



Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning



Lysanne S. Post^{a,*}, Tamara van Gog^a, Fred Paas^{a,b}, Rolf A. Zwaan^a

^a Institute of Psychology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

^b Faculty of Education, University of Wollongong, Wollongong NSW 2522, Australia

ARTICLE INFO

Article history:

Available online 5 March 2013

Keywords:

Animations
Instructional design
Embodied cognition
Cognitive load theory

ABSTRACT

This study examined whether simultaneously observing and making gestures while studying animations would lighten cognitive load and facilitate the acquisition of grammatical rules. In contrast to our hypothesis, results showed that children in the gesturing condition performed worse on the posttest than children in the non-gesturing, control condition. A more detailed analysis of the data revealed an expertise reversal effect, indicating that this negative effect on posttest performance materialized for children with lower levels of general language skills, but not for children with higher levels of general language skills. The finding that for children with lower language ability, cognitive load did not decrease as they saw more animations provided additional support for this expertise reversal effect. These findings suggest that the combination of observing and making gestures may have imposed extraneous cognitive load on the lower ability children, which they could not accommodate together with the relatively high intrinsic load imposed by the learning task.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Although instructional animations are widely used in education, they are not always effective for learning, because the information presented is transient (Ayres & Paas, 2007). Information appears and then disappears and one is often required to keep the disappeared information in mind in order to comprehend the next piece of information. This is a highly demanding task for working memory, which is limited in capacity (e.g., Cowan, 2001; Miller, 1956). According to Cognitive Load Theory (CLT; Sweller, 1988; Sweller, Van Merriënboer, & Paas, 1998) this causes a high cognitive load. CLT describes three types of cognitive load that play a role in learning (Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Sweller et al., 1998). Intrinsic load is determined by the difficulty of the content of what is to be learned. The higher the number of interacting information elements, the more difficult the material is for the learner and the higher the intrinsic load (Sweller, 1994). Note that this also depends on learner expertise – with increasing expertise more information elements are combined into schemata, which reduces the intrinsic load of a task. Extraneous load is caused by the design of instruction and does not contribute to learning. Germane load on the other hand is also caused by the design of instruction, but is beneficial for learning. Thus, the last

two types of cognitive load can be altered by instructional designers, depending on the instructional format used. With instructional animations, for instance, it has been found that counteracting negative effects of transience by means of cueing (De Koning, Tabbers, Rikers, & Paas, 2009, 2010a) or segmenting (Spanjers, Van Gog, Wouters, & Van Merriënboer, 2012) makes animations more effective for learning.

Regarding the negative effect of transience on learning from instructional animations, there is an exception: When they demonstrate human movement tasks, dynamic visualizations such as videos or animations are often effective (Höffler & Leutner, 2007; Van Gog, Paas, Marcus, Ayres, & Sweller, 2009). It has been proposed (Van Gog et al., 2009) that this might be due to the mirror neuron system that is activated when one sees someone else perform an action – this is assumed to form the basis of the human capability to learn through imitation (Rizzolatti & Craighero, 2004). As human neurons respond to observing actions as a basis for learning, it might be that transience poses less of a problem in terms of working memory load, and procedures are acquired more easily when human movement tasks are depicted in animations.

In line with this notion of the mirror neuron system, embodied cognition theories also put forth an involvement of the motor system in learning. Embodied accounts of cognition postulate that cognitive processes are grounded in perception and bodily actions (Barsalou, 1999; Wilson, 2002). Thus, cognitive representations of symbols like numbers and letters are ultimately based on sensorimotor codes within a generalized system that was originally devel-

* Corresponding author. Address: The Institute of Psychology, Faculty of Social Sciences, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Tel.: +31 104088732.

E-mail address: l.s.post@fsw.eur.nl (L.S. Post).

oped to control an organism's motor behavior and perceive the world around it. In line with this view, memory for action phrases (e.g., 'Lift the pen.') has been shown to be better when participants had performed the action themselves (Engelkamp & Zimmer, 1997). Moreover, the semantics of such action phrases influenced behavior in another study, with faster reading times when meaning and motion were congruent (e.g., 'He started the car'; Zwaan, Taylor, & De Boer, 2010).

