# The effectiveness of multimedia programmes in children's vocabulary learning

## Joana Acha

*Joana Acha is predoctoral researcher from Universitat de València. Address for correspondence: Joana Acha, Departamento de Metodología, Facultad de Psicología, Av. Blasco Ibáñez, 21, 46010-Valencia, Spain. FAX: +34 963 864697. Email: amorjo@uv.es*

**Abstract**

The present experiment investigated the effect of three different presentation modes in children's vocabulary learning with a self-guided multimedia programmes. Participants were 135 third and fourth grade children who read a short English language story presented by a computer programme. For 12 key (previously unknown) words in the story, children received verbal annotations (written translation), visual annotations (picture representing the word), or both. Recall of word translations was better for children who only received verbal annotations than for children who received simultaneously visual and verbal annotations or visual annotations only. Results support previous research about cognitive load in e-learning environments, and show that children's learning processes are hindered by limited working memory. This finding implies a challenge for multimedia programmes designed for children and based on self-regulated learning.

Which presentation mode of a multimedia programme will help in learning more vocabulary in an unknown language: one with pictures, one with words or one combining both representations? The growing use of multimedia programmes in schools (eg, via the Internet or via ad hoc programmes) has conveyed a great support to individual learning in the sense that it allows the student to access different kinds of information at the same time (words, pictures, audio) and at their own pace. Thus, the designers of multimedia programmes face the challenge of not only adapting programmes to the learner's certain characteristics but also of fitting the requirements for these programmes to be effective (Ávila & Sadoski, 1996).

Previous research about presentation modes in different learning environments has shown that, in some cases, adding a picture to the word may enhance the comprehension of school texts and the information storage process (see Tindall-Ford, Chandler & Sweller, 1997). This result may be attributed to the fact that the picture makes recall

easier by identification of previous experience stored in long-term memory, or alternatively, the word next to the picture reduces the interpretation range (ie, it makes it more reliable). The facilitative effect obtained when the picture is presented next to the word (Mayer, 1997) is a result of dual coding processes involved in associative learning. Hence, when a word and a picture are presented simultaneously, the verbal representation and the corresponding pictorial representation are processed at the same time in working memory, facilitating the organisation and integration of both picture and word through the corresponding connections (see Paivio, 1971), for a theoretical description of dual coding processes).

If one assumes that any design for a multimedia programme should be based on solid educational theory and evidence, it is important to examine the potential limits to this facilitative effect in an e-learning environment (Wild & Quinn, 1998). Indeed, simultaneous presentation of verbal and pictorial stimuli has not always proved to be useful in self-paced learning, probably because of limited working-memory capacity (see Baddeley, 1986, 1997), for a detailed description of working memory). The learning difficulties resulting when information in multimedia programmes overloads working memory can imply *cognitive load* (see Brünken, Plass & Leutner, 2003). Cognitive load may occur when two types of stimuli that supply the same information (eg, a written word and a picture related to the same concept) are perceived through the same information processing channel, which is a visual channel in this case. Reducing cognitive load involves cutting the unneeded duplication of the essential information. Thus, to create effective multimedia programmes, the combined verbal and pictorial presentation mode should be used only in case when its effectiveness has been clearly demonstrated (Mayer, Heiser & Lonn, 2001; Mayer & Moreno, 2002).

The issue is more complex, though. The study of the effectiveness of presentation modes in multimedia programmes has also focused on factors related to the learner's characteristics that may involve greater degree of difficulty in learning, such as the individual's visual or verbal ability. Prior studies have shown that, in certain learning situations, some learners can benefit more than others from the simultaneous presentation of word and picture (ie, dual coding implying a facilitative effect) (Riding & Grimley, 1999). In this light, using a comprehension task with self-paced multimedia programmes, learners with high ability (verbal and visual) benefit more than learners with low ability from the simultaneous presentation (word and picture) (Mayer & Sims, 1994). Therefore, it seems necessary to examine the conditions in which the learner's cognitive resources may not allow him to maintain the presented information in working memory, in order to reduce (or even eliminate) cognitive load when implementing any learning multimedia programme (eg, see Mayer & Moreno, 2003; Moreno & Mayer, 2002). Moreover, most research in this area has been conducted with adults. This hinders the generalisation of these findings to other populations, such as primary school children, where the use of self-paced multimedia programmes is more frequent in the classroom environment. On the positive side, the combined use of a word and a picture may reinforce the memory trace in children, but (on the negative side) it may

well force the child to handle an additional element, thus increasing the experienced cognitive load.

