Journal of Experimental Psychology: Applied
1996, Vol. 2, No. 4, 330-342

1076-898X/96/$3.00

Learning by the Keyword Mnemonic: Looking for Long-Term Benefits

Margaret H. Thomas and Alvin Y. Wang
University of Central Florida

Although using the keyword mnemonic to learn new vocabulary items enhances performance on tests of immediate cued recall when compared with control strategies, the reverse is true after a delay. The authors examined long-term retention of second language vocabulary with 2 variants of the keyword method: self-generated keywords (Experiment 1) and mnemonic pictures (Experiment 2). Results showed that keyword generation did not attenuate forgetting, but provision of pictures of the keyword and translation referent during study improved long-term retention. Pictures may increase the visual detail of interactive images and hence the stability of the memory trace (Experiment 3). However, designers of instructional modules should be aware that immediate gains evidenced by learners who use the standard keyword technique may dissipate rapidly.

There is a considerable literature that argues for the beneficial effect of the keyword mnemonic on learning and immediate retention. Originally developed as a study strategy for second language vocabulary learning (Atkinson, 1975; Raugh & Atkinson, 1975), the keyword method has been applied to a wide range of learning tasks such as the acquisition of science concepts (Morrison & Levin, 1987; Rosenheck, Levin, & Levin, 1989), the names of state capitals (Levin, Shriberg, Miller, McCormick, & Levin, 1980), and factual information presented in prose format (Shriberg, Levin, McCormick, & Pressley, 1982). The recent literature abounds with reports of further successful applications; for example, keyword instruction has been demonstrated to be helpful to children with emotional disorders, students with learning disabilities, and the elderly (Dretzke, 1993; King-Sears, Mercer, & Sindelar, 1992; Mastropieri, Scruggs, & Fulk, 1990). Also, the keyword mnemonic has been implemented in recent computer-based, language-learning applications (Rabinowitz & Steinfeld, 1995; Sutherland, 1991).

The keyword mnemonic is based on the use of interactive visual imagery, an encoding strategy widely acknowledged to result in effective memory codes (e.g., Bower, 1970; Paivio, 1971). Its application to learning Tagalog, which is the national language of the Philippines, is presented in this article, in particular as it relates to learning the meaning of several vocabulary items. For example, the Tagalog word *salamin* means *eyeglasses* in English. Formation of the keyword-based interactive image proceeds across a two-step process. First, a keyword is selected on the basis of an acoustic or orthographic quality of the

Margaret H. Thomas and Alvin Y. Wang, Department of Psychology, University of Central Florida. The authors shared equally in the research and writing of this article, and the order of authorship was determined arbitrarily.

Portions of this research were presented at the following annual meetings: Psychonomic Society, San Francisco, November 1991, and Southeastern Psychological Association, Knoxville, Tennessee, March 1992, and New Orleans, Louisiana, March 1994.

We thank Shawn Burke, Jerry Crutchfield, Kelly French, Kathleen Guess, Madison Johnson, Deborah Little, Jeannine Mielke, Alesa Neighbarger, Deanna Sykes, and Julie Schulte for help with data collection and Jeannine Mielke, the artist who drew the pictures used in Experiments 2 and 3.

Correspondence concerning this article should be addressed to Margaret H. Thomas, Department of Psychology, University of Central Florida, Orlando, Florida 32816-1390. Electronic mail may be sent via Internet to mthomas@pegasus. cc.ucf.edu.

Tagalog stimulus term. Ideally, the keyword should be a highly familiar and concrete English word; perhaps, in this case the word *salmon.* Second, the learner is instructed to form a vivid interactive image combining the keyword *salmon* with the English response term *eyeglasses.* Under such circumstances, many people may imagine a salmon wearing eyeglasses. Later, when asked to recall the meaning of *salamin,* they retrieve the keyword *salmon* along with the interactive keyword-based image of a "salmon wearing eyeglasses."

Theoretically, the beneficial claims for use of the keyword method are not surprising. Paivio's (1975) dual coding theory proposed that an image provides a second and independent memory code in addition to the ubiquitous verbal code. Thus, retrieval should be more likely when images have been generated at encoding because there are two memory codes for the event, either of which could support recall. Alternatively, Marschark and colleagues (Marschark & Cornoldi, 1991; Marschark & Hunt, 1989; Marschark & Surian, 1989), although rejecting the notion of a dual memory system, suggested that imaginal processing enhances recall by increasing the relative relational and distinctiveness information generated at encoding and subsequently available at retrieval. Regardless of their differences, both of these accounts seem to imply that keyword learning is particularly long-lasting because of the increased variety of information in long-term store.

Recently, however, the long-term effectiveness of the keyword mnemonic has been questioned (Wang, Thomas, Inzana, & Primacerio, 1993; Wang & Thomas, 1995; Wang, Thomas, & Ouellette, 1992). Wang et al. (1992) reviewed keyword method studies that included measures of long-term retention and observed that studies showing a keyword advantage after a delay were based on within-participants comparisons. However, when contrasts are based on repeated testing of the same individuals, both initial acquisition rates and levels of immediate recall are potential confounding variables (Slamecka & Katsaiti, 1988; Underwood, 1954, 1966; Wang, 1991). Thus, in three experiments reported by Wang et al. (1992), the keyword method was applied to learning second language vocabulary words where the delay interval was manipulated as a between-participants variable. In this paradigm, regardless of whether initial acquisition levels were allowed to vary or whether they were held constant, an unanticipated pattern of results was consistently obtained: Greater forgetting was found under conditions of keyword learning compared with controls who were encouraged to engage in rote rehearsal. For example, in Experiment 3 on an immediate retention test, recall of keyword learners was 80% as compared with 69% for rote learners. However, for those tested after a 1-week delay, the recall level of keyword learners was only 10%, whereas the recall performance of controls was significantly higher (30%).

