986491	13.693	0.483	0.637	– 1.643614	2.596935
	[HL = 1.00]* Time	2.093718	1.395108	13.693	1.501	0.156	–0.904803	5.092239
	[HL = 2.00]* Time	0	0	–	–	–	–	–

Dependent variable = scores; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

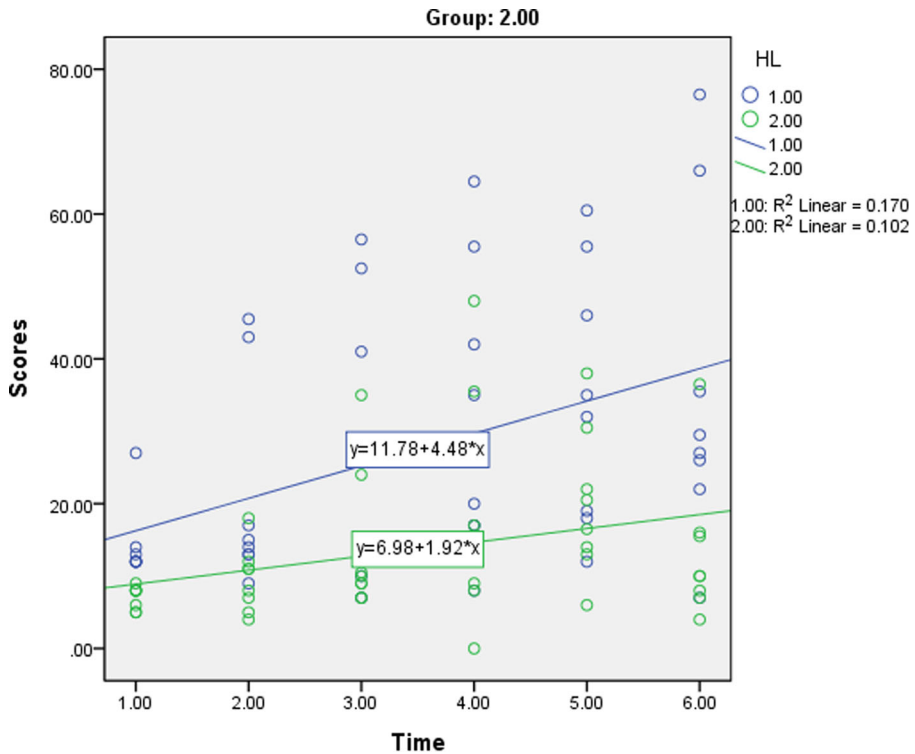


**Fig. 7** Regression lines between test time and scores by level for the Kinect group

The results of the second liner mixed effects model for the Second Life group (see the middle part of Table 6) show that the intercept for the high-achievement students is significantly higher (coefficient = 9.86) than for the low-achievement students (coefficient = 4.67) ( $t = 2.30$ ,  $p = 0.039$ ) at the pretest. Moreover, the time coefficient of 2.69 is significant ( $t = 2.28$ ,  $p = 0.038$ ), suggesting that the low-achievement group gained 2.69 points on average in the test score for each subsequent learning phase as well as the two delay tests. The learning gain for the low-achievement group over time is 0.17–5.21 points. However, there is no significant difference in slope for the high-achievement group compared to the low-achievement group, suggesting that although the high-achievement students also improved as the experiment progressed, both groups improved at a similar rate.

The results of the third liner mixed effects model for the Paper group also show that the intercept for the high-achievement level is significantly higher (coefficient = 14.06) than for the low-achievement level students (coefficient = 6.81) ( $t = 2.50$ ,  $p = 0.023$ ) at the pretest (see the bottom part of Table 6). In addition, the time coefficient of 0.48 was not significant, indicating that the low-achievement group did not make significant improvements over the learning phase and delay tests. There is no significant difference in slope for the high-achievement group compared to the low-achievement group, suggesting that the high-achievement group did not differ from the low-achievement group in learning rate although the average gain by the high-achievement group was 2.09 points compared with the low-achievement group.