These embodied cognition studies suggest a link between semantics and the motor system, and it has been proposed that animations can be improved by activating the motor system by showing gestures (to which mirror neurons would respond) or asking learners to make gestures, even for non-human movement tasks (as in mathematical procedures or grammar; e.g., De Koning & Tabbers, 2011). Importantly, making gestures has been shown to lower cognitive load during math problem solving (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001) and to foster learning: When instructed to gesture while explaining math problems, children added new problem-solving strategies to their repertoire and remembered more from a subsequent lesson from the teacher (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007) and this beneficial effect was retained after four weeks (Cook, Mitchell, & Goldin-Meadow, 2008). Observation of gestures was also found to be effective for children's learning (Ping & Goldin-Meadow, 2008). Children had higher learning benefits when they saw guiding gestures (indicating sizes of objects) while learning Piagetian conservation tasks than when they did not observe gestures.

The present study focuses on the role of gestures in learning first-language grammar rules from animations, more specifically the grammatical rules for transforming an active sentence into a passive sentence. Considering language acquisition, research on the effects of gestures has mainly focused on second language learning and on concrete topics such as word learning. For instance, a study on word learning found that French children who were instructed to imitate gestures during word learning produced more English words on a test than children who were not instructed to gesture (Tellier, 2008). However, little research has been done considering the use of gestures in first language acquisition and learning more abstract concepts, such as grammar rules. Thus, it is unknown whether effects of gestures extend to learning abstract concepts in one's native language. Although, both observing gestures and making gestures have been shown to positively affect learning, the effects of the combined use of both techniques are unknown. We would predict learning benefits of both observing and making gestures through activation of the motor system and lightening of cognitive load. It is plausible that the effects would add up to an even higher learning benefit than of each of them separately. However, we have not found any literature examining this combined effect of simultaneously observing and making gestures. Moreover, very little research has been conducted on learning such abstract content as grammar rules from animations. Most research on instructional animations has focused on biological (e.g., how the heart works; De Koning, Tabbers, Rikers, & Paas, 2010b), natural (e.g., how lightning develops; Mayer & Moreno, 1998), or mechanical processes (e.g., how a piano works; Boucheix & Lowe, 2010), or on human-movement (e.g., origami; Wong et al., 2009) and problem-solving procedures (e.g., probability calculation; Spanjers et al., 2012). To the best of our knowledge, there are some studies on second language acquisition from animations (e.g., Roche & Scheller, 2008), but none on first language learning.

In sum, based on the above review of the literature, we propose that the effectiveness of grammar animations could be enhanced through gestures. Gestures are assumed to activate the motor system, thereby lightening cognitive load and enhancing learning. It should be noted that this beneficial effect on cognitive load and learning is not necessarily expected for children with all levels of

expertise on the subject matter. First, it is plausible that learning of grammar rules is better for children with higher general language skills than for children with lower general language skills. Second, the effect of gestures could potentially differ between children depending on their level of language skills. That is, research on the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003) has shown that an instructional format may cause different effects on cognitive load and learning for learners with different levels of expertise. For instance, in a study on acquiring skills to use a database program, novices were found to benefit more from worked examples, whereas more experienced learners had equal learning benefits from worked examples as from exploration; the worked-out steps were redundant for them and no longer contributed to their learning (Tuovinen & Sweller, 1999; see also Kalyuga, Chandler, Tuovinen, & Sweller, 2001).

Considering the present study this could mean that instructions to gesture might be effective for children with lower, but not for children with higher levels of language skills, for whom gestures might be redundant. The opposite might also be possible, that instructions to make gestures impose additional load, which might be beneficial for higher level learners (i.e., germane load) but might cause such high load for lower level learners, that it impairs their learning. Given that there is no prior research in this area, it is hard to predict whether language skills have an effect and if so, in which direction, but research on the expertise reversal effect suggests it is important to consider level of language skills as a factor (instead of a covariate).

