Lenguas Modernas 30 (2004-2005), 191-200
Universidad de Chile

THE EFFECTS OF SELF-GENERATED VERSUS SUPPLIED INTERACTIVE IMAGES ON THE LONG-TERM RETENTION OF EFL VOCABULARY BY COLLEGE STUDENTS*

MANUEL COLÓN
Universidad de Oriente, Venezuela
MÁXIMO RODRÍGUEZ
Universidad de Los Andes, Venezuela

The present study, using a between-participants factorial ANOVA design, investigated the extent to which self-generated interactive images (standard keyword) were more effective than researcher-supplied images (keyword + interaction) in the long-term retention of EFL vocabulary by college students. It also assessed the extent to which these two versions of the keyword method were more effective than the rote method. The results indicated that the effectiveness of self-generated images was statistically comparable to that of researcher-developed ones. Additionally, the results indicated that only the standard version of the keyword method produced significantly superior long-term recall to the rote method.

KEY WORDS: long-term retention of EFL vocabulary; keyword method; self-generated images; researcher-supplied images

In the field of foreign language (FL) teaching, a renewed interest in the development of vocabulary has been observed the last few years. However, consensus has not been reached as to the most effective way to teach FL vocabulary. One of the most effective methods seems to be the keyword method (Ellis 1995, Stahl and Fairbanks 1986).

The keyword method was developed by Atkinson and his collaborators (Atkinson 1975, Atkinson and Raugh 1975, Raugh and Atkinson 1975, Raugh, Schupbach and Atkinson 1977) to facilitate recall of FL definitions. It is a two-stage method. In

* The findings reported here are based on the first author's master's thesis in ESL, which was completed under the guidance of the second author and submitted to La Universidad de Oriente, Cumaná, in January 2002. The authors are grateful to Mark Sadoski for his insightful comments on an earlier version of this manuscript.

Correspondence concerning this article may be addressed to Máximo Rodríguez, Departamento de Lenguas Modernas, Universidad de Los Andes, Trujillo, Venezuela. Internet: <maxrr52@hotmail.com>

the first stage, the FL word is recoded into an acoustically similar and concrete word or phrase in the learner's native language. This word or phrase is known as the keyword. In the second stage, a mental image is created in which the referent of the keyword is interactively linked to the meaning of the FL word. For example, a native speaker of Spanish could learn that the English word *empennage* means *airplane tail*, by using *empanada* (meat pie) as the keyword and creating a mental image of a meat pie hanging on an airplane tail. It is assumed that presentation of the to-be-learned word (here, *empennage*) will evoke the keyword (here, *empanada*), which in turn will evoke the interactive image (here, the meat pie hanging on the airplane tail) that includes the FL definition (here, *airplane tail*).

The keyword method has proven more effective on the immediate recall of FL definitions than a number of control methods, including rote, context, and a no-strategy control in which participants are encouraged to use any method they wish. This superiority has been shown in experiments using college students (e.g., Atkinson and Raugh 1975, Godley, Fournet and Estes 1987, Kasper and Glass 1988, Raugh and Atkinson 1975), children (e.g., Avila and Sadoski 1996, Elhelou 1994, Merry 1980, Pressley 1977, Pressley, Levin and Miller 1981), and elderly (Gruneberg and Pascoe 1996). However, the long-term benefits of the keyword method have been recently questioned by Wang, Thomas, and their collaborators (Thomas and Wang 1996, Wang and Thomas 1995, Wang, Thomas and Ouellette 1992, Wang, Thomas, Inzana and Primicerio 1993). In a series of experiments using between-participants designs, these researchers have shown that the keyword method exhibited a greater forgetting rate than either rote or context and that the initial advantage of the keyword either disappeared or reversed after a delay of a few days.

The success of the keyword method depends to a great extent upon the learning of both the keywords and the interactive images (Atkinson 1975). However, recent research (e.g., Carney and Levin 1998, Thomas and Wang 1996, Experiment 1, Wang et al. 1992, Experiments 1-3) suggests that recall of the keywords does not guarantee successful recall of FL definitions.