To obtain a relatively pure measure of forgetting, Wang and Thomas (1992) used an incidental learning paradigm so that participants in the delay groups would be unlikely to rehearse the materials during the delay interval. Differential rehearsal can be a potential confound when different learning strategies are compared because strategies vary according to the ease with which they can be reinstated by learners during the delay interval. Because everyday learning and memory tasks are usually intentional, Wang, Thomas, Inzana, and Primicerio (1993) decided to replicate this phenomenon under an intentional learning set. Nevertheless, controls still recalled significantly more (62%) than keyword learners (43%) after a 2-day delay even though the keyword and the rote rehearsal control conditions had comparable levels of immediate recall (~80%).

Furthermore, Wang and Thomas (1995) obtained similar results when comparing keyword learning to a semantic context control condition where students learned second language vocabulary words within the context of meaningful sentences. When semantic context learners were compared with keyword learners (Experiment 2), immediate retention of the keyword group was higher (77%) than that of semantic context learners (36%). However, after a 2-day delay, semantic context learners recalled more translation equivalents (32%) than keyword learners (16%).

These findings suggest caution in recommending application of the keyword method to real-life settings where long-term retention is of considerable importance. However, there are interesting variations of the standard keyword technique that incorporate features that might lead to increased

resistance to long-term forgetting. In this article, we report on our examination of two such procedures: study with self-generated keywords and picture-enhanced keyword learning.

Wang et al. (1992) suggested that self-generated keywords might lead to more stable encodings and better long-term retention than keywords supplied by an experimenter. In fact, retention of paired associates is higher when learners actively generate their own mediators, as opposed to when these mediators are provided by an experimenter (Pelton, 1969; Schwartz, 1971; Wang et al., 1983; see Experiment 3). Presumably, successful self-generated mediators involve encodings that learners spontaneously produce during acquisition and then can reinstate during recall.

Another variation of the keyword mnemonic calls for providing learners with pictures during study. The effectiveness of an imagery-based mnemonic like the keyword method is probably at least partially from visual-like cues that increase the vividness and concreteness of item pairs (Wang & Thomas, 1995). Although images are potent initially, they may be less durable over time (Hasher, Griffin, & Johnson, 1977); consequently they fail to support retrieval after a delay. Thus, the provision of pictures of the keyword and the translation referent during study might lend greater stability to such encodings. In Experiments 1 and 2, we investigated the potential robustness of these variations of the standard keyword method with unconfounded (between-participants) comparisons over time. In both of these experiments, an incidental learning set was induced in order to obtain relatively pure measures of recall over time.

Experiment 1

Although naturalistic application of the keyword method usually requires that learners generate their own keywords, keywords are provided in most laboratory studies. Indeed, it has been reported that on immediate recall tests, learners who generate their own keywords do not perform as well as those whose keywords were provided (Desrochers, Wieland, & Cote, 1991; Hall, Wilson, & Patterson, 1981; Patton, D'agaro, & Gaudette, 1991). This is particularly likely in instances where good keywords are relatively difficult to generate. However, it is possible that self-generated keyword mediators may be more resistant to forgetting over a delay and hence prove more effective than experimenter-supplied keywords in the long term.

Method

Materials, Participants, and Design

The stimuli were 24 Tagalog concrete nouns. The words were chosen on the basis of amenablity to the generation of obvious concrete English keywords (see Wang et al., 1992; Appendix B). For example, some of the items were *araw–arrow–sun; pana–pan–bow; salamin–salmon–eyeglasses;* and *baboy–boy–pig.* A total of 85 introductory psychology students participated in partial fulfillment of course requirements. All were native English speakers with no prior knowledge of Tagalog. They were divided into groups of 2 to 5. A 2 × 3 factorial design was used with learning strategy (keyword–generate, keyword-provided, and rote repetition) and retention interval (5-min and 2-day) as between-participants factors. The *n*s in each treatment group ranged from 11 to 16.

Procedure

Participants were told that we were interested in comparing reactions to different learning strategies. We told them that they would be asked to use different strategies to learn foreign language materials and to rate the strategies according to how well they liked them. Then we taught them how to use one of the three study strategies.

Keyword–generate condition. Students were first instructed in the rationale and use of the keyword technique. The characteristics of good keywords were discussed, and several examples were given for Spanish language stimuli. Each Tagalog word was then presented, and learners were asked to write down the best keyword they could come up with. To gauge the stability of these encodings, we collected the first sheet and reminded the participants, "A good keyword not only sounds like the foreign word but also refers to an object that can be easily imagined." The Tagalog items were again presented in a different order; again we instructed participants to write down their best keyword. They were told that if

they thought of a better keyword than they had on Trial 1 they should change it.

We then distributed a list of the Tagalog words, their English equivalents, and a blank for the keyword. Learners were given 6 min to study the list and were told to write their keywords in the appropriate blanks. We reminded them to form vivid, interactive images that combined keywords and English words. They were asked to make a check next to each language unit after they had successfully formed an interactive image.[1] To encourage efficient use of the acquisition period, we informed participants when 3 min and 1 min remained for this task. Following the study period, a general knowledge test was administered (e.g., "Which river is longer: the Nile or the Amazon?"). Participants in the 5-min retention group were then either given a cued recall test that listed all 24 Tagalog words with blanks for writing the keyword and the translation equivalent. We allowed 5 min for completion of this task. Those in the 2-day delay group were dismissed and asked to return in 2 days. Upon their return, these participants took the cued recall test. At Session 1, these students were told that they would be taught a different learning strategy at Session 2 where they would use new materials. No mention was made of a delayed test on the material studied in Session 1. None of the participants reported on our debriefing questionnaire that they had attempted to rehearse the materials during the delay interval.