In sum, the present study combines the use of gestures and animations in language acquisition. The question that is being examined is whether simultaneously observing and making gestures while studying animations, contributes to grammar learning. This experiment focuses on teaching children which grammatical rules are involved in the transformation of an active sentence (e.g., 'Pete is petting the dog') into a passive sentence (e.g., 'The dog is being petted by Pete') through animations. It is hypothesized that children will experience lower cognitive load and perform better on both an immediate and delayed (after one week) posttest when they saw and made gestures while studying animations. Depending on the amount of forgetting, an interaction of Condition and Posttest might occur. That is, it could be that gestures lead to less forgetting, which would become evident through an interaction effect. However, it can also be that both groups show similar forgetting. In that case, there will be no interaction. Motivation and perceived difficulty were assessed as a check, as these variables might provide alternative explanations for possible cognitive load and learning effects when they would differ between conditions. Finally, in light of the expertise reversal effect, effects of levels of language skill will be explored.

2. Method

2.1. Participants and design

Sixty-nine Dutch primary school children in grade 6 participated in the experiment, they came from four classrooms in two schools. Two participants were excluded from all analyses because teachers stated that their IQ was extremely low (≤ 70). The age of the 67 remaining participants ranged from 10 to 13 years ($M = 11.57$, $SD = .70$) and 34 of the participants were boys. All children were born in the Netherlands and were sufficiently fluent in Dutch to understand the instructions and participate in the experiment. Fifteen children had one or two parents who were not born in the Netherlands. Five participants were absent during the second session and were therefore excluded from analyses concerning the delayed posttest (i.e. all performance measures).

This experiment had a $2 \times 2 \times 2$ design with two between-subjects factors (Condition: Gesture, $n = 33$ and Control, $n = 29$; Language Skills: High, $n = 31$ and Low, $n = 31$) and one within-subjects factor (Posttest: Immediate, Delayed; $N = 62$). Children were pseudo-randomly assigned to an animation condition (Gesture, Control), matching for general language skills of which the experimenter had received an index from the teacher for each child based on a national standardized test. These tests result in a category score of A, B, C, D, or E. Children with an A are among the best 25% of all Dutch children that have done that specific test; B stands for the next 25%; C for the 25% after B; D for the 15% after C; and E for the lowest-scoring 10% of children. Both schools used such a standardized language ability test; however, they used different versions. Therefore, the scores on the pretest (a general language test constructed for this study) were used to assign children to the High and Low language ability conditions, because this measure was the same for all participants. A regression analysis was conducted on the index of language skills provided by the teachers, which verified that the pretest actually measured language skills ($F(4, 62) = 9.90$, $p < .0001$, $R^2 = .62$), with a higher index of language skills resulting in a higher score on the pretest.

There were no significant differences in the numbers of boys and girls between conditions ($\chi^2(3) = 3.04$, $p = .385$).

2.2. Materials

A pretest was constructed, consisting of 31 questions. This general language test was constructed with two purposes. The main goal of this test was to assess prior knowledge of the concepts relevant to the topic of sentence transformation (active and passive sentences). Because the questions regarding the relevant concepts were part of a larger general language test, these concepts were not specifically primed. The second aim of pretesting the children with this language test was to determine their general language skills.

The animations used for this experiment were built in Microsoft PowerPoint and lasted 62 s each. The first two of the four animations were preceded by a slide showing the begin state (i.e. active) and end state (i.e. passive) of the sentence that was being transformed in the animation, so that children had a little preview of what they were going to see in the animations. During each animation, a voiceover explained every step of the transformation. In the Gesture condition, a human arm was visible throughout the entire animation (see Fig. 1 for an example) that moved the words to the right places. The movements of the arm are the observed gestures that are examined in this study. These gestures contain procedural information about the grammar rules. This is similar to the gestures in mathematical problem solving in Broaders et al. (2007) study. In the control condition there was no arm present and words just moved from one place to the other in a straight line. Because our participants were required to make arm movements, the sentences in the animations were deliberately not about making a

movement of the arm in any kind (e.g. 'Kim is reading the book.') to prevent semantic interference (i.e., semantics of action phrases can influence one's motion; e.g. Zwaan et al., 2010). After each animation children were asked to rate the effort invested in understanding it. Mental effort was measured on a 9-point-scale developed by Paas (1992), ranging from 'very very low effort' to 'very very high effort'. This measure gives an indication of the cognitive load children experience during animation study.

