

Visual and Spoken Texts in MCALL Courseware: The Effects of Text Modalities on the Vocabulary Retention of EFL Learners

Dr. Masoud Khalili Sabet (Corresponding author)

Faculty of Literature and Humanities, Guilan University

PO Box 41635-3988, Rasht, Iran

Tel: 98-0131-772-9535 E-mail: sabetma2002@yahoo.com

Hamed Babaie Shalmani

Young Researchers' Club, Faculty of Letters and Humanities

Islamic Azad University-Rasht Branch

PO Box 129, Rasht, Iran

Tel: 98-0131-322-9424 E-mail: babaie@iaurasht.ac.ir

Abstract

The present study sought to explore the effects of Multimedia Computer-Assisted Language Learning (MCALL) programs drawing on two different text modalities on the vocabulary retention of Iranian EFL learners. The two groups under study received treatment on vocabulary items under two multimedia conditions: The first group received treatment on the vocabulary items using a multimedia environment comprising streaming video and visual texts, and the second group received treatment on the same items through a similar environment drawing on streaming video and spoken texts. After the experiment, the two groups took an immediate post-test and a delayed post-test. The study revealed that those students who received treatment on the items through visual texts and video outperformed the ones who received treatment on the same items through spoken texts and video. This appears not to corroborate the view that the modularity of the working memory always results in a more efficient learning.

Keywords: MCALL courseware, Visual text, Spoken text, Modularity, Vocabulary retention

1. Background

Through the years, a good many studies have shown the negative impact of the working memory limitations in information processing on performance on cognitive tasks (Norman & Bobrow, 1975; Just & Carpenter, 1992; Anderson, Reder, & Lebiere, 1996). The adverse effect of such limitations on learning is quite palpable in multimedia environments where learners are to integrate different information elements, such as streaming video, pictures, texts, etc. in the instruction. Here, a mental representation of one element has to be kept active in the working memory while searching for the corresponding element. Particularly, in the absence of prior knowledge and schemata to guide the search process, cognitive overload is a serious menace to learning (Sweller, Van Merriënboer, & Paas, 1998).

Another property of the working memory germane to multimedia learning is the existence of separate memory modules for different input modalities. There is a consensus that the modularity of the working memory capacity might help minimize cognitive overload that comes about when different pieces of information are processed within a single module. According to Baddeley's (1997) Multiple-Components Theory, the working memory comprises a "central executive" and two slave systems, the "visuospatial sketchpad" and the "phonological loop". While the former is dedicated to processing visual and spatial information, the latter is allotted to acoustic and verbal information. The central executive serves as an intermediate device that connects two or more mental representations of information that are encoded in separate memory modules. It is contended that when information is presented in two sensory modalities rather than one, the working memory total capacity is utilized more efficiently, as both slave systems are addressed concurrently. Consequently, relative to the available resources, the cognitive load of multimedia instruction is reduced.

Sweller (1999) argued that available cognitive resources to learners should be directed to the learning process itself and not to the irrelevant features of instructional materials. In his well-renowned Cognitive Load Theory, he differentiated between "intrinsic" and "extrinsic" loads of instruction, contending that while the former refers to the intricacy of instruction and the learning tasks, the latter applies to the way information is presented to

learners. In other words, intrinsic load depends highly on the learning content and the learners' expertise while extrinsic load is caused by the format of instruction. Accordingly, since in multimedia learning the necessary mental integration of information leads to a high cognitive load, instruction should be given in such a way that they keep extrinsic load as low as possible.

Mayer (2001) proposed the Generative Theory of multimedia learning based on the result of an experiment focusing on the use of multimedia instructional messages. The study explored how lightning storms develop, how car braking systems work, and how bicycle tire pumps work. In his theory, two main assumptions were made on the way people process these kinds of instructions. First, learners engage in the active processing of the instructional material. Therefore, a coherent mental representation of information is created as learners select information, organize it and integrate it with existing knowledge structures. Second, humans have separate processing channels for aural and visual information. Mayer (2001) related this dual-channel assumption to the phonological loop and the visuospatial sketchpad of Baddeley's (1997) working memory model, thus implying that visual words and spoken words are initially processed in different channels, but are subsequently represented in the same verbal system. This helps learners utilize memory sources optimally, and cognitive load decreases consequently.