The main aim of this study was to examine which presentation mode is more effective for primary school children who are learning new vocabulary in a second language with a self-paced multimedia programme: either presenting only one stimulus (which involves only one connection, either verbal or pictorial) or presenting two stimuli simultaneously, both a word and a picture (which allows the integration of the two connections, verbal and pictorial). Previous research on vocabulary learning (using adult participants) has underlined the importance of the learning context in a multimedia environment (Duquette, Renié & Laurier, 1998) and has primarily focused on factors like learner's preference and his spatial and verbal ability. In the first case, when learners were able to choose the presentation mode according to their own preference, presenting words and pictures simultaneously was more effective than presenting only words or only pictures (Mayer, Plass, Chun & Leutner, 1998). However, a recent study by Plass, Chun, Mayer and Leutner (2003), in the context of English-speaking college students learning German, showed that for the students with low cognitive ability (either visual or verbal), recall and comprehension were worse in the condition with simultaneous presentation of verbal and pictorial information than in the condition that presented verbal or pictorial information separately. Given that the children's cognitive ability is lower than the adults' (eg, Pressley, 1982; see also Cain, Oakhill & Lemmon, 2004), the simultaneous presentation of word and picture may produce in the children the same effect as in the previously mentioned low-ability adults (ie, higher cognitive load and thus, lower recall rate and poorer comprehension). If this is so, one session of self-paced reading with a single mode presentation (either visual or verbal) should be more effective for vocabulary recall in short and medium term than a complex presentation mode (visual and verbal). Thus, the obtained data will have direct implications on the use of words and pictures to create valuable multimedia tools with educational aims.

## Method

*Participants and design*

The participants were 135 Spanish children (67 female, 68 male), 66 in Grade 3 (age $M = 8$ years) and 69 in Grade 4 (age $M = 9$ years). They attended three primary schools in Vizcaya and Guipúzcoa, Spain, which served middle-to-low socioeconomic population from urban zones. The schools were generally representative of primary schools in Spain. Children were exposed to the typical curriculum and medium of English instruction. Because of limits on the amount of time children could miss classroom activities, no English test was applied for selection purposes. However, each child was judged by his or her respective teacher as making normal progress in English language (ie, at grade-level expectations). All of the children were tested at the beginning of the school year.

*Materials and apparatus*

We employed PCs that presented an interactive multimedia short story designed for this purpose. The story consisted of 101 English words presented in one page. Twelve words were inside a button: donkey, drawer, penknife, hammer, bricks, ladybird, waistcoat,

bonnet, mittens, jug, tray and napkin. All of them were substantives and their frequency ranged between 1 and 18 per million (mean = 5.4) in the Kučera and Francis (1967) English frequency dictionary.

For each of these 12 words, different types of multimedia annotations were available. Verbal annotations consisted of a text translation of the word. Visual annotations consisted of a picture illustrating the word. All the pictures were taken from the standardised picture-set of Pérez and Navalón (2003). To access these annotations, children moved the mouse over the marked word and pressed the mouse button. A new page with the corresponding annotation appeared at the screen. The annotations were designed to aid the children in understanding the meaning of individual vocabulary items.

The paper and pencil materials consisted of an English vocabulary pretest and posttest, and a test of verbal and spatial ability. The vocabulary pretest consisted of 60 English words, and the posttest consisted of the 12 annotated English words (all included in the pretest and unknown to the children). Students were instructed to produce a Spanish translation of each word. The verbal and spatial ability tests were taken from the Evaluación Factorial de Aptitudes Intelectuales (Santamaría, Arribas, Pereña & Seisdedos, 2005). The internal consistency values for the spatial ability test and the verbal ability test were 0.86 and 0.83 respectively.