A short questionnaire, consisting of seven statements, was used to examine what children thought of this way of learning. Statements 1, 2 and 6 were about whether children liked this kind of learning method and whether they would like to do this more often (i.e. 'I thought learning from these animations was interesting', 'I thought learning from these animations was useful', and 'I want to learn from animations more often'). Statements 3, 4, 5, and 7 concerned perceived difficulty of the learned material (i.e. 'I thought learning from these animations was difficult', 'I think I understand the animations', 'I would like to see more animations', and 'I think that I am capable of making a passive sentence on the test'). Children rated their agreement with each statement on a 7-point scale ranging from 'I do not agree at all' to 'I totally agree'.

The posttest was programmed in E-Prime. First, three easy, randomly presented sentences, similar to the ones in the animations (e.g. 'Tessa calls Mirjam.'), had to be transformed from active to passive. Then, three more difficult sentences (e.g. 'Tom puts on his shoes for the soccer game.') were also randomly presented. Next, three moderately difficult sentences had to be transformed back from passive to active (e.g. 'The heavy furniture is being moved by Bianca.'). After that, three comprehension questions were posed ('What is the object of this sentence?', 'What should you do with it when you make a passive sentence?', and 'When you have made a passive sentence it is no longer the object of the sentence. What is it called now?'). Finally, children were asked to describe how they would explain the procedure they just learned to a classmate. Two versions of the posttests were created (the same structure, but with different sentences), and the order of posttest versions was counterbalanced across participants.

The standard curriculum in the Netherlands does contain the topic of passive sentence, but it is discussed quite briefly. Since there was no extensive standard teaching method available for the instruction or testing of active–passive sentence transformations, both the animations and the posttests were constructed for this study in collaboration with primary school teachers.

2.3. Procedure

First, the children were pretested on prior knowledge of the concept of active–passive sentence transformations and on their general language skills; all children within one class were tested at once in their classroom. Four children were absent during the pre-test session. They filled out the pretest just before they participated in the study phase.

The study phase took place in individual sessions starting with the first participant right after the pretest. This phase consisted of a short introduction and the presentation of the animations. Children were pseudo-randomly assigned to an animation condition prior to this session, matching for general language skills of which the experimenter had received an index from the teacher for each child (based on a national standardized test). In the introduction, the experimenter made sure the child had sufficient knowledge of the concepts 'verb', 'subject', 'object', and 'past participle'. As soon as everything was clear to the child, he or she was instructed to watch and listen carefully to the animations. Depending on the condition children were assigned to, they were instructed to gesture along with the hand that moved the words in the animations (Gesture condition) or watch the animations in which words



Fig. 1. Screenshot of animation.

moved ‘automatically’ while sitting on their hands (Control condition). After each animation children were asked to rate the amount of effort it took to understand the animation by saying a number from 1 to 9 out loud (this procedure and the meaning of the numbers had been explained to them during the introduction). Four animations were presented and after that the questions were posed regarding interest and perceived difficulty.

Directly following the study phase, the immediate posttest about passive sentences was completed. Children were instructed to answer the questions as accurately and as quickly as possible and to keep the animations in mind while answering the questions. Seven days later, the delayed posttest was completed. Due to practical difficulties at the school, for nine participants there was a delay of six days between the posttests and for one participant it was ten days. At the end of the last phase, children received a small gift (a pencil or an eraser).

2.4. Data analysis

The dependent variables were the performance on the different parts of the posttests (Sentence transformations, Comprehension questions, Explanation, and Descriptions of movements in the explanation). The ‘Explanation’ score is the number of correct steps of the transformation procedure participants reported. The ‘Descriptions of movements in the explanation’ score is the number of steps reported that described the movement of a part of the sentence. Movement of parts of the sentences is crucial in the transformation of an active sentence into a passive one. Four of such movements could be mentioned in the explanation. These are (1) object to the beginning, (2) subject to the end, (3) adding the Dutch verb ‘worden’ (meaning: ‘is being’), and (4) adding the Dutch word for ‘by’ (‘door’). This resulted in a separate ‘movements in explanation’ score. Mental effort, Interest and Perceived difficulty are three other dependent variables that were examined with Condition as the between subjects factor ($N = 67$).