The two properties of the working memory, limited capacity and modularity, have accordingly intrigued many enthusiasts to explore the likely effects these properties have on learning. Since picture and text are readily integrated in multimedia environments, one can easily challenge the aforementioned theories by authoring multimedia courseware incorporating visuals and different text modalities. One assumption is that when visuals, such as pictures, streaming video, etc. and visual texts are presented to learners simultaneously, the visual module or the visuospatial sketchpad in Baddeley's term is overload, resulting in a less efficient learning. The cognitive overload, however, can be minimized by presenting texts as narrations so that both visual and auditory channels are engaged.

Studies by Mousavi, Low and Sweller (1995) and Jeung, Chandler and Sweller (1997) revealed that students receiving multimedia instruction with spoken text spent less time on subsequent problem-solving tasks as opposed to those receiving visual-text instructions. Furthermore, in studies by Kalyuga, Chandler and Sweller (2000), students receiving spoken-text instruction had higher scores on various retention and transfer tests, and in experiments by Tindall-Ford, Chandler and Sweller (1997) students not only obtained higher test scores but also reported less mental effort during the instruction. On the whole, these results strongly underpinned the design guideline for the use of spoken text in multimedia instruction.

2. Purpose of the Study

Inspired by the modularity theories, this research explored the effects of two multimedia programs drawing on two text modalities, i.e., visual and spoken texts, on the vocabulary retention of EFL learners. The study aimed to ascertain whether spoken texts coupled with streaming video would offer any superiority over a combination of visual texts and streaming video in helping learners better retain the vocabulary items being introduced.

3. Research Question and Hypothesis

This study sought to find an empirically justified answer to the following question:

Is there any significant difference between the use of the multimedia program drawing on visual texts and visuals and the multimedia program using spoken texts and visuals in helping EFL learners better retain vocabulary items?

A null hypothesis is as follows:

There is no statistically significant difference in the use of the multimedia program using visual texts and visuals and the one using spoken texts and visuals in helping EFL Learners better retain vocabulary items.

4. Methodology

4.1 Participants

The subjects involved 180 students who were majoring in English translation at the Islamic Azad University-Rasht Branch, Iran. They were identified as intermediate-level students based on their overall band score on an IELTS test of proficiency and were randomly assigned to two equivalent groups of subjects comprising male and female participants.

4.2 Instruments

The instruments in this study fell into two categories: There were two types of multimedia courseware applying the treatment, and a recognition vocabulary test that served as both the pre- and the post-tests. The multimedia

programs, developed by one of the researchers, introduced 50 vocabulary items through either visual or spoken texts. Both programs also used video segments to help subjects better surmise the meanings of the words being introduced.

The vocabulary test administered at the beginning and the end of the experiment was used to measure the subjects' prior knowledge of the words being introduced, as well as their degree of learning through the two types of treatments.

4.3 Procedure

At the beginning of the experiment, a proficiency test of receptive skills based on the UCLES IELTS examination papers was administered to 400 sophomores majoring in English translation at the Islamic Azad University-Rasht Branch, Iran. To standardize this eighty-item test, SIMSTAT, an item analyzer was used. The result of the analysis revealed that all items had desirable IF indexes ranging from 0.37 to 0.73 and ID indexes well above 0.40. Using Cronbach's Alpha, the reliability index turned out to be 0.80.

Next, an exploratory factor analysis was used to help the researchers determine the number of factors involved, as well as the extent to which the items on the test modules correlated with the underlying constructs. The result of the analysis revealed that only one factor was involved, as only one component (scree) had an eigenvalue well beyond unity (Figures 1 & 2 below).

Once the construct validity of the test was established, 180 participants who got five on the IELTS test were identified as intermediate-level students following the rating scheme developed by the Local Examination Syndicate at Cambridge University. According to the scheme, all candidates who obtain an overall band score of five are identified as "modest users" or those who are at the intermediate-level of language proficiency.

The participants were then randomly assigned to three equivalent groups of subjects: a pilot group and two experimental groups. To randomize the subjects, a randomizer called SuperCool Random Number Generator was used. Once the subjects were assigned a number from 1 to 180, the program randomized them by generating random sets of numbers from within the range. Afterwards, the first 60 subjects whose numbers fell under the first column were put in the pilot group and the second and the third 60 subjects were put in the experimental groups. The subjects comprised mixed groups of males and females.