As for the interactive images, concern has been raised regarding the efficacy of self-generated versus researcher-supplied images. Atkinson (1975), based upon the results of some pilot studies, argued that it was more beneficial for the learners to generate their own interactive images than using those supplied by the researcher. The argument seems to have been taken in good faith by most keyword researchers because the issue has been scarcely investigated. A review of the literature indicated that with the exception of the studies by Pressley and Levin (1978) and Carney and Levin (1998, Experiment 4) the issue has not been investigated directly. Furthermore, it is interesting to point out that the purpose in the Pressley and Levin's study was to examine two developmental constraints associated with children's use of the keyword method rather than to compare the effects of self-generated and imposed interactive images. Additionally, it is interesting to note that only immediate retention was assessed.

Pressley and Levin (1978) asked sixth and second graders to learn a list of 18 Spanish words and a randomly selected sample of 12 of these words, respectively.

The participants were randomly assigned to one of five experimental conditions. Of these, three were imagery-link variations: Interacting Picture, where learners were shown colored lines drawings of the keyword and the Spanish translation referents interacting; Separate Picture, where learners were shown separate drawings of the two referents; and Separate Word, where the keyword and its Spanish translation were displayed in print. In the two separate conditions, learners were given instructions to generate interactive images involving the keyword and the Spanish translation. The other two conditions were regarded as controls: Picture Control, where the Spanish translation and its picture were shown and Word Control, where learners were shown just the Spanish translation. The results indicated that for second graders it was more effective to use the pictorial interactive images supplied by the researchers than to generate their own, while for sixth graders it did not make any significant difference whether they used their own self-generated images or those supplied by the researchers.

Carney and Levin (1998, Experiment 4) was one of five experiments conducted to investigate various aspects of the contention set forth by Wang, Thomas, and their associates (Thomas and Wang 1996, Wang and Thomas 1995, Wang et al. 1992, Wang et al. 1993) that the delayed recall of the keyword method is not impressive, especially in comparison to that of the rote method and in the absence of an immediate test. Specifically, Experiment 4 examined the effect of two versions of the keyword method: standard keyword (self-generated interactive images) and keyword + interaction (researcher-supplied images) relative to that of the rote method on the long-term retention of 24 Tagalog nouns (the official language of the Philippines). College students were given two passes at 15-s intervals. Three sets of eight items were randomly selected from the list of 24 words. One set of eight words was tested immediately, after a 2-day delay, and after a 5-day delay. A second set was tested after a 2-day delay, and after a 5-day delay. Finally, the third set was tested only after a 5-day delay. The analyses yielded statistically nonsignificant difference between the two versions of the keyword method. Moreover, they indicated that both versions of the keyword method were superior to the rote method on each testing occasion, a finding that ran counter to those of Wang, Thomas, and their colleagues.

The present study sought to expand the ecological validity of Carney and Levin's (1998, Experiment 4) findings by using actual EFL instructors to teach the stimuli to their own students in their regular classrooms. Specifically, the present study addressed two research questions:

1) To what extent are self-generated interactive images more effective than researcher-supplied ones in the long-term retention of EFL vocabulary by college students?
2) To what extent do the two versions of the keyword method produce superior long-term recall to the rote method?

Method

Design and participants

In the present study a 2 (retention interval: immediate, delayed) x 3 (treatment: standard keyword, keyword + interaction, rote) between-participants design was used. One hundred thirty-six first-year undergraduates from four EFL classes at a northern university in Venezuela were randomly assigned to one of three learning conditions: standard keyword, keyword + interaction, or rote. Recall was assessed either immediately after the treatment or after a one-week delay. Three EFL instructors also participated in the study. Each instructor taught the experimental words in the three learning conditions and administered the cued-recall test.