Language skills were included as a between-subjects factor in the analysis. Level of language skills of the children was based on the performance on the pretest (which was highly related to the index provided by the teachers; see ‘participants’ section). Children with a performance score of 55% or lower were classified as having a low level of language skills, those with a score above 55% were classified as having a high level of language skills.

3. Results

3.1. Interest, perceived difficulty, and mental effort

The average scores on Interest and Perceived difficulty are shown in Table 1. The statements on Interest and Perceived difficulty were rated similarly by children in the four conditions (all $ps > .05$ for each of the seven 2 (Condition) \times 2 (Skills) ANOVAs),

Table 1
Mean (SD) agreement on interest and perceived difficulty statements.

Statement	Gestures ($n = 33$)	Control ($n = 29$)
1. Interest	5.58 (1.20)	5.35 (1.20)
2. Useful	5.82 (1.13)	5.76 (.96)
3. Difficulty	2.36 (1.85)	2.74 (1.81)
4. Understood	6.18 (1.18)	6.21 (1.01)
5. Want more	4.52 (1.77)	4.35 (1.32)
6. Want more often	5.45 (1.60)	5.85 (1.28)
7. Capable	5.64 (1.34)	5.74 (1.19)

Note: Agreement was measured on a 7-point scale.

so these factors cannot have influenced mental effort (i.e. cognitive load) ratings.

A 2 (Condition) \times 4 (Animation) \times 2 (Skills) ANOVA revealed no effect of Condition ($F(1, 63) = .01$, $p = .936$, $\eta_p^2 < .001$), nor any interaction effects (all $ps > .05$) on the mental effort scores (see Table 2). So there was no difference between the Gesture group and the Control group in how much effort they invested to understand the animations and this was true for all animations. However, there was a significant main effect of Animation (multivariate Wilks’ $\lambda = .67$, $F(3, 61) = 9.86$, $p < .0001$, $\eta_p^2 = .33$), with a trend showing that each subsequent animation was rated lower on mental effort: Animation 2 was rated significantly lower than Animation 1 ($F(1, 65) = 19.99$, $p < .001$, $\eta_p^2 = .24$), 3 was not rated significantly lower than 2 ($F(1, 65) = 2.71$, $p = .105$, $\eta_p^2 = .04$), but 4 was rated significantly lower than 3 ($F(1, 65) = 6.45$, $p = .014$, $\eta_p^2 = .09$). There was also a main effect of Skills ($F(1, 63) = 9.48$, $p = .003$, $\eta_p^2 = .13$), indicating that children with low levels of language skills had to invest more mental effort to understand the animations, as one would expect.

3.2. Performance

The pretest questions on active–passive sentence transformation were analyzed to get an indication of prior knowledge regarding the topic of the animations. Data showed that none of the children knew what an ‘active’ or a ‘passive’ sentence was prior to the study phase.

In contrast to our hypothesis, a 2 (Condition) \times 2 (Posttest) \times 2 (Skills) ANOVA showed that overall, the Control group performed better than the Gestures group ($F(1, 58) = 6.09$, $p = .017$, $\eta_p^2 = .10$) on the transformation of sentences (see Table 3). There was no main effect of posttest or any interaction effects between the variables Condition, Skills, and Posttest (all $ps > .05$). There was a main effect of Skills ($F(1, 58) = 15.23$, $p < .001$, $\eta_p^2 = .21$, which (not surprisingly) indicated that children with high language skills performed better than children with low language skills. The performance data on the comprehension questions show a similar pattern of results (main effect of Skills: $F(1, 58) = 8.29$, $p = .006$, $\eta_p^2 = .13$, and a marginally significant effect of Condition: $F(1, 58) = 3.33$, $p = .073$, $\eta_p^2 = .05$). The answers to the explanation question of the posttests resulted in two scores as is visible in Table 3. Both scores were analyzed with similar ANOVAs as the sentence transformations and comprehension questions just described. Regarding the Explanation data, only a main effect of Skills was found ($F(1, 58) = 16.44$, $p < .001$, $\eta_p^2 = .22$). Similar results were found for the mention of movements in the explanation (only a main effect for Skills, $F(1, 58) = 11.53$, $p < .01$, $\eta_p^2 = .17$). Again, children with higher language skills scored higher.