**Fig. 8** Regression lines between test time and scores by level for the SL group

Based on the results listed in Table 6, producing motions (Kinect group) while listening to English phrases benefits the students with high EFL level rather than those with low EFL level. In contrast, in the SL group, students with both high and low EFL levels were benefited by watching their avatars' motions while listening to the identical English materials. The students in the paper group seemed to be not benefited from watching the paper-based materials and not doing any body motions while learning to the EFL phrases.

To synthesize the results described above, although the differences in EFL performances among the three groups did not reach a significant level, the improvements made by the students in both Kinect and SL groups were significant after the treatment. However, the improvements made by each group were different; the SL group demonstrated the greatest improvement, followed by the Kinect group. On the contrary, the students in the Paper group did not make significant improvement along the experiment. Furthermore, a further analysis on the scores of students with different EFL levels (high- vs. low-achievement) showed that the improvements made by the Kinect group were contributed only from those with high-achievement while the students with either high- or low-achievement in the SL group contributed to the significant improvements although the progression rate was not significant from the two levels. In contrast with the other two groups, no significant improvements were identified in both achievement levels.

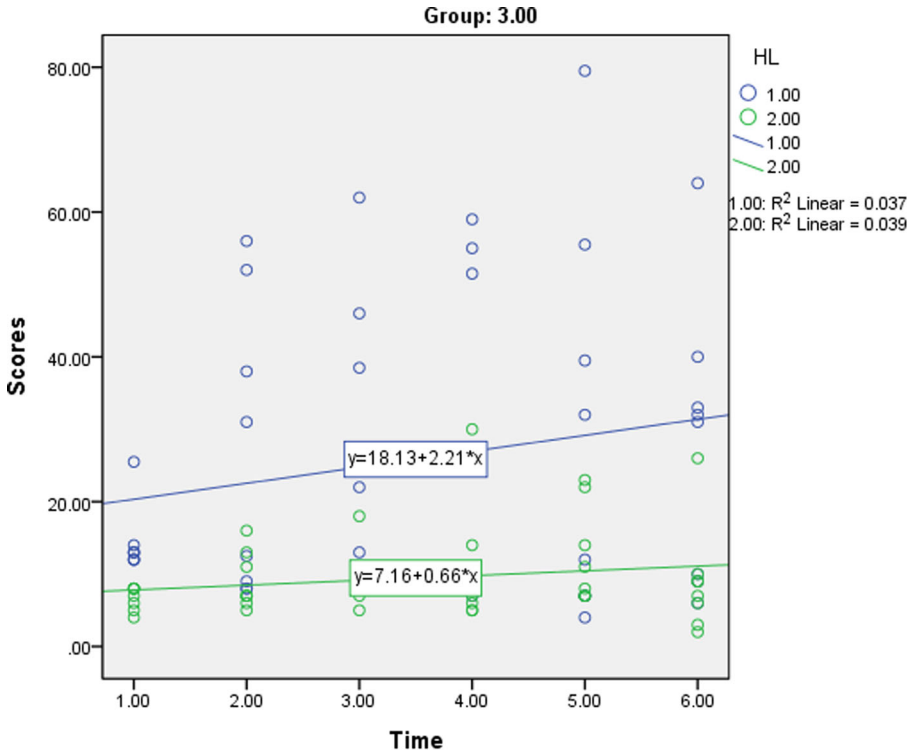


Fig. 9 Regression lines between test time and scores by level for the Paper group

### Discussion

The purposes of this study were to understand the effects of using body motions (real or virtual) on learning English sport-related phrases by elementary school students. Two different types of embodied learnings were adopted in the current study: producing physical body movement (gesture-based), and watching 3D avatar-based movement, and both were compared with non-embodied learning. It was found that the learning effects of using and without using the different embodied learnings were not significant although the participants in both the SL and Kinect groups received higher scores than those in the paper group. The finding here seems to be inconsistent with Macedonia and Knösche’s study (2011), which found that training of vocabulary through enactment led to better memory performance than the training of vocabulary through audiovisual. However, the treatment periods between the current study and Macedonia and Knösche (2011) was different, that of the former is much shorter than the latter. Therefore, whether the results of the current study would be similar to those of Macedonia and Knösche’s (2011) study if the treatment period was extended needs further investigation.