The next step involved designing a recognition vocabulary test in the multiple-choice format that would serve as both the pre- and the post-tests. The test comprised 60 concrete vocabulary items that fell under two general themes: animals and tools. To standardize the pre-test, it was first administered to the pilot group under study. Each item correctly answered would receive a score of one mark, and the total score possible would be 60. The item analyzer utility revealed that 10 items malfunctioned and these were excluded from the test. The subjects' papers in this group were then re-scored and the reliability index was computed. It turned out to be 0.74, which was significant. Next, the construct validity of the test was established through a factor analysis that showed that items highly correlated with the latent construct, i.e., the vocabulary recognition ability (Figure 3 below).

In the next step, the vocabulary test was administered to the experimental groups under study. The purpose of pre-testing was twofold: to ascertain the participants' prior knowledge of the words to be introduced and to determine the homogeneity of the groups at the beginning of the experiment. The pre-test result appeared under table 1 below.

As shown in the table, the subjects delivered a poor performance on the test. This implied that they needed to receive treatment on the vocabulary items.

Moreover, the t-test statistic (table 2) revealed that the two groups were homogeneous concerning the vocabulary items being introduced ($p > 0.05$).

Once the homogeneity of the groups was determined, the two groups received treatment on the vocabulary items through two Multimedia Computer-Assisted Language Learning (MCALL) courseware authored by one of the researchers. The first experimental group received treatment on the vocabulary items using a multimedia environment comprising streaming video and visual texts, and the second group received treatment on the same items through a similar environment drawing on streaming video and spoken text. The two multimedia conditions differed in that while in the first group the subjects could see the texts appearing on the screen, in the second group the participants were required to wear headsets and listen to the passages. There was no visual text and the students could only hear the researcher's voice playing in the background.

The texts through which the vocabulary items were introduced were all excerpts taken from Microsoft Encarta providing a meaningful context for vocabulary learning. For instance, as far as the animal theme was concerned, the passages provided information as to the physical characteristics of the very animal, its diet, habitat, etc.

Likewise, in order to introduce terms referring to tools, the passages gave information on the physical shape of the tool, e.g., what a “chisel” was like, and where it was normally used. The programs were designed in such a way that they would automatically run once inserted in the CD-ROM drivers and introduce 50 vocabulary items within a span of 50 minutes.

After the experiment, the two groups took an immediate post-test and a delayed post-test two weeks later. The results of the post-tests appeared under tables 3 and 4 below. As shown in the tables, the experimental groups obtained a higher mean on both tests in comparison to the pre-test scores. This implies that both kinds of treatments significantly expanded the subjects’ vocabulary repertoire.

In a similar vein, tables 5 and 6 show the results of the Levene’s Test of equality of variances and t-test for the immediate and delayed post-tests respectively. A glimpse at the results reveals that there was no significant difference between the mean scores on the immediate post-test ($p > 0.05$), but a significant difference was found between the means on the delayed post-test ($p < 0.05$). Hence, the present results favor the use of visual texts in multimedia environments, as the subjects receiving treatment through such texts could more readily remember the vocabulary items on the delayed post-test as compared with those who received treatment on the same items through spoken texts.

5. Results and Discussion

The purpose of this study was to find an empirically justified answer to the following question:

Is there any significant difference between the use of the multimedia program drawing on visual texts and visuals and the multimedia program using spoken texts and visuals in helping EFL learners better retain vocabulary items?

The answer is “yes”, as the experiment revealed that the difference between the mean scores was statistically significant on the delayed post-test albeit no major difference was found between the means on the immediate post-test. The null hypothesis formulated a priori was accordingly rejected assuming no significant difference in the use of the two types of treatments. The fact that both groups performed equally well on the immediate post-test appears not to confirm the previous research findings (Norman & Bobrow, 1975; Just & Carpenter, 1992; Anderson et al., 1996) on the limited capacity of the working memory that corroborated the negative impact of such limitations on learning. Furthermore, such results seem to stand in opposition to Baddeley (1997) and Mayer’s (2001) view postulating that the modularity of working memory necessarily yields a more efficient learning, which results from the optimal utilization of memory resources. The study showed that although for the first experimental group the information was only presented visually, the cognitive overload did not come about as expected.