Materials

A modified Rodríguez and Sadoski (2000) list of 15 obscure and low frequency concrete English nouns was used in the present study. Two of their experimental words: *costard* and *hurdle* were replaced by their two practice words: *empennage* and *poteen*. The former was replaced because its definition was unknown to the participants, whereas the latter because it did not have an obvious keyword. Additionally, four new words: *lock*, *cart*, *pot*, and *tap* were included as practice words. The words were presented on transparency on an overhead projector. Different transparencies were prepared for each treatment condition. For the rote condition, each transparency included the list of English words and their Spanish definitions. For the keyword conditions, in addition to the Spanish definitions, each transparency included the keywords. In the keyword + interaction condition, the researcher-developed interactive images were verbally given to the participants (see Appendix). A total of 16 transparencies (8 for the training phase and 8 for the learning phase) was used (see Procedure below). A cued-recall test was administered to all the students. It consisted of a list of 15 English words with blank spaces provided to write the Spanish definitions.

Procedure

Before the actual experiment began, the instructors were trained in the keyword method and made familiar with the procedures of the experiment. They received detailed oral and written descriptions of the methods and their respective procedures, using some of the practice words. Once their mastery was evident, as reflected by their teaching of one of the to-be-learned words to the researchers, the experiment was launched.

After students in each class were randomly assigned to (a) one of three learning conditions: standard keyword, keyword + interaction, and rote, and (b) to either an immediate or delayed recall condition, students in each learning condition met with the instructors in their classroom at their regular class time. There, the instructors described the appropriate procedures for the method and used the four practice words

to help students become familiar with it. To ensure comprehension, instructions were given in Spanish.

The practice words were presented one at a time. The total presentation time was held constant across the learning conditions. Participants in the rote condition were paced through four randomized lists of the practice words at a rate of 15 seconds per word, while those in both keyword conditions were paced through two randomized lists of words and their corresponding keywords at a rate of 30-second intervals. Immediately after the practice session, recall of the four practice items was informally checked out. In the learning phase, participants were asked to learn the 15 experimental words following the same procedures used during the practice session. They were not forewarned that a test would follow. As in the practice session, participants in the rote condition were paced through four randomized lists of the experimental words at a rate of 15-second intervals, while those in the keyword conditions were paced through two randomized lists of the words and their respective keywords at a rate of 30 seconds per word. Immediately after the presentation of the words, participants in the immediate retention condition were administered a 2-minute mental exercise in which they were asked to find the differences between two drawings. Then they were administered the cued-recall test. Five minutes were allotted to complete the test. Participants in the delayed condition were administered the cued-recall test a week later and therefore were not administered the distractor task.

Results

The cued recall data were submitted to a 2 (retention interval: immediate and a 1-week delay) x 3 (treatment: rote, standard keyword, and keyword + interaction) between-participants ANOVA, using the general linear model (GLM) procedure. Means, standard deviations, and group ns for the cued recall data are shown in Table 1. Statistically significant main effects were found for treatment, $F(2, 130) = 6.70$, $p < .002$, and retention interval, $F(1, 130) = 23.86$, $p < .001$. No statistically significant effect was found for their interaction, $p > .35$.

The main effect for treatment indicated that the effect of the three learning methods on the participants' recall was significantly different. The main effect for retention interval indicated that performance on immediate cued recall was superior to that on delayed cued recall. The statistically nonsignificant interaction indicated that the treatment effects on the participants did not differ under the two retention interval conditions. Inspection of the means suggests the superiority of the standard keyword method over the other methods. However, these results were compromised because the test of homogeneity of variance between groups proved to be statistically significant, $F(5, 130) = 8.90$, $p < .0001$.

To address this statistical problem, a second analysis was performed after randomly deleting two to four cases from five of the six experimental conditions until equal ns were achieved, since ANOVA has been shown to be robust for equal sample sizes (Glass and Hopkins 1996: 524). Given the few cases deleted, it was expected that the analysis with the restricted data set would yield similar results.