Additionally, the findings of the current study also seem not to be supported by the perspectives of embodied language processing which states that the movement made by the learner’s body would assist language learning (Rueschemeyer et al. 2010) if we only pay attention to the difference among the three treatment groups. However, if the focus is shifted to the improvements made by the participants in different motion groups, the results

then not only can be viewed as being in line with those in Zwaan and his colleagues' study (Zwaan et al. 2002), but also expand the arguments made by Mahon and Caramazza (2008) because it was found that learning with embodied motion, either by using physical body or watching one's avatar, appears to enhance elementary EFL learners' comprehension of the motions. From the findings obtained from a meta-analysis on the effects of gestures (embody motion) on foreign language learning, Hostetter (2011) also confirmed that children benefit from gestures when learning a foreign language. Additionally, the findings of the current study also echo the argument about the benefits of listening along with motions to verbal output (Wan et al. 2011). Wan et al. (2011) found that non-verbal autism children learning word along with motions made a significant improvement in verbal output. Furthermore, the participants in this study were only provided with only three learning sessions. The differences among the three groups might be significant if more learning opportunity is provided as described in the previous paragraph. In sum, it is suggested that a longer period should be taken in the future study to further confirm the effects of different embodied motions on EFL learners' listening performance.

In addition to the abovementioned confirmation of the effects of embodied motion on elementary students' English performance, some interesting findings were also identified in this study. First, it was found that the students watching their avatar motions while learning made the most improvements compared with others, both those learning by moving their bodies and not, especially for the students with low-achievement. The results seem to imply that the effects of involving ones' avatars in 3D virtual worlds is not limited to positively influencing their social presences as argued by Di Blas and Poggi (2007) and De Lucia et al. (2009), but is also extended to the cognitive domain. According to the Indexical Hypothesis, the possible explanation is that viewing one's avatar helped activate children's embodied experiences, which were then mapped onto the phrases they were learning. This process makes phrases learning meaningful without asking learners to actively generate meaningful associations. More future research, such as research on neuroscience, should be conducted to obtain more practical evidence to puzzle out the whole picture of the effects of virtual involvement on human's life.

Second, the students with a lower EFL level apparently learned better while watching their avatars doing the motions than by moving their own bodies or doing nothing as listening to the English phrases. The finding is very interesting. According to Lan et al. (2015b) study, learners engaged in a 3D virtual world outperformed their peers who learned with a 2D web-based environment in learning Chinese vocabulary. In their study, the participants in the 3D group moved their avatars in the virtual world while learning. On the other hand, those in the web-based group watched the 2D line drawings shown on the screen without any body movement. In that study (Lan et al. 2015b), neither group involved producing real body motion. Furthermore, according to relevant research on gesture-based study (e.g., Chao et al. 2013), learning involving gestures benefits learners' performances. Yet, the finding of this study proposes that learning by watching one's avatar motion benefits EFL students' learning more than by moving their own bodies, especially for those with low-achievement. The results might be caused by the degree of concentration of learner's attention. As suggested by Lan et al. (2015b), learners' gaze to the learned objects in 3D virtual worlds seemed to enhance their attentional control to the learning targets and thus helped them to ignore the influence from the surrounding. It was found that the participants in the SL group concentrated on gazing upon their 3D avatars' motions while those in the Kinect group focused on doing the motions. However, the gaze behavior did not work for those in the paper group although they gazed on the 2D line drawings on the paper yet did not perform as did those in the SL group. In a similar vein,

Glenberg et al. (2011) found that having children manipulate images of toys on a computer screen benefits reading comprehension better than physical manipulation of the toys immediately after reading the texts. One of the explanations is that manipulating images on the screen may encourage children to focus more on the texts, while manipulating the physical toys can turn children's attention away from the texts.

Following cognition perspective, Sanchez and Wiley (2006) argued that participants with higher working memory capacity have better controlled attention than participants with lower working memory and thus produce better learning outcomes. Based on the results obtained from this study, it is worthy of researchers' attentions to figure out whether learning in a well-designed and target-oriented 3D virtual world is likely to expand learners' working memory and thus helps them better control their attention. Although the findings described above are interesting, further investigation is needed, especially on the effects of different embodied learning on different levels EFL students' EFL learning due to the very small scale of the participants in the current study.