*Procedure*
We tested the children in their English language classes in groups of 15 per session during their normally scheduled class hour. Classes met for 45 minutes per day and the entire procedure required a 40-minute class period in each class for three non-consecutive days. On the first day, the students filled out the vocabulary pretest at their own rates. This test consisted of a list of 60 English words, in which children had to write the meaning of the ones they knew. On the second day, students fulfilled the verbal and spatial ability test. It took them 15 minutes to fill out the first one and another 15 minutes to fill out the second one. On the third day, children read the story with the multimedia programme.

Children were randomly assigned to the three groups ('word-only', 'picture-only', or 'word and picture'; group mean age $M = 8.5$). Children in the 'word-only' group, when selecting the word, received a text translation of the word in the middle of the screen next to the word written in English. Children in the 'picture-only' group, when selecting the word, saw a picture representing the meaning of the word in the middle of the screen. Children in the 'word and picture group' saw both the translation and the image representing the word in the middle of the screen.

Each child sat in front of a separate PC system in the school's computer lab. For 10 minutes, the children were given a brief demonstration of the programme with a preview on each computer. They were given instructions on how to look up a marked word, hold the mouse button down, and, when the annotation was available, how to release the mouse button. They were also given the instruction that upon selecting the word they

*Table 1: Means and standard deviations (in parenthesis) of word-recall mean percentages on the vocabulary test for groups of children across the different conditions*

| Experimental group | Immediate posttest | Delayed posttest |
|---|---|---|
| Word only | 39.73 (33.14) | 27.82 (25.32) |
| Picture only | 30.40 (23.70) | 19.26 (13.89) |
| Word and picture | 29.60 (21.19) | 18.90 (15.10) |

would see a cue that would aid them in remembering the meaning, and that they had to look up all of the marked words while reading the story because the learning task was to remember the meaning of the words. When they had finished the task, they were asked to quit the programme. They were given 20 minutes to read the story twice and after that, they had another 10 minutes to fill out the vocabulary posttest. Afterwards, the experimenter recorded the meanings that were correctly remembered by each child on the English vocabulary posttest (the 'immediate' posttest). Two weeks after the experiment, the children fulfilled the same vocabulary posttest again (the 'delayed' posttest). Eight children dropped from this second posttest because they reported ill.

## Results

The word-recall mean percentages for each group in the immediate and delayed posttest are displayed in Table 1.

An Analysis of Covariance (ANCOVA) was performed to determine the effects of the presentation mode in each group. As spatial and verbal ability are related to skilled processing of pictorial material and vocabulary learning respectively[1] (eg, see Plass *et al*, 2003), in order to obtain higher statistical power, data were submitted to an ANCOVA on the percentage of the recalled words in the vocabulary test as dependent variable, group (word-only, picture-only, word and picture) as a between-subjects factor, and verbal ability and spatial ability as covariates.[2] Separate analyses were conducted for the immediate and the delayed posttest.

---

[1] Indeed, the Pearson correlation coefficient between word-recall mean percentages and verbal ability was significant, both in the immediate test ($r = 0.37$, $p < 0.001$), and in the delayed test ($r = 0.38$, $p < 0.001$).

[2] An Analysis of Variance (ANOVA) was also performed to determine the effects of the group and ability. Data were submitted to an analysis of variance 3 (group: word, picture, word and picture) $\times$ 3 (ability: high visual and low verbal, low visual and high verbal, visual and verbal average) with recalled words as a dependent measure. The ANOVA showed a reliable effect for group both in the immediate, $F(2,133) = 3.60$, $p < 0.05$, $n^2 = 0.054$; and in the delayed posttest, $F(2\ 124) = 4.74$, $p < 0.05$, $n^2 = 0.074$; for ability only in the delayed posttest, $F(2\ 124) = 2.74$, $p = 0.068$, $n^2 = 0.042$; and $F(2\ 124) = 4.65$, $p < 0.05$, $n^2 = 0.073$, but it did not show an interaction between group and ability, $F(2\ 133) = 1.54$, $p = 0.19$; and $F(2\ 124) = 1.94$, $p = 0.10$. Nonetheless, the ANOVA results must be taken with caution, as most children had low scores (0.5 quantile or less) in verbal and visual ability, hence, the groups were not balanced.