Analysis with the restricted data set

As expected, the pattern of results of the restricted data set was identical to that of the original data set. Means, standard deviations, and group ns are shown in Table 2. Statistically significant main effects were found for treatment, $F(2, 114) = 5.70$, $p < .004$, and retention interval, $F(1, 114) = 21.59$, $p < .0001$. The treatment by retention interval interaction was not statistically significant, $F < 1$.

Follow-up tests were performed separately for immediate and delayed recall using the Scheffé method (Huck and Cormier 1996). For immediate recall, none of the comparisons were statistically significant. However, for delayed recall, the standard keyword method was superior to rote, $p < .01$. No other comparisons were statistically significant. These findings are illustrated in Figure 1.

The effect sizes for the difference between the mean of the standard keyword method and that of the rote method for both immediate and delayed recall confirmed the superiority of the standard keyword method. The computed effect sizes were .66 and .96 standard deviation units for immediate and delayed recall, respectively. These effect sizes ranged from medium to large (Cohen 1988).

Discussion

The major purpose of the present study was to compare the effects of supplied versus self-generated interactive images on the long-term retention of English vocabulary by college EFL students. Specifically, two research questions were addressed in the study. The first research question assessed the extent to which self-generated interactive images (standard keyword) produced superior recall to supplied interactive images (keyword + interaction), while the second examined the extent to which the two versions of the keyword method were more effective than the rote method.

The results indicated that self-generated interactive images were statistically as effective as researcher-supplied ones in promoting long-term recall of FL definitions. This finding concurs with those of Carney and Levin (1998, Experiment 4), but it is at odds with Atkinson's (1975) prediction. Carney and Levin reported that in all of their measures the performance of students in the standard keyword and keyword + interaction conditions was strikingly similar. Indeed, in the between-participants part of their design (the part of their design that is comparable to ours), the largest mean recall difference between the two versions of the keyword method on any single measure did not exceed 3%, whereas in the present study it reached 12% on the delayed test. However, the numerous methodological differences between their experiment and ours may account for such inconsistency.

Furthermore, the results indicated that the standard keyword method was significantly superior to the rote method after a one-week delay as attested by the statistical significance of their means difference and the obtained effect size of nearly one standard deviation unit. However, the performance of the students in the keyword + interaction condition was statistically comparable to that of those in the rote condition. These findings only partly support those of Carney and Levin (1998,

Experiment 4) who found that the mean recall of the learners in the two keyword conditions was statistically significantly superior to that of rote learners on each testing occasion, even on the third occasion following a 5-day delay.

The finding that the standard keyword method was significantly superior to rote in delayed recall is inconsistent with Wang et al. (1992) and Thomas and Wang's (1996) claim that in between-participants designs, the keyword method fails to produce superior long-term retention recall to other methods (usually rote). Furthermore, the statistically nonsignificant treatment by retention interval interaction indicated that the forgetting rate in the keyword learning conditions was comparable to that in the rote learning condition. This finding is contrary to that of Wang et al. (1992) and Thomas and Wang's (1996) claim that in between-participants designs, the keyword method exhibits greater forgetting rate than rote. Indeed, inspection of the data as reflected in Figure 1 indicated that the forgetting rate of the rote group was almost twice as much as that of the standard keyword group (18.33% and 9.33%, respectively). Mean recall difference of the standard keyword learners over that of the rote learners increased from about 8% to about 17% over the one-week interval. This finding is even more remarkable given the fact that the keyword learners had only half the study trials of their rote counterparts.

In sum, the results of the present study, taken together with those of Pressley and Levin (1978) and Carney and Levin (1998, Experiment 4) suggest that older children and adults might use the keyword method effectively either generating their own interactive images or using those supplied to them. Nevertheless, the results suggest that relative to the rote method, only self-generated interactive images produce superior long-term recall. Finally, the results suggest that for the keyword method to be effective, especially in between-participants designs, keyword participants should be provided with at least two study trials and granted a fair amount of time to practice their interactive images. However, what a fair amount of time amounts to and whether it interacts with the number of learning trials, remains an empirical question and warrants further investigation.