Even though no interactions were found, the main effect of Skills indicates that performance differs for the two groups of language skills. As can be seen in Table 3, the effect of Skills on Sentences and Comprehension seems to be caused mainly by one group (Low Skills Gesture condition). Separate ANOVAs on low and high language skills, reveal that the Control condition performed better than the Gesture condition on Sentences ($F(1,$

Table 2
Mean (SD) mental effort ratings.

Animation	1	2	3	4
Low, control ($n = 15$)	4.06 (1.39)	3.29 (1.79)	2.88 (1.83)	2.88 (2.06)
Low, gestures ($n = 16$)	3.88 (2.42)	3.38 (2.45)	3.13 (1.93)	2.88 (2.13)
High, control ($n = 14$)	3.35 (1.22)	2.24 (.90)	1.94 (.97)	1.59 (.62)
High, gestures ($n = 17$)	2.53 (1.55)	2.12 (1.05)	2.29 (1.31)	1.82 (1.13)

Note: Mental effort ratings were measured on a 9-point scale.

Table 3
Mean (SD) performance scores.

		Low (<i>n</i> = 31)		High (<i>n</i> = 31)	
		Control (<i>n</i> = 15)	Gestures (<i>n</i> = 16)	Control (<i>n</i> = 14)	Gestures (<i>n</i> = 17)
Sentences	Posttest 1	0.76 (.19)	0.58 (.29)	0.90 (.17)	0.85 (.18)
	Posttest 2	0.76 (.24)	0.54 (.34)	0.90 (.14)	0.82 (.25)
Comprehension	Posttest 1	0.58 (.37)	0.45 (.35)	0.76 (.38)	0.70 (.38)
	Posttest 2	0.58 (.44)	0.26 (.31)	0.74 (.35)	0.64 (.35)
Explain	Posttest 1	0.41 (.26)	0.32 (.24)	0.54 (.17)	0.59 (.26)
	Posttest 2	0.37 (.25)	0.27 (.30)	0.57 (.17)	0.58 (.22)
Movements in explanation	Posttest 1	0.50 (.38)	0.41 (.41)	0.71 (.29)	0.65 (.40)
	Posttest 2	0.41 (.40)	0.27 (.39)	0.68 (.33)	0.71 (.37)

Note: Performance is displayed in proportions.

29) = 5.36, $p = .028$, $\eta_p^2 = .16$) and marginally on Comprehension ($F(1, 29) = 3.47$, $p = .073$, $\eta_p^2 = .11$) only for the children with low levels of language skills. No main effects or interaction effect between Performance and Condition were found for Explanation and Movements in Explanation (all $ps > .05$) for high or low Skills.

4. Discussion

This study examined whether simultaneously observing and making gestures while studying animations, would contribute to grammar learning, more specifically the transformation of active into passive sentences. Surprisingly, however, we found the opposite: Children who observed and made gestures during the study phase performed worse on a subsequent test than children who did not. Differences in interest and perceived difficulty cannot account for this effect, because there were no differences between conditions on those variables. Moreover, mental effort ratings showed that children had to invest less effort in each subsequent animation, regardless of experimental condition.

However, an exploratory analysis suggests that these overall effects did not apply to children with higher and lower language ability alike. Rather, the negative effects of gesturing on learning only applied to children with low levels of language skills (children with high language skills performed equally well in both conditions), whereas the clear decrease in mental effort ratings over the animations applied only to children with higher skill levels. The combination of these findings suggests that potentially, the additional instruction to simultaneously observe and make gestures was too demanding for these children with low language skills.

Even though an expertise reversal effect (Kalyuga et al., 2003) would explain why we found the opposite of what we expected, at first sight it does not seem to fully explain why we did not find what we predicted (i.e., that gestures would lighten the cognitive load and facilitate learning) for children with higher language skills. A potential explanation of this lack of effect for higher expertise children would be that *all* children experienced extraneous (ineffective) cognitive load due to the instructions to simultaneously observe and make gestures, but that the children with a higher level of expertise could handle this without detrimental effects on learning, because the learning task was lower in intrinsic load for them.