In previous studies, the multimedia instruction primarily focused on teaching subjects from technical domains, such as geometry (Mousavi et al., 1995; Jeung et al., 1997), scientific explanations of how lightning develops (Moreno & Mayer, 1999), reading a technical diagram (Kalyuga et al., 2000), and electrical engineering (Tindall-Ford et al., 1997) where the format of instructions played a key role in how well learners would perform on the tests. Notwithstanding, as far as vocabulary learning is concerned, this study implies that the format of instruction is not of great significance and drawing on a single working memory module might not always lead to Sweller’s (1999) extrinsic load, resulting in a less efficient learning. One rationale is that since vocabulary teaching in this experiment centered on introducing general vocabulary to the subjects, the visual memory was not overloaded, as the explanations given on the vocabulary items were easy to process and hence might not have consumed the memory resources excessively. If this is the case, then it can be argued that the technicality of information might correlate with the degree to which it places a high demand on the memory modules. Further studies, however, are required to corroborate this view. Yet, another justification for such contradictory results might stem from the assumption that the format of instruction does not necessarily correlate with cognitive overload irrespective of the technicality of information. In other words, whether or not the piece of information is technical, the way through which it is presented to learners may not serve as the causal variable in determining how well it is processed within the modules. It might, then, be intriguing to replicate the current study where the focus of instruction would be teaching discipline-specific vocabulary (vocabulary in different disciplines, including medicine, physics, etc.) through visuals only and to explore whether the format of instruction would truly matter.

Additionally, the mean scores on the delayed post-test further corroborate that the spoken text does not necessarily offer any superiority over visual text. The experimental group receiving treatment on the vocabulary items through visuals only outperformed the one who received treatment through a combination of visuals and spoken texts. This shows that the students in the first group could more readily remember the vocabulary items at the examination time. One rough justification, however, is that when information is presented in a single format,

the elaboration and rehearsal processes occur more effectively as opposed to when it is presented through various modalities, addressing different memory modules. Hence, the piece of information is more likely to be coded effectively in long-term memory, leading to a more convenient retrieval of information on learners' part.

Another possibility is that visual texts, like other visuals, might more effectively focus learners' attention on the subject matter being introduced. Convictions are strong that visuals (pictures or streaming video) have the potential to sustain learners' attention during the learning process (Al-Seghayer, 2001). As a result, learners' sustained attention during information processing might then lead to a more effective coding of information.

The present study thus favored the use of visual texts in multimedia environments as the best mode of introducing vocabulary that might significantly aid in the memorization and retrieval of words. Notwithstanding, due to a paucity of research on the role of text modality in multimedia environments, it is rather difficult to refute extant theories of multimedia learning through a single study. Further experiments are required to substantiate such a claim.

6. Conclusion and Pedagogical Implications

This study showed that visual texts might prove more effective in the memorization and retrieval of vocabulary when combined with streaming video in multimedia environments. Accordingly, instruction should center on the use of such texts where fragments of visual texts might persist in learners' visual memory, thus making vocabulary learning a more memorable experience. The use of such texts, together with streaming video, might indeed make information encoded in the visual memory (here the context through which vocabulary items are introduced) more elaborate and hence more memorable. Teachers or teachers as designers can, then, author customizable courseware where vocabulary of interest can be introduced through multimedia environments integrating visuals to help maximizing vocabulary learning efficiency.

7. Suggestions for Further Research

This study focused on the vocabulary retention of intermediate-level learners in an EFL contexts. Further experiments should investigate the vocabulary retention among learners at different proficiency levels. This experiment was a one-shot study. It is not clear whether visual texts always appear more effective than spoken texts. Therefore, follow-up studies are to be longitudinal so as to further substantiate such a claim. Moreover, the rather paradoxical results of the current study can be accounted for by the fact that the process of learning was somewhat system-based or system-controlled. The subjects in this study had no control over the instruction process, as the MCALL programs would automatically introduce the vocabulary items. Studies showed that there might be a difference between system-controlled and learner-paced learning, where learners themselves control the pace of instruction (Tabbers, 2002). Accordingly, future experiments can be learner-controlled so as to help researchers determine whether or not the mode of learning might have any impact on learning through different text modalities. Furthermore, the participants in this research comprised mixed groups of males and females. Hence, future studies should explore whether "gender" too, as a moderating variable, may affect the way male and female students learn vocabulary through visual and spoken texts in multimedia environments.