The ANCOVA revealed a significant effect of group, both in the immediate posttest, $F(2,130) = 3.16$, $p < 0.05$, $\eta^2 = 0.046$; and in the delayed posttest, $F(2\ 122) = 4.14$, $p < 0.05$, $\eta^2 = 0.064$. This main effect of group reflects that the 'word-only' group shows a higher percentage of recalled words than the 'word and picture' group, both in the immediate and the delayed posttests, $t(1,89) = 1.76$, $p < 0.05$; and $t(1,82) = 1.93$, $p < 0.05$, respectively. Furthermore, the percentage of recalled words was higher for the 'word-only' group than for the 'picture-only' group; this difference was significant in the immediate posttest, $t(1,81) = 1.68$, $p < 0.05$; and it approached significance in the delayed posttest, $t(1,79) = 1.92$, $p = 0.057$. The performance in the 'picture-only' and the 'word and picture' groups was quite similar ($p > 0.25$ in the two posttests).

## Discussion

The main question in this study was to examine which presentation mode was more effective in a self-paced vocabulary learning multimedia programme for primary school children. Specifically, we examined the effects of three presentation modes on word-recall for previously unknown (concrete) English words: adding a word, adding a picture, or (simultaneously) adding a word and a picture. The recall rate of children in the 'word-only' group (ie, the group exposed to one stimulus through one information-processing channel: visual) was better than the recall rate of children in the 'word and picture' and the 'picture-only' groups. This difference occurred both in an immediate posttest and in a delayed posttest which took place 2 weeks later.

It may be of interest to examine why the 'word-only' group performed better than the 'picture-only' group. The picture involves a higher cognitive load than the word and, in our case, led to less effective learning—indeed, naming times of pictures are substantially higher than the naming times of words (eg, Ferrand, Segui & Grainger, 1996). We must keep in mind that adding a word allows establishing a direct connection between the verbal representation of the word in the native language and its foreign equivalent. The picture involves establishing an additional connection between the representation of the picture and the corresponding representation of the word. Thus, children who process the picture can be exposed to a higher cognitive load than those who only process the word.

With respect to the effect of adding a picture in the context of a self-paced vocabulary learning programme, the present data strongly suggest that adding the picture to the word involves extra cognitive resources, as deduced by the better performance in the 'word-only' group than in the 'word and picture' group. To process a given picture, children have to allocate more of their cognitive resources and this causes a high cognitive load. Bear in mind that the child has to translate an ambiguous pictorial representation into an unambiguous verbal representation (meaning). Furthermore, in the 'word and picture' group, the child's processing might also have been hindered by the simultaneous presentation of a word and a picture through the visual information-processing channel, resulting in an even lower performance. Interestingly, the obtained findings are consistent with the observation that, in adults with low cognitive abilities, learning may be worse when a picture is added in a multimedia programme (see Plass

*et al*, 2003). Nonetheless, although in the present vocabulary learning experiment adding a picture did not enhance vocabulary learning, this does not preclude the effectiveness of a 'word and picture' presentation mode in other learning contexts (Mayer & Anderson, 1991; Mayer, Bove, Bryman, Mars & Tapangco, 1996).

The reported findings have important implications for designers of educational programmes for children. Our results clearly suggest that the instructional material used to enhance learners' experience should consider the processes involved in each task to impose the minimum cognitive load. For example, our task required the establishment of a simple association between an orthographic representation and a representation that already existed in the mental lexicon. This work shows that this association is easily established with orthographic information only. Yet, other studies show that when high-order processes are involved, pictures, schemes and conceptual maps are very useful in self-pace learning. Concretely, when the task requires text comprehension and integration of information, pictures have been useful to interpret text and construct deep understanding (see Ainsworth, 1999).