Keyword-provided condition. To equate exposure to the Tagalog words and keywords to that of learners in the keyword-generate condition, we presented participants in this group with each Tagalog word with its experimenter-provided keyword across two trials at 5-s intervals. Then, just as in the keyword–generate condition, they studied the entire list during a 6-min period, checking each unit after they had formed an interactive image. The general knowledge test and cued recall tests were administered, as appropriate for each participant's delay group.

Rote repetition condition. Rote repetition learners were presented with each Tagalog–English language unit once at a 15-s rate. Booklets were provided in which each pair was printed at the top of the page followed by 10 blank lines. To discourage use of spontaneous elaborative strategies, we asked participants to think of the Tagalog word while writing down the appropriate English word as many times as possible within the 15 s. The general knowledge test and the test of cued recall, either immediately or after 2 days, followed.

Table 1
Cued Recall Scores as a Function of Learning Condition and Length of Retention Interval in Experiment 1

Condition	5-min		2-day	
	M	*SD*	*M*	*SD*
Keyword–generate	14.9	5.3	4.0	2.7
Keyword–provided	20.4	1.8	2.1	1.4
Rote repetition	15.3	5.4	12.8	6.9

Results and Discussion

Self-generation of keywords for the set of Tagalog words used in the present study appears to have been a reasonable task. In the keyword–generate condition, participants were able to generate keywords for virtually all 24 items ($M = 23.5$). The condition means for cued recall are shown in Table 1. A 2 × 3 analysis of variance (ANOVA) showed significant main effects for retention interval, $F(1, 79) = 110.15$, $MSE = 21.23$, $p < .001$, and learning strategy, $F(2, 79) = 7.21$, $p < .002$. The interaction was also highly significant, $F(2, 79) = 21.56$, $p < .001$. A Fisher's least significant difference of 4.06, $p < .02$, which was performed on these data, indicated that the 5-min cued recall performance for the keyword-provided condition was superior to the performance of the keyword–generate and rote repetition conditions, which did not differ statistically. However, after 2 days, rote rehearsal participants recalled significantly more than either of the two keyword groups. Specifically, the delayed cued-recall performance of both key-

[1] These check marks served two functions: They allowed keyword learners to self-monitor their progress during the unpaced acquisition period, and, they were later tallied to determine the extent to which participants were able to complete the task of forming images for all language units. This analysis revealed that keyword participants in both keyword learning groups successfully formed images for over 90% of the language stimuli.

word conditions was statistically comparable (10% to 20%). In contrast, the long-term performance of rote repetition learners was much higher (65%).

As in previous articles (Wang et al., 1992, 1993; Wang & Thomas, 1995), keyword forgetting cannot account for the striking long-term retention decrements observed for all keyword learners. Students in the keyword–generate condition recalled 21.6 and 21.2 keywords after 5 min and after 2 days, respectively. Similarly, those in the keyword-supplied group recalled 23.9 keywords (5-min delay) and 20.3 keywords (2-day delay).

Apparently, allowing keyword learners to generate their own keywords does not appreciably alter the rapid forgetting that has been observed in previous studies. Although the students had no trouble coming up with keywords, they rarely changed them. In fact, fewer than 4% were changed on the second keyword generation trial. Experimenter-provided keywords were clearly more effective than either self-generated keywords or rote rehearsal when recall was tested soon after study. However, after a 2-day delay, performance of all keyword learners was inferior to participants who used the rote repetition method, regardless of the source of their keywords.

Experiment 2

Several researchers have reported benefits on immediate tests when learners are presented with mnemonic pictures of the keyword and the information that they need to recall. These effects have been demonstrated particularly with learners who have difficulty forming visual images that are vivid and interactive. For example, several studies have reported benefits for both young and elderly learners (Dretzke, 1993; Shriberg, Levin, McCormick, & Pressley, 1982; Levin, Shriberg, & Berry, 1983). Also, illustrative mnemonics are helpful for students with learning disabilities (Mastropieri et al., 1990).

Because Dretzke (1993) did not find that pictures enhanced the keyword learning of college students, we must conclude that young adults are probably more adept at forming their own visual images. Thus, little or no benefit might be expected for these learners on an immediate test of recall; indeed, a ceiling effect may have occurred in this condition. However, it is possible that benefits might be observed for long-term retention. Study with pictures might enhance long-term retention for at least two reasons: (a) the visual detail present in a picture may result in a more elaborate memory code than a totally self-generated image, and (b) the fragility of keyword memories may be due to the existence of alternative representations for many of the keywords beyond the ones evoked in the mnemonic image (Wang et al., 1992). That is, the keyword-based image might not use a representation of the keyword ordinarily produced by the learner. To the extent that different encodings are elicited during acquisition and attempted recall, decrements in recall performance can be expected (Hasher, Griffin, & Johnson, 1977; Hasher & Johnson, 1975). For example, in the case of a participant who generated an image of a "salmon wearing eyeglasses" when given the keyword *salmon* and the English word *eyeglasses* (even if this keyword-based image were highly interactive, distinctive, and vivid), the cue did not evoke an image of a *salmon* that this individual would have spontaneously produced under normal conditions. Instead, this person may consistently think of "orange-colored fish steaks" when seeing the word *salmon*. Thus, recall of the appropriate keyword *salmon* when presented with *salamin* may be of little use to the learner if "orange fish steaks" is elicited rather than "a whole fish" wearing eyeglasses. Even less extreme variations in the interpretation of the keyword may have deleterious effects on long-term recall. Perhaps the word *salmon* later evokes an image of a realistic fish swimming upstream rather than the more cartoonlike rendition that the learner imagined wearing eyeglasses during the study trials. The likelihood of the evocation of an alternate representation of the keyword should be particularly pronounced after an extended delay between study and subsequent testing.