So how might the combination of simultaneously observing and making gestures cause extraneous load? In hindsight, because students had to follow the arm making the transformations, they may have paid a lot of attention to the arm itself, distracting their attention – at least temporarily – away from the words that were being moved and the verbal explanation of the moves. High expertise students might have had sufficient working memory capacity available for dealing with this ‘dual task’, but not lower expertise

children. If this indeed explains our results, then it is still possible that only observing gestures (i.e., without moving along), or only making gestures (i.e., following the movement of the words, instead of the arm) would be beneficial for learning, and future research should establish whether this indeed is the case.

However, it might also be that observing and imitating gestures is just too passive to facilitate learning (De Koning & Tabbers, 2011). In that case it would be essential that the gestures children make are more active. One way to accomplish this is to let children generate the gestures themselves (i.e., enactment), instead of imitating gestures. For instance, enactment of a story (through manipulation of objects) has been found to lead to beneficial effects on learning (Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004); whether it would also lead to learning benefits with animations remains to be examined. It would seem that simultaneous enactment might further aggravate effects of transience for low language level children, but possibly it might be effective for higher-level children. Another possibility is that letting children create their own gestures will lead to more natural gestures, turning the instruction to gesture in a less conscious task. This could reduce extraneous load and enhance learning.

Another way in which our gesturing instruction might have caused extraneous load, is that it also included some redundant movements. When one word was moved, the arm moved to the next word that had to be moved. While this focuses learners' attention on the next upcoming action, this second movement in itself does not mean anything for the procedure of the sentence transformation, because it does not involve moving a word to a certain location. As it was the task to make the same gestures as the arm in the animations, children also made several of these ‘meaningless’ gestures. Maybe the mapping between the movement required in the procedure and the gesturing would be stronger or facilitated when every single gesture was meaningful (for a discussion of mapping of movements, see De Koning & Tabbers, 2011).

It should be noted that the effects applied to the production of passive sentences and the answers to comprehension questions. There were no effects found on both Explanation scores. This is probably due to the difficulty of the task. For children with low language skills, this task was so difficult that performance was very low in both conditions. Highly skilled children performed better on this task, but just as with the other performance measures, did not experience facilitation of gesturing.

Finally, an alternative explanation for the present results relates to Mayer & Moreno's (2003) coherence principle, which states that “students understand a multimedia explanation better when interesting but extraneous material is excluded rather than included” (Mayer & Moreno, 2003, p. 48). Although we assumed the movements and gestures that were shown to be relevant to the transformation of sentences, the current experimental design does not allow us to make a conclusive decision about their relevance. The relevance of the material in the animations seems to be supported

by the fact that all children learned at least something about the concept of active–passive transformations, but we cannot exclude the possibility that animation with a hand moving words without the instruction to gesture along, or even a static presentation format would have resulted in higher learning outcomes. Future studies could examine the relevance of animations and gesture to acquire strategies to learn grammatical structures by directly comparing static and animated instruction methods, and animated instructional methods with and without gestures.

5. Conclusions

In sum, simultaneously observing and making gestures had a negative influence on learning grammar rules for children with a low level of language skills. It seems plausible that this task was too demanding for lower skilled children, and several explanations have been provided for why this might be the case. However, even though it was not too demanding for higher skilled children, they did not experience facilitation of gestures either. One should take great caution in drawing conclusions with respect to educational practice from a single study, but at least our results suggest that the combination of observing and making gestures while learning from animations might be ineffective. Future research should address the potential limitations of our study mentioned above (i.e., the ‘dual-task’, too passive learning, and redundant movements), in order to examine under which circumstances observing and/or making gestures could be effective for children with varying levels of language skills.

Acknowledgements

This research was funded by the Netherlands Organization for Scientific Research (NWO; Project number 411–10–907). We thank Bonnie van Huik and Jacqueline de Nooijer for their assistance during preparation of stimulus material.