## Conclusion

How humans' bodies are involved in their learning has been an interesting issue and thus it has attracted researchers' attention as the technologies of both virtual reality and gesture recognition advanced in recent years (Lan et al. 2015a; Monahan et al. 2008). To add to the knowledge of the effect of human's motion on their EFL listening, two kinds of embodied learnings: real physical body (Kinect), and 3D virtual avatar (SL), were compared with non-embodied learning (paper) in this study. The evidences obtained from the current study favor 3D avatar-based embodied learning according to the overall outcomes. In addition, it is also found that gaze behaviors only work for students' learning when they gaze at their 3D avatar's motions rather than 2D line drawings, especially for those with low-achievement. Additionally, for those with high-achievement, by watching their 3D avatar moving, students remembered better than by moving their own bodies or doing nothing. The findings of this study encourage language educators to adopt 3D virtual reality (3D VR) worlds as one potential option when gesture-based approach is considered, especially for low-achievement students. Both a longer period of learning time and a larger scale of participants were suggested for future research for obtaining more solid evidences to understand the effects of motion on learning. Cross-discipline research, such as the cooperation among the experts in e-learning, second language acquisition, and neuroscience research, is also encouraged to gain knowledge from different perspectives to uncover the secret of the interaction of mind and body on humans' learning.

**Acknowledgements** The authors would like to thank the Ministry of Science and Technology, Taiwan, R.O.C., for financially supporting this research under Grant Nos. NSC 101-2511-S-003-031-MY3, MOST 104-2911-I-003-301, MOST 105-2511-S-003 -018 -MY3, MOST 106-2511-S-003 -015 -MY3, and MOST 103-2628-S-003-002-MY3. They are also thankful that this research was partially supported by "Chinese Language and Technology Center" and "Higher Education Sprout Project" of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan.

**Compliance with ethical standards**

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

## References

- Asher, J. J. (1969). The total physical response approach to second language learning. *Modern Language Journal*, 53, 3–17.
- Asher, J. J. (1977). *Learning another language through actions: the complete teacher's guidebook*. USA: Sky Oaks Productions.
- Asher, J. J., & Price, B. S. (1967). The learning strategy of the total physical response: Some age differences. *Child Development*, 38(4), 1219–1227.
- Atkinson, D. (2010). Extended, embodied cognition and second language acquisition. *Applied Linguistics*, 31(5), 599–622.
- Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology*, 102, 35–39.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Becker, A. S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., & Ananthanarayanan, V. (2017). *NMC horizon report: 2017 higher education edition*. Austin: The New Media Consortium.
- Borghini, A. M., & Cimatti, F. (2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia*, 48(3), 763–773.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulated the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24, 355–363.
- Chao, K. J., Huang, H. W., Fang, W. C., & Chen, N. S. (2013). Embodied play to learn: Exploring Kinect-facilitated memory performance. *British Journal of Educational Technology*, 44(5), E151–E155.
- Chen, N.-S., & Fang, W.-C. (2014). Gesture-Based Technologies for Enhancing Learning. In R.-H. Huang, Kinshuk, & N.-S. Chen (Eds.), *The New Development of Technology Enhanced Learning* (pp. 95–112). Berlin Heidelberg: Springer.
- Cowart, M. (2005). *Embodied Cognition*. Retrieved April 23, 2014 from <<http://www.iep.utm.edu/embodcog/>>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- De Lucia, A., Francese, R., Passero, I., & Tortora, G. (2009). Development and evaluation of a virtual campus on Second Life: The case of SecondDMI. *Computers & Education*, 52(1), 220–233.
- Derry, S. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist*, 3, 163–174.
- Deutschmann, M., Panichi, L., & Molka-Danielsen, J. (2009). Designing oral participation in second life: A comparative study of two language proficiency courses. *ReCALL*, 21(2), 206–226.
- di Blas, N., & Poggi, C. (2007). European virtual classrooms How to build effective 'virtual' educational experiences. *Virtual Reality*, 11, 129–143.
- Gao, M. (2009). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Chinese Journal of Phonetics*, 2, 43–50.
- Gibbs, R. W. (2005). *Embodiment and cognitive science*. New York: Cambridge University Press.
- Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2011). Improving early reading comprehension using embodied CAI. *Instructional Science*, 39(1), 27–39.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565.
- Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5(6), 664–674.
- Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, 9(5), 234–241.
- Holme, R. (2012). Cognitive linguistics and the second language classroom. *TESOL Quarterly*, 46(1), 6–29.
- Hostetter, A. B. (2011). When do gestures communicate?. A meta-analysis. *Psychological Bulletin*, 137(2), 297–315.
- Hung, I. C., Hsu, H. H., Chen, N. S., & Kinshuk. (2015). Communicating through body: A situated embodiment-based strategy with flag semaphore for procedural knowledge construction. *Educational Technology Research and Development*, 63(5), 749–769.
- Hung, I.-C., Lin, L.-I., Fang, W.-C., & Chen, N.-S. (2014). Learning with the body: An embodiment-based learning strategy enhances performance of comprehending fundamental optics. *Interacting with Computers*, 26(4), 360–371.
- Johnson, L., Adams, B. S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC horizon report: 2016 Higher education edition*. Austin: The New Media Consortium.

- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New Jersey: Prentice-Hall International.
- Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7, 54–60.
- Lan, Y. J. (2014). Does second life improve Mandarin learning by overseas Chinese students? *Language Learning & Technology*, 18(2), 36–56.
- Lan, Y. J. (2015). Contextual EFL learning in a 3D virtual environment. *Language Learning & Technology*, 19(2), 16–31.
- Lan, Y. J., Chen, N. S., Li, P., & Grant, S. (2015a). Embodied cognition and language learning in virtual environments. *Educational Technology Research & Development*, 63(5), 639–644.
- Lan, Y. J., Fang, S. Y., Legault, J., & Li, P. (2015b). Second language acquisition of Mandarin Chinese vocabulary: context of learning effects. *Educational Technology Research & Development*, 63(5), 671–690. <https://doi.org/10.1007/s11423-015-9380-y>.
- Lan, Y.-J., Kan, Y.-H., Hsiao, I. Y. T., Yang, S. J. H., & Chang, K.-E. (2013). Designing interaction tasks in Second Life for Chinese as a foreign language learners: A preliminary exploration. *Australasian Journal of Educational Technology*, 29(2), 184–202.
- Lan, Y. J., Kan, Y. H., Sung, Y. T., & Chang, K. E. (2016). Oral-performance language tasks for CSL beginners in Second Life. *Language Learning & Technology*, 20(3), 60–79.
- Lin, T. J., & Lan, Y. J. (2015). Language learning in virtual reality environments: Past, present, and future. *Educational Technology & Society*, 18(4), 486–497.
- Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, 20(12), 1507–1515.
- Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, 5(4), 196–211.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal. *Journal of Physiology*, 102, 59–70.
- McCafferty, S. (2002). Gesture and creating zones of proximal development for second language learning. *The Modern Language Journal*, 86, 192–203.
- McNeill, D. (1992). *Hand and mind: What the hands reveal about thought*. Chicago, IL: University of Chicago.
- Monahan, T., McArdle, G., & Bertolotto, M. (2008). Virtual reality for collaborative e-learning. *Computers & Education*, 50(4), 1339–1353.
- Morett, L. M., & Chang, L. Y. (2015). Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, 30(3), 347–353.
- Negueraela, E. J., & Lantolf, J. P. (2004). The “private function” of gesture in second language communicative activity: A study on motion verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics*, 14, 113–147.
- Regenbrecht, H., Schubert, T., & Friedmann, F. (1998). Measuring the sense of presence and its relations to fear of heights in virtual environments. *International Journal of Human-Computer Interaction*, 10, 233–249.
- Rueschemeyer, S. A., Lindemann, O., van Rooij, D., van Dam, W., & Bekkering, H. (2010). Effects of Intentional Motor Actions on Embodied Language Processing. *Experimental Psychology*, 57(4), 260–266. <https://doi.org/10.1027/1618-3169/a000031>.
- Ryu, D. (2013). Play to learn, learn to play: Language learning through gaming culture. *ReCALL*, 25(2), 286–301.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34(2), 344–355.
- Schubert, T., Friedmann, F., & Regenbrecht, H. (1999). Embodied presence in virtual environments. In R. Paton & I. E. Neilson (Eds.), *Visual Representations and Interpretations* (pp. 268–278). Berlin: Springer-Verlag.
- Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior*, 29(4), 217–234.
- Taipei City Government’s Department of Education (2000). *Curriculum guidelines of elementary English language teaching and learning*. Retrieved Dec 09, 2017 from [http://www.google.com.tw/url?sa=t&rc=t=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDYQFjAB&url=http%3A%2F%2Fwiki.estmue.tp.edu.tw%2Fimages%2F3%2F3e%2F%25E8%258B%25B1%25E6%2596%2587.doc&ei=5\\_i9UKWeAe7smAX1ioC4Bw&usq=AFQjCNFH892OVzBhifHrboQIUkRt8afiQ&sig2=Zc6kUIIoAnTwKrGD E\\_Q8pw](http://www.google.com.tw/url?sa=t&rc=t=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDYQFjAB&url=http%3A%2F%2Fwiki.estmue.tp.edu.tw%2Fimages%2F3%2F3e%2F%25E8%258B%25B1%25E6%2596%2587.doc&ei=5_i9UKWeAe7smAX1ioC4Bw&usq=AFQjCNFH892OVzBhifHrboQIUkRt8afiQ&sig2=Zc6kUIIoAnTwKrGD E_Q8pw)

- Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, 8(2), 219–235.
- Tsai, R. (2011). Teaching and learning the tones of Mandarin Chinese. *Scottish Language Review*, 24, 43–50.
- Wan, C. Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., et al. (2011). Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: a proof of concept study. *PLoS ONE*, 6(9), e25505. <https://doi.org/10.1371/journal.pone.0025505>.
- Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, 2, 116.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.
- Yang, H. W. (2010). *On writing Chinese characters and building Chinese character perception (zì gǎn) in teaching Chinese as a second language (Master's thesis)*. Taipei: National Chengchi University.
- Yoon, H., & Park, J. (2013). Avatar animation using skeleton tracking with Kinect sensor. *International Journal of Advancements in Computing Technology*, 5(12), 339.
- Zinszer, B. D., & Li, P. (2012). Predictors of native-like lexical categorization in Chinese learners of English. Paper presented at the *Second language Research Forum 2012*, Pittsburgh, PA, October, 2012.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, 13(2), 168–171.

**Yu-Ju Lan** is a Professor in the Department of Chinese as a Second Language at National Taiwan Normal University. Her research interests include technology enhanced foreign language learning, language learning in virtual worlds, mobile learning, and online synchronous teacher training.

**Wei-Chieh Fang** is a doctoral student in Applied Linguistics in Education at Washington University in St Louis. He holds a Master of Science in Cognitive Psychology. His recent research explores the potential of using smartphones to facilitate communicative language teaching and the development of strategic competence in speaking. He was also on several projects exploring the effects of motion-sensing technology on language learning.

**Indy Y. T. Hsiao** received his PhD in Computer Science and Information Engineering from National Central University, Taiwan. His research interests include virtual world learning, game-based learning, task-based learning, virtual reality, and Internet-of-Things.

**Nian-Shing Chen** is Chair Professor in the Department of Applied Foreign Languages at the National Yunlin University of Science and Technology, Taiwan. He has published over 400 academic papers in the international referred journals, conferences and book chapters. One of his papers published in *Innovations in Education and Teaching International* was awarded as the top cited article in 2010. He is the author of three books with one textbook entitled “e-Learning Theory & Practice”. He has received the national outstanding research awards for three times from the National Science Council in 2008, 2011–2013 and the Ministry of Science and Technology in 2015–2017. His current research interests include assessing e-Learning course performance; online synchronous teaching & learning; mobile & ubiquitous learning; gesture-based learning and educational robotics. Prof. Chen is serving as editorial board members for many international journals and guest editors for special issues of international journals. He has also organized and chaired numerous international conferences and workshops in the area of advanced learning technologies. Professor Chen is a golden core member of IEEE, ACM and the former Chair of the IEEE Technical Committee on Learning Technology (<http://ltf.ieee.org>). He is Editor-In-Chief of the SSCI-indexed *Journal of Educational Technology & Society*.