In Experiment 2, we explored the possibility that the positive effects of the keyword method observed soon after study could be extended across the delay interval by increasing the availability of vivid and distinctive visual cues at the time of encoding. In this study, we supplemented the standard keyword learning procedure by providing participants with separate drawings of the keyword and of the translation word during acquisition. They were instructed to imagine an interactive image of the two objects as depicted

in these pictures. For keyword–picture learners, in addition to the standard test of cued recall, we also included a second measure of recall, which was cued this time by the picture of the keyword itself. If keyword learners have difficulty in reinstating the original encoding of the keyword after a delay, then this type of cue should be a potent retrieval aid.

Method

Materials, Participants, and Design

Separate drawings in black ink of a 24-item set of keywords and the translation referents were created by a professional artist. Nine of the 24 items were new, and the other 15 were from the stimulus set used in Experiment 1. The design was a 2 × 3 factorial with two between-participants variables: retention interval (5-min and 2-day) and learning strategy (keyword, keyword–picture, and rote repetition). Introductory psychology students (n = 103) naive to the Tagalog language participated. There were 8 to 21 participants in each of the six treatment conditions.

Procedure

Keyword condition. Participants were instructed as to the rationale of the keyword method. We placed emphasis on the importance of generating vivid, interactive images. Several examples of Spanish–English word pairs were given. The 24 Tagalog–English keyword pairs were then presented on cards. The entire list was presented twice in varying different order at a rate of 10 s. Before the second list presentation, participants were reminded again to form vivid, interactive visual images for the keyword and the English word.

After the study trials, participants in the 5-min delay group took a 5-min general knowledge test. Then, we administered the cued recall test. Participants were given a list of the Tagalog words with their keywords and were asked to write the appropriate English translation in the blanks provided. We allowed 5 min for participants to complete the test. Students in the 2-day delay condition were treated similarly, except that they were not given the recall tests until Session 2.

Keyword–picture condition. Participants were instructed in the keyword method just as those in the standard keyword condition. However, on each card displaying the Tagalog–English keyword pairs, there also was a drawing of the object or entity denoted by keyword and the English word. These drawings were positioned directly above the printed words (see Figure 1). The participants were told to imagine the items drawn on the cards in an interactive image of their own design. As in the standard keyword group, the entire list was presented twice in varying orders at a 10-s rate with the experimenter displaying the appropriate cards on each trial. The testing sequence was identical to that in the standard keyword group with one exception. Participants in the keyword–picture group were given a final picture-cued recall test after the standard cued recall test. The cards displaying the Tagalog word, its keyword, and the picture of the keyword were presented in random order at a 10-s rate. The participants were asked again to write the English translation equivalent in the appropriate blank on an answer sheet.

Rote repetition condition. The control condition was designed so that the initial performance level would be comparable with that of the mnemonic group. This can be accomplished by giving nonmnemonic learners a greater number of list presentations at a faster rate (Hall, Owens, & Wilson, 1987; Wang et al., 1992). Thus, on the basis of pilot testing, rote repetition learners were exposed to the Tagalog–English word pairs for five trials. The pairs were printed on cards and presented at a 2-s rate with the item order random across the trials. Total study time was held equivalent to that for the keyword–picture group. Half of the participants were given the standard cued recall test (a list of the Tagalog words) immediately after taking the 5-min test of general knowledge. The others were dismissed after the general knowledge test and were given the recall test 2 days later.

Results and Discussion

Standard Cued Recall

The means for the six treatment conditions are presented in Table 2. A 2 × 3 ANOVA revealed

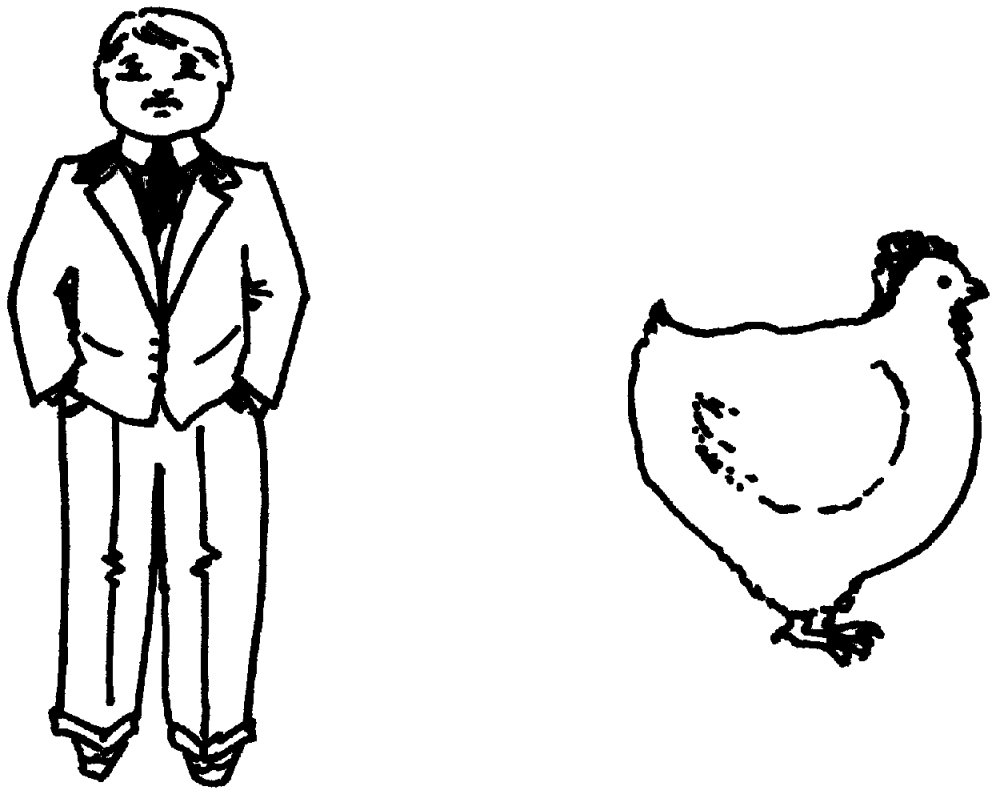

Figure 1. Sample of illustrations from Experiment 2 shown in conjunction with a Tagalog–English keyword pair.