References

- Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology*, 21, 811–820. <http://dx.doi.org/10.1002/acp.1351>.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660. <http://dx.doi.org/10.1017/S0140525X99002149>.
- Boucheix, J., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, 20, 123–135. <http://dx.doi.org/10.1016/j.learninstruc.2009.02.015>.
- Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, 136, 539–550. <http://dx.doi.org/10.1037/0096-3445.136.4.539>.
- Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, 106, 1047–1058. <http://dx.doi.org/10.1016/j.cognition.2007.04.010>.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185. <http://dx.doi.org/10.1017/S0140525X01003922>.
- De Koning, B. B., & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualizations: An embodied perspective. *Educational Psychology Review*, 23, 501–521. <http://dx.doi.org/10.1007/s10648-011-9173-8>.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, 21, 113–140. <http://dx.doi.org/10.1007/s10648-009-9098-7>.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010a). Learning by generating vs. Receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education*, 55, 681–691. <http://dx.doi.org/10.1016/j.compedu.2010.02.027>.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010b). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, 20, 111–122. <http://dx.doi.org/10.1016/j.learninstruc.2009.02.010>.
- Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, 96, 43–60. [http://dx.doi.org/10.1016/S0001-6918\(97\)290005-X](http://dx.doi.org/10.1016/S0001-6918(97)290005-X).
- Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, 96, 424–436. <http://dx.doi.org/10.1037/0022-0663.96.3.424>.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12, 516–522. <http://dx.doi.org/10.1111/1467-9280.00395>.
- Höfler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17, 722–738. <http://dx.doi.org/10.1016/j.learninstruc.2007.09.013>.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23–31. http://dx.doi.org/10.1207/S15326985EP3801_4.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93, 579–588. <http://dx.doi.org/10.1037/0022-0663.93.3.579>.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90, 312–320. <http://dx.doi.org/10.1037/0022-0663.90.2.312>.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43–52. http://dx.doi.org/10.1207/S15326985EP3801_6.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97. <http://dx.doi.org/10.1037/h0043158>.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434. <http://dx.doi.org/10.1037/0022-0663.84.4.429>.
- Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38, 63–71. doi: http://dx.doi.org/10.1207/S15326985EP3801_8.
- Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, 44, 1277–1287. <http://dx.doi.org/10.1037/0012-1649.44.5.1277>.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192. <http://dx.doi.org/10.1146/annurev.neuro.27.070203.144230>.
- Roche, J., & Scheller, J. (2008). Grammar animations and cognition. In F. Zhang & B. Barber (Eds.), *Handbook of research on computer-enhanced language acquisition and learning* (pp. 205–218). Hershey: IGI Global.
- Spanjers, I. A. E., Van Gog, T., Wouters, P., & Van Merriënboer, J. J. G. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers & Education*, 59, 274–280. <http://dx.doi.org/10.1016/j.compedu.2011.12.024>.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. <http://dx.doi.org/10.1023/A:1022193728205>.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. http://dx.doi.org/10.1207/s15516709cog1202_4.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4, 295–312. [http://dx.doi.org/10.1016/0959-4752\(94\)90003-5](http://dx.doi.org/10.1016/0959-4752(94)90003-5).
- Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, 8, 219–235. <http://dx.doi.org/10.1075/gest.8.2.06tel>.
- Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, 91, 334–341. <http://dx.doi.org/10.1037/0022-0663.91.2.334>.
- Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review*, 21, 21–30. <http://dx.doi.org/10.1007/s10648-008-9094-3>.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9, 625–636. <http://dx.doi.org/10.3758/BF03196322>.
- Wong, A., Marcus, N., Ayres, P., Smith, L., Cooper, G. A., Paas, F., et al. (2009). Instructional animations can be superior to statics when learning human motor skills. *Computers in Human Behavior*, 25, 339–347. <http://dx.doi.org/10.1016/j.chb.2008.12.012>.
- Zwaan, R. A., Taylor, L. J., & De Boer, M. (2010). Motor resonance as a function of narrative time: Further tests of the linguistic focus hypothesis. *Brain and Language*, 112, 143–149. <http://dx.doi.org/10.1016/j.bandl.2008.11.004>.