In line with this reasoning, designers should work on a framework that considers three main points. First, definition of the cognitive processes required by the task (ie, association and recall as in our case, or higher order processes such as comprehension and synthesis). Second, identification of the functions of the elements provided to trigger the domain specific knowledge (ie, do they help to complement, categorise or integrate information). Third, organisation of the information in a temporal (ie, frame-to-frame learning) or spatial configuration (conceptual maps, schemes) according to each stage of the learning process (Chandler & Sweller, 1991; Morrison & Anglin, 2005).

In sum, this experiment provides empirical evidence that shows that, in a second-language vocabulary learning multimedia programme for children, presenting a word is more effective than presenting the word together with the picture or only the picture. This result casts some doubts on the (direct) efficiency of adding pictorial information in educational multimedia programmes for children, in which learning is self-paced and no direct coaching is provided. The growing demand of these programmes in education and the difficulty in generalising the empirical results to different ages and learning contexts, make it necessary to test any multimedia programme before implementing it. Furthermore, the new possibilities of e-learning with hypermedia formats and the individualisation of the learning process calls for integrated educational and technological research as a means to assure effective learning results. Future research should focus on this point: how to suit multimedia technology development to children's learning possibilities.

## Acknowledgements

programme, as well as Edurne Goikoetxea (Universidad de Deusto) and Manuel Perea (Universitat de València) for their comments on the paper.

## References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, *33*, 131–152.

Ávila, E. & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary language learning. *British Journal of Educational Technology*, *46*, 379–395.

Baddeley, A. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. (1997). *Human memory. Theory and practice* (ed. rev.) East Sussex, England: Psychology Press.

Brünken, R., Plass, J. L. & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*, 53–61.

Cain, K., Oakhill, J. & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: the influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*, 671–681.

Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332.

Duquette, L., Renié, D. & Laurier, M. (1998). The evaluation of vocabulary acquisition when learning French as a second language in a multimedia environment. *Computer Assisted Language Learning*, *11*, 3–34.

Ferrand, L., Segui, J. & Grainger, J. (1996). Masked priming of word and picture naming: the role of syllabic units. *Journal of Memory and Language*, *35*, 708–723.

Kučera, H. & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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

Mayer, R. E. & Anderson, R. B. (1991). Animations need narrations: an experimental test of a dual coding hypothesis. *Journal of Educational Psychology*, *83*, 484–490.

Mayer, R. E. & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, *14*, 87–99.

Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52.

Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*, 389–401.

Mayer, R. E., Bove, W., Bryman, A., Mars, R. & Tapangco, L. (1996). When less is more: meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, *88*, 64–73.

Mayer, R. E., Heiser, J. & Lonn, S. (2001). Cognitive constraints on multimedia learning: when presenting more material results in less understanding. *Journal of Educational Psychology*, *93*, 187–198.

Mayer, R. E., Plass, J. L., Chun, D. M. & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, *90*, 25–36.

Moreno, R. & Mayer, R. E. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, *14*, 87–99.

Morrison, G. R. & Anglin, G. J. (2005). Research on cognitive load theory: application to e-learning. *Educational Technology Research and Development*, *53*, 94–104.

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart & Winston.

Plass, J. L., Chun, D., Mayer, R. E. & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*, 221–243.

Pressley, M. (1982). Elaboration and memory development. *Child Development*, *53*, 296–309.

Pérez, M. A. & Navalón, C. (2003). Spanish norms for 290 new pictures: Name agreement, image agreement, familiarity, visual complexity and image variability. *Psicológica*, *24*, 215–241.

Riding, R. & Grimley, M. (1999). Cognitive style, gender and learning from multimedia materials in 11-year-old children. *British Journal of Educational Technology*, *30*, 43–56.

Santamaría, P., Arribas, D. Pereña, J. & Seisdedos, N. (2005). *Evaluación Factorial de Aptitudes Intelectuales*. Madrid, Spain: TEA Ed.

Tindall-Ford, S., Chandler, P. & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: applied*, *3*, 257–287.

Wild, M. & Quinn, C. (1998). Implications of educational theory for the design of instructional multimedia. *British Journal of Educational Technology*, *29*, 73–82.