a significant main effect of retention interval, $F(1,97) = 126.82$; $MSE = 21.27$; $p < .001$, reflecting decline in performance over time. Also, the Retention Interval × Learning Strategy interaction was significant, $F(2, 97) = 3.26$; $p = .04$. This interaction was due to the fact that even though immediate performance was not reliably affected by learning strategy, learners in the standard keyword group recalled less than controls after a delay of 2 days (Fisher's least significant difference = 3.75, $p < .02$).[2] However, the performance of the picture–key-

Table 2
Cued Recall as a Function of Learning Condition and Length of Retention Interval in Experiment 2

Condition	5-min		2-day	
	M	*SD*	*M*	*SD*
Keyword	19.46	5.02	5.75	2.54
Keyword–picture	19.15	3.72	7.81	4.73
Rote repetition	17.33	5.40	9.65	4.72

[2] The only nonsignificant effect in the ANOVA was the main effect of learning strategy. This suggests that the rote rehearsal procedure, which was designed to generate an initial acquisition level equivalent to that of the keyword conditions, was successful. Because of the inherent unpredictability of participants keeping their appointments for the small-group test sessions, there was a smaller number of participants in the standard keyword conditions ($n = 13$, 5-min retention; $n = 8$, 2-day retention) than in the other treatment groups (all $ns = 20$). With these sample sizes, a small effect of strategy would not be detected, given the power of the analysis (Cohen, 1969). However, power was sufficient to detect a large effect. Because large differences are usually noted when comparing the keyword method to other strategies on immediate recall, we are satisfied that they were largely eliminated here.

word group did not differ reliably from the other two groups.

Picture-Cued Recall

For the keyword–picture learners, picture-cued recall means were 20.60 (SD = 3.51) for the 5-min retention group and 9.80 (SD = 5.14) for the 2-day delay group. A 2 (recall measure) × 2 (retention interval) ANOVA was performed on their cued recall and picture-cued recall scores. Not surprisingly, there was a significant main effect of retention interval, $F(1, 39) = 69.22$; $MSE = 36.24$; $p < .001$, indicating that recall reliably declined over time. Of primary interest to us was the finding that the picture cue reliably increased recall by about two items, $F(1, 39) = 39.32$; $MSE = 1.55$; $p < .001$. The interaction did not approach significance ($F < 1$).

Provision of distinctive pictures to keyword learners at study retarded the rate of long-term forgetting. In contrast to previous research with the standard keyword method and in our Experiment 1 where keyword learners were given pictures at encoding, their level of cued recall after a delay did not differ statistically from that of controls. Also, when recall of the picture–keyword learners was prompted with a picture of the keyword from the original study period, performance was significantly enhanced. It should be noted that the picture-cued recall measure followed a test of standard cued recall. If the standard cued recall test served as an occasion to elicit alternative representations of the keyword, it might be that our data have underestimated the potential of the picture cue to retrieve the keyword-linked image. Future research should include unconfounded measures of picture-cued recall. In this way, the extent to which recall might be improved by such cuing could be explored.

Experiment 3

As suggested above, in addition to the relational cues provided by interactive images, the visual nature of images confer more elaborate and detailed information in memory than verbal encodings. In Experiment 3, we attempted to examine the importance of visual cues as compared with relational cues in the keyword imagery mnemonic. To simplify procedures, we focused on the paired associate learning link in the keyword method by omitting the Tagalog word. That is, participants were presented with English word pairs, that is, the keyword–translation referents from our prior studies. We compared learners who were provided with interactive pictures in acquisition with those who were given a meaningful sentence relating the two words during the study period. Presumably, sentence stimuli emphasized the relational attributes of the word pairs, whereas pictorial stimuli promoted perceptually based encodings in addition to relational information. To accentuate the role of perceptual versus relational cues, we introduced two different distractor tasks during the retention interval. Therefore, half of the participants in each acquisition group were given one of two distractor tasks. One task was designed to interrupt visually based cues, and the other was targeted to interfere with relational information. If interactive visual processing enhances retention primarily by its additional perceptually based information, learners in the interactive pictures group should have been especially hindered by the distractor task designed to interrupt visual cues, whereas those who encoded the items in terms of their relation in a sentence should have been more negatively affected by interference with relational cues. To obtain relatively uncontaminated measures of recall performance, we used an incidental learning set as in Experiments 1 and 2.

Method

Materials, Participants, and Design

The 24 word pairs were the keyword–translation referents from Experiment 2. Interactive drawings of each item pair were created by a professional artist. Participants (n = 76) were assigned randomly to one of the four treatment groups of a 2 (interactive pictures and relational sentences) × 2 (picture distractors and sentence distractors) design. The data for each condition were collected in groups of 2 to 5 participants; ns in each treatment condition ranged from 13 to 23.