References

Atkinson, R. C. 1975. Mnemotechnics in second-language learning. *American Psychologist* 30: 821-828.

Atkinson, R.C. and M. R. Raugh. 1975. An application of the mnemotechnic keyword to the acquisition of Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory* 104: 126-133.

Avila, E. and M. Sadoski. 1996. Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning* 46: 379-395.

Carney, R. N. and J. R. Levin. 1998. Do mnemonic memories fade as time goes by? Here's looking anew! *Contemporary Educational Psychology* 23: 276-297.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Elhelou, M. A. 1994. Arab children's use of the keyword method to learn English vocabulary words. *Educational Research* 36: 295-302.

ELLIS, N. C. 1995. The psychology of foreign language vocabulary acquisition: implications for CALL. *Computer Assisted Language Learning* 8: 103-128.

GLASS, G. V. and K. D. HOPKINS. 1996. *Statistical methods in education and psychology* (3rd edition). Boston: Allyn and Bacon.

GODLEY, R. M., G. P. FOURNET, and R. E. ESTES. 1987. The effects of administration, treatment, and pacing factors on the keyword method of foreign vocabulary learning. *Educational and Psychological Research* 7: 173-181.

GRUNEBERG, M. M. and K. PASCOE. 1996. The effectiveness of the keyword method for receptive and productive foreign vocabulary learning in the elderly. *Contemporary Educational Psychology* 21: 102-109.

HUCK, S. W. and W. H. CORMIER. 1996. *Reading statistics and research* (2nd edition). New York: Harper Collins College Publishers.

KASPER, L. F. and A. L. GLASS. 1988. An extension of the keyword method facilitates the acquisition of simple Spanish sentences. *Applied Cognitive Psychology* 2: 137-146.

MERRY, R. 1980. The keyword method and children's vocabulary learning in the classroom. *British Journal of Educational Psychology* 50: 123-136.

PRESSLEY, M. 1977. Children' use of the keyword method to learn simple Spanish vocabulary words. *Journal of Educational Psychology* 69: 465-472.

PRESSLEY, M. and J. R. LEVIN. 1978. Developmental constraints associated with children's use of the keyword method of foreign language vocabulary learning. *Journal of Experimental Child Psychology* 26: 359-372.

PRESSLEY, M., J. R. LEVIN, and G. E. MILLER. 1981. The keyword method and children's learning of foreign vocabulary with abstract meanings. *Canadian Journal of Psychology* 35: 283-287.

RAUGH, M. R. and R. C. ATKINSON. 1975. A mnemonic method for learning a second language vocabulary. *Journal of Educational Psychology* 67: 1-16.

RAUGH, M. R., R. D. SCHUPBACH, and R. C. ATKINSON. 1977. Teaching a large Russian language vocabulary by the mnemonic keyword method. *Instructional Science* 6: 199-221.

RODRÍGUEZ, M. and M. SADOSKI. 2000. The effects of rote, context, keyword, and context/keyword on retention of vocabulary in EFL classes. *Language Learning* 50: 385-412.

STAHL, S. A. and M. M. FAIRBANKS. 1986. The effects of vocabulary acquisition: a model-based meta-analysis. *Review of Educational Research* 56: 72-110.

THOMAS, M. H. and A. Y. WANG. 1996. Learning by the keyword mnemonic: looking for long-term benefits. *Journal of Experimental Psychology: Applied* 2: 330-342.

WANG, A. Y. and M. H. THOMAS. 1995. Effect of keywords on the long-term retention: help or hindrance? *Journal of Educational Psychology* 87: 468-475.

WANG, A. Y., M. H. THOMAS, and J. A. OUELLETTE. 1992. Keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology* 84: 520-528.