References

- Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), 202-232. Retrieved February 15, 2010, from <http://llt.msu.edu/vol5num1/alseghayer/default.pdf>.
- Anderson, J. R., Reder, L. M. & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, 30, 221-256.
- Baddeley, A. D. (1997). *Human memory: theory and practice* (revised ed.). Hove, United Kingdom: Psychology Press.
- Jeung, H., Chandler, P. & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, 17, 329-343.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kalyuga, S., Chandler, P. & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126-136.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
- 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.

Moreno, R. & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91, 358-368.

Mousavi, S. Y., Low, R. & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319-334.

Norman, D. A. & Bobrow, D. C. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7, 44-64.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

Sweller, J., Van Merriënboer, J. J. G. & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.

Tabber, H. K. (2002). *The modality of text in multimedia instructions*. Heerlen: Educational Technology Expertise Centre.

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

Table 1. Scores on the pre-test

	Experimental Groups	N	Mean	Std. Deviation	Std. Error Mean
Pre-test Scores	A	60	4.5333	2.30303	.29732
	B	60	5.1000	2.56905	.33166

Table 2. T-test and the Levene's Test of equality of variances

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower
Pre-test Scores	Equal variances assumed	1.524	.219	-1.272	118	.206	-.56667	.44542	-1.44872	.31539
	Equal variances not assumed			-1.272	116.618	.206	-.56667	.44542	-1.44883	.31550

Table 3. Scores on the immediate post-test

	Experimental Groups	N	Mean	Std. Deviation	Std. Error Mean
Immediate Post-test Scores	A	60	46.2667	5.40830	.69821
	B	60	47.2167	3.83601	.49523

Table 4. Scores on the delayed post-test

	Experimental Groups	N	Mean	Std. Deviation	Std. Error Mean
Delayed Post-test Scores	A	60	41.8333	10.01722	1.29322
	B	60	26.4667	14.78027	1.90813

Table 5. T-test and the Levene’s Test results for the immediate post-test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower
Immediate Post-test Scores	Equal variances assumed	3.346	.070	1.110	118	.269	.95000	.85601	-.74512	2.64512
	Equal variances not assumed			1.110	106.374	.270	.95000	.85601	-.74704	2.64704

Table 6. T-test and the Levene’s Test results for the delayed post-test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower
Delayed Post-test Scores	Equal variances assumed	21.950	.000	6.666	118	.000	15.36667	2.30507	10.80200	19.93133
	Equal variances not assumed			6.666	103.758	.000	15.36667	2.30507	10.79550	19.93783

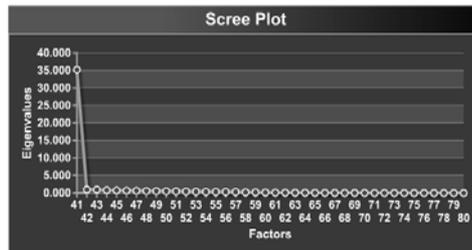


Figure 1. Listening module

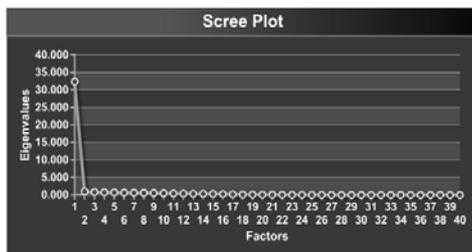


Figure 2. Reading module

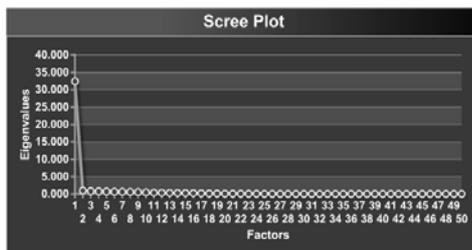


Figure 3. Vocabulary test