Acquisition. Learners in the interactive pictures group were told to look at the pictures and to visualize them vividly and distinctly. Each picture was presented for 10 s, and the entire list

was presented twice in random order. Learners in the relational sentences condition were presented with a sentence printed in large type on a card that included both words, with the relevant ones underlined. For example, for the item *MAN–CHICKEN,* there was an interactive picture of a man holding a chicken (see Figure 2). The actual sentence was read as follows: "The MAN scattered the feed around the farmyard for the CHICKEN." Each sentence was presented for 10 s, and the entire list was presented twice in random order.

Distractor tasks. The picture distractor task required participants to examine a set of 48 drawings that showed the items in different pairings. For example, the chicken was then shown standing inside a hot dog bun; the original pairing for bun was *BUN–DOG,* which showed a dog in a hot dog bun. The participants were asked to rate each picture as to how easy or difficult they would be to draw. These instructions were designed to accentuate the visual qualities of the pictures rather than the relationships depicted. The sentence distractor test was composed similarly; that is, 48 sentences were created by using the original words in different pairings (e.g., "The CHICKEN sandwich was served on a sesame seed BUN"). Participants in the sentence distractor condition were asked to rate each sentence on the degree of reasonableness of the relationship expressed. In this task, attention was drawn deliberately to a consideration of the relational attributes of the paired items.

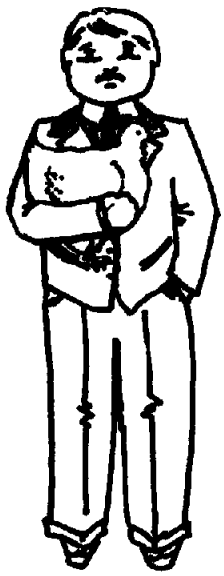

Figure 2. Sample of illustration from Experiment 3.

Table 3
Cued Recall as a Function of Learning Condition and Type of Distractor Task in Experiment 3

Condition	Pictures		Sentences	
	M	*SD*	*M*	*SD*
Interactive pictures	15.41	5.57	18.62	5.44
Sentences	13.91	6.92	11.39	6.05

Procedure. After the interactive picture or the relational sentence condition test, half of the participants in each acquisition group were exposed to the picture distractor task and the other half to the sentence distractor task. Then, all participants took the 5-min general knowledge test, with the test of cued recall following immediately. Each stimulus word was printed on a sheet of paper with a corresponding blank. We allowed 5 min for participants to complete the test.

Results and Discussion

The mean cued recall scores for participants in each condition are presented in Table 3. These scores were analyzed by a 2 × 2 completely randomized ANOVA. The main effect of study method was significant, $F(1, 72) = 9.22$; $MSE = 37.26$; $p = .003$, whereas the overall effect of the type of distractor was not ($F < 1$). The interaction was also significant, $F(1, 72) = 3.98$; $p < .05$.

Picture distractors were more disruptive than relational sentence distractors for learners who had studied interactive pictures during acquisition (see Table 3). In contrast, reading new relational verbal encodings was more disruptive for those individuals who had studied sentences than was viewing pictures of these items in new pairings. To examine this outcome more fully, we classified recall errors either as errors of commission or errors of omission.[3] The mean proportion

[3] The effectiveness of the distractors is evidenced by the presence of errors of commission. In contrast, in Experiments 1 and 2, participants rarely made errors of commission regardless of treatment condition. Instead, incorrect responses were failures to respond at all.

of total errors that were errors of commission for each group are presented in Table 4. Errors of commission predominated when the distractor task was similar to the encoding task. That is, interactive picture learners were more likely to confuse the pairings and make errors of commission when they had viewed distractor pictures than when they had read relational sentences. In contrast, sentence learners were less likely to pair items incorrectly when they were presented with picture distractors than when they were presented with sentences expressing new relations. A 2×2 ANOVA confirmed this interpretation. Only the interaction was significant, $F(1, 68) = 10.92$, $MSE = .09$, $p < .001$. For all other Fs, ps $> .18$.

General Discussion

Under standard implementation procedures, vocabulary learning by the keyword method is prone to rapid forgetting (Wang et al., 1992, 1993; Wang & Thomas, 1995). The present research investigated the robustness of two variants of the standard keyword mnemonic with respect to long-term retention. In Experiment 1, self-generation of keywords was found to provide no benefit as compared with use of experimenter-provided keywords over a 2-day retention interval. Both groups of keyword learners, self-generated and experimenter-provided keywords, forgot more over the delay interval than learners in the control group. Also, consistent with prior research (Desrochers et al., 1991; Hall et al., 1981; Patton et al., 1991), we found that learners who used experimenter-provided keywords performed better on an immediate retention test than students who generated their own keywords. Thus, self-generation of keywords appears not to facilitate immediate or long-term retention.

In Experiment 2, picture-enhanced keyword learning was examined in comparison with the standard keyword method. Consistent with earlier reports (Wang et al., 1992, 1993; Wang & Thomas, 1995), students using the standard keyword method performed more poorly than the rote repetition control group after a delay of 2 days. However, students provided with pictures of the keyword and the translation referent during acquisition had a forgetting rate that did not exceed that of the control group. Perhaps pictures increase the visual detail of the interactive image at encoding and thus produce a memory trace that is more robust over time than a totally self-generated image. This interpretation is supported by results of Experiment 3. Students studied paired associates by viewing interactive pictures of the words or by reading sentences relating the words. A distractor task where they were either shown pictures depicting new pairing of the words or new sentences incorporating the words followed. Different instructional sets were used for each distractor task so that attention to visual detail would be emphasized for pictures and attention to relational information would be emphasized for sentences. For students who had originally studied interactive pictures, recall of the original pairing was disrupted more by picture distractors than by sentence distractors. This suggests that distinctive visual detail is a more important factor than relational information for this type of encoding. In contrast, relational information proved more important for those who initially studied the word pairs in the context of meaningful sentences. For these students, recall was more disrupted by new sentence distractors than by pictures. Thus, the similarity of the distractor code to the original stimulus encoding predicted forgetting.