WANG, A. Y., M. H. THOMAS, C. M. INZANA, and L. J. PRIMICERIO. 1993. Long-term retention under conditions of intentional learning and the keyword mnemonic. *Bulletin of Psychonomic Society* 31: 545-547.

Table 1. Cued Recall by Group, Full Sample

	Retention Interval					
	Immediate			Delayed		
Treatment	M	SD	n	M	SD	n
Rote 12.35	2.06	23	9.58	3.16	24	
Keyword + Interaction	12.77	1.97	22	10.13	3.98	24
Standard Keyword	13.52	1.50	23	12.25	2.51	20

Note. Maximum score = 15

Table 2. Cued Recall by Group, Restrictive Sample

	Retention Interval					
	Immediate			Delayed		
Treatment	M	SD	n	M	SD	n
Rote 12.45	2.19	20	9.70	2.79	20	
Keyword + Interaction	12.80	1.91	20	10.45	3.82	20
Standard Keyword	13.65	1.42	20	12.25	2.51	20

Note. Maximum score = 15

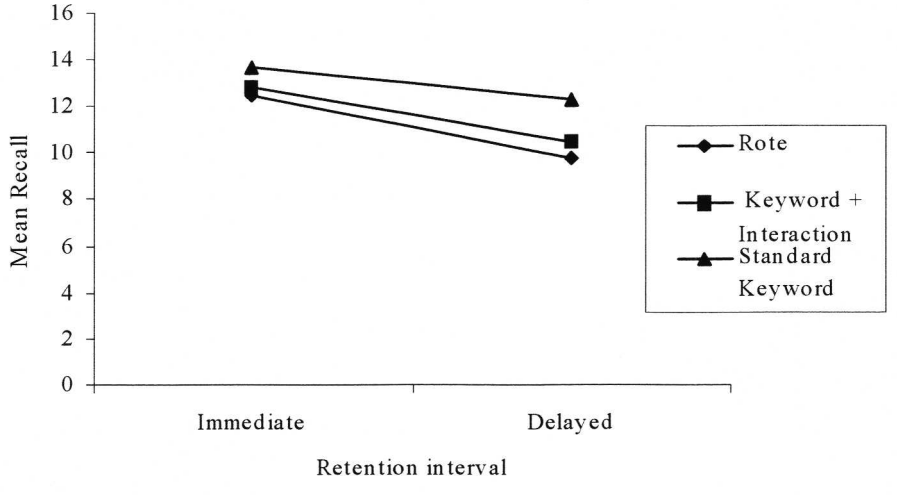

Figure 1. Mean Recall as a Function of Treatment and Retention Interval

APPENDIX

English Words, Keywords, and Interactive Images Used in the Keyword + Interaction Condition

English words	Keyword	Interactive image
lock	loco (crazy)	madman hitting himself with a lock
cart	carta (postal letter)	postal letter inside a cart
pot	pato (duck)	duck sitting in a pot
tap	tapa (lid)	lid hitting a faucet
empennage	empanada (meat pie)	meat pie hanging from airplane tail
poteen	patín (skate)	bottle of whisky sitting on a skate
jarvey	llave (key)	driver with car keys in his hands
catkin	casco (helmet)	bunch of flowers in a helmet
skillet	esqueleto (skeleton)	skeleton frying bones in a frying pan
carlin	carbón (coal)	old woman with coal all over her face
grampus	grapa (staple)	picture of a dolphin stapled on the bulletin board
manchet	mancha (stain)	piece of bread with an enormous stain
ratine	rata (rat)	rat wrapped with a piece of cloth
cordite	cordero (lamb)	lamb towing a box of explosives
hairpin	harpa (harp)	hairpin holding some harp strings
baskin	bastón (cane)	pair of boots tied to a walking cane
tweezers	tiza (chalk)	pair of tweezers holding a piece of chalk
rubble	rabo (tail)	mouse tail resting on a stone
goblet	cable (wire)	a wire plugged into an electric goblet