Table 4
Proportion Errors of Commission as a Function of Learning Condition and Distractor Task in Experiment 3

Condition	Pictures	Sentences
Interactive pictures	.37	.15
Sentences	.22	.50

Although we did not design our research to compare directly the different theoretical accounts of imagery-based, paired-associate learning offered by Paivio and colleagues (e.g., Paivio, Walsh, & Bons, 1994) and by Marschark and Hunt (1989), some implications of the data are suggested. The two approaches differ primarily in that Paivio's (1975) dual-coding theory assumes that relational information from verbal or visual encoding is independent of the encoding of visual detail. Hence, recall is aided by the existence of two separate memory codes. In contrast, Marschark and Hunt (1989) argued that imagery enhances the distinctiveness of the words and of their relations. This distinctiveness allows retrieval of a particular encoding from the larger set

of related associates that might be elicited by the cue word. According to Marschark and Hunt, the effects of imagery and relational encoding are not independent, and imagery enhances cued recall when relational processing is present at encoding and retrieval.

Of relevance to the present results, we must note that both approaches have emphasized the memorial benefits of the use of imagery. The principal question relates to how they might account for the surprising fragility of these encodings over time. Initial benefits of imagery-based encodings dissipate rapidly and can thereby be encompassed within dual coding theory framework if it is assumed that a visual memory code, as compared with a verbal code, decays more quickly or becomes more difficult to retrieve over time. The argument for two separate memory codes also is compatible with the finding that picture distractors are more disruptive of the immediate benefit in cued recall resulting from studying pictorial materials during acquisition than are verbal relational distractors. This pattern of results could be viewed as supporting Paivio's (1971) position that visual encodings are independent of relational encodings and are thus less susceptible to interference by relational information.

Marschark and Hunt's (1989) analysis implies that imagery-based encodings confer an advantage in cued recall only when the concomitant relational information is first accessed. In Experiment 3, the sentence distractor condition was designed to focus attention on new relational encodings. It seems that the view of Marschark and Hunt predicts that sentence distractors will disrupt encodings that are based on pictures, as well as those based on relational sentences, because relational information is a necessary first step in the retrieval process in both cases. However, our data show that when sentence distractors were given to the picture learners, the rate of errors of commission, which suggested interference by new encodings, was the lowest among all groups.

Long-term recall after picture-enhanced learning was bolstered by provision of a picture of the keyword. This suggests that retrieval of keyword-linked images may also be hindered by alternative representations of the keyword that may be evoked after a delay. In part, this type of interference may also at least partially explain the vulnerability of keyword learning to rapid forgetting. Such a process is compatible with Marschark and Hunt's (1989) emphasis on the importance of relational processing in cued recall. However, the benefit of picture cues would also be compatible with the notion of dual memory traces because such cues would be particularly effective in accessing the visual memory trace. However, it should be reiterated that because the present research was not designed to compare specifically these two theoretical positions, these comments are tentative and are meant to be suggestive of further research.

The present results, taken together with the earlier reports of Wang and colleagues (Wang et al., 1992, 1993; Wang & Thomas, 1995), suggest several factors that should be considered in applications of the keyword mnemonic in naturalistic settings. First of all, it is doubtful that use of this mnemonic will be successful when students are given instruction in its implementation and then expected to incorporate it into their study regimen without further guidance. Students apparently have difficulty in generating keywords that will be as effective as experimenter-provided ones. This is particularly true with stimuli for which good keywords are difficult to generate. Because Experiment 1 replicated this result with stimuli designed to be keyword hospitable suggests particular caution. Successful application of the keyword method may depend on provision of good keywords to students or on more extensive training in the process of generating good keywords. Secondly, the high initial recall rates after study with the keyword method may engender inaccurate metamemory. Students should be made aware that rapid forgetting is quite probable without more rehearsal (Wang & Thomas, 1995).

However, the results of Experiment 2 suggest that the stability of learning with the keyword technique may be improved when pictures supplement the standard visual imagery instructions. Certainly, with the increasing capability of computers to generate and present pictures, this enhancement is now technologically feasible. The finding that provision of the picture of the keyword additionally aided retrieval suggests that frequent presentation of this cue during training could be an effective learning aid.

We encourage further work in applications of

the keyword mnemonic to naturalistic learning situations, and we agree with other researchers (Levin & Pressley, 1985; Levin, Levin, Glasman, & Nordwall, 1992) that its potential may be positive. However, it should be noted that as yet no clear demonstration of a benefit on long-term retention with this mnemonic has been reported. Although we found use of pictures to be a helpful addition to the keyword method, long-term retention afterwards was still no better than that of a control group. Finally, research has suggested that people who are given training in the use of mnemonics rarely continue to use the techniques in their everyday lives (Bellezza, 1983; Higbee, 1988; Lapp, 1983; Roberts, 1983; Wood & Pratt, 1987). Indeed, memory researchers themselves are no more likely to report that they use mnemonics than highly educated individuals in other fields (Park, Smith, & Cavanaugh, 1990). Clearly there is a gap between the promise and practice of mnemonic learning that will be bridged only by demonstrations of naturalistic implementations that achieve positive long-term outcomes.

References

Atkinson, R. (1975). Mnemotechnics in second language learning. *American Psychologist, 30,* 821–828.

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 126–133.

Bellezza, F. S. (1983). Mnemonic-device instruction with adults. In M. Pressley & J. R. Levin (Eds.), *Cognitive strategy research: Psychological foundations* (pp. 51–73). New York: Springer-Verlag.

Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist, 58,* 496–510.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Desrochers, A., Wieland, L. D., & Cote, M. (1991). Instructional effects in the use of the mnemonic keyword method for learning German nouns and their grammatical gender. *Applied Cognitive Psychology, 5,* 19–36.

Dretzke, B. J. (1993). Effects of pictorial mnemonic strategy usage on prose recall of young, middle-aged, and older adults. *Educational Gerontology, 19,* 489–502.

Hall, J. W., Owens, W. L., & Wilson, K. P. (1987). Presentation rates and keywords in vocabulary learning. *Bulletin of the Psychonomic Society, 25,* 179–181.

Hall, J. W., Wilson, K. P., & Patterson, R. J. (1981). Mnemotechnics: Some limitations of the mnemonic keyword method for the study of foreign language vocabulary. *Journal of Educational Psychology, 73,* 345–357.

Hasher, L., Griffin, M., & Johnson, M. K. (1977). More on interpretive factors in forgetting. *Memory & Cognition, 5,* 41–45.

Hasher, L., & Johnson, M. K. (1975). Interpretive factors in forgetting. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 567–575.

Higbee, K. L. (1988). *Your memory* (2nd ed.). New York: Prentice-Hall.

King-Sears, M. E., Mercer, C. D., & Sindelar, P. T. (1992). Toward independence with keyword mnemonics: A strategy for science vocabulary instruction. *Remedial and Special Education, 13,* 22–33.

Lapp, D. (1983). Committment: Essential ingredient in memory training. *Clinical Gerontologist, 2,* 58–60.

Levin, J. R., Levin, M. E., Glasman, L. D., & Nordwall, M. B. (1992). Mnemonic vocabulary instruction: Additional effectiveness evidence. *Contemporary Educational Psychology, 17,* 156–174.

Levin, J. R., & Pressley, M. (1985). Mnemonic vocabulary instruction: What's fact, what's fiction. In R. F. Dillon (Ed.), *Individual differences in cognition* (Vol. 2, pp. 145–172). Orlando, FL: Academic Press.

Levin, J. R., Shriberg, L. K., & Berry, J. K. (1983). A concrete strategy for remembering abstract prose. *American Educational Research Journal, 20,* 277–290.

Levin, J. R., Shriberg, L. K., Miller, G. E., McCormick, C. B., & Levin, B. (1980). The keyword method in the classroom: How to remember the states and their capitals. *Elementary School Journal, 80,* 185–191.

Marschark, M., & Cornoldi, C. (1991). Imagery and verbal memory. In C. Cornoldi & M. A. McDaniel (Eds.), *Imagery and cognition* (pp. 133–182). New York: Springer-Verlag.

Marschark, M., & Hunt, R. R. (1989). A reexamination of the role of imagery in learning and memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 710–720.

Marschark, M., & Surian, L. (1989). Why does imagery improve memory? *European Journal of Cognitive Psychology, 7,* 251–263.

Mastropieri, M. A., Scruggs, T. E., & Fulk, B. M. (1990). Teaching abstract vocabulary with the keyword method: Effects on recall and comprehension. *Journal of Learning Disabilities, 23,* 92–96.

McCormick, C. B., & Levin, J. R. (1987). Mnemonic prose-learning strategies. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 392–406). New York: Springer-Verlag.

Morrison, C. R., & Levin, J. R. (1987). Degree of mnemonic support and students' acquisition of science facts. *Educational Communication and Technology Journal, 35,* 67–74.

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

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1196–1204.

Park, D. C., Smith, A. D., & Cavanaugh, J. C. (1990). Metamemories of memory researchers. *Memory & Cognition, 18,* 321–327.

Patton, G. W. R., D'agaro, W. R., & Gaudette, M. D. (1991). The effect of subject-generated and experimenter-supplied code words on the phonetic mnemonic system. *Applied Cognitive Psychology, 5,* 135–148.

Pelton, L. H. (1969). Mediational construction vs. mediational perception in paired-associate learning. *Psychonomic Science, 17,* 21–22.

Rabinowitz, M., & Steinfeld, R. (1995). SIVL: The instructional design underlying a foreign language vocabulary tutor. In C. N. Hedley, P. Antonacci, & M. Rabinowitz, (Eds.), *Thinking and literacy: The mind at work* (pp. 141–152). Hillsdale, NJ: Erlbaum.

Raugh, M. R., & Atkinson, R. C. (1975). A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology, 67,* 1–16.

Roberts, P. (1983). Memory strategy instruction with the elderly: What should memory training be the training of? In M. Pressley & J. R. Levin (Eds.), *Cognitive strategy research: Psychological foundations* (pp. 75–100). New York: Springer-Verlag.

Rosenheck, M. B., Levin, M. E., & Levin, J. R. (1989). Learning botany concepts mnemonically: Seeing the forest and the trees. *Journal of Educational Psychology, 81,* 196–203.

Schwartz, M. (1971). Subject-generated versus experimenter-supplied mediators in paired-associate learning. *Journal of Experimental Psychology, 74,* 238–247.

Shriberg, L. K., Levin, J. R., McCormick, C. B., & Pressley, M. (1982). Learning about "famous" people via the keyword method. *Journal of Educational Psychology, 74,* 238–247.

Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 716–727.

Sutherland, R. L. (1991). The mnemonic keyword with USAF Academy cadets. *Dissertation Abstracts International, 52,* (6-A), 2082.

Underwood, B. J. (1954). Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin, 51,* 276–282.

Underwood, B. J. (1966). *Experimental psychology.* New York: Appleton-Century-Crofts.

Underwood, B. J., & Postman, L. (1960). Extraexperimental sources of interference in forgetting. *Psychological Review, 67,* 73–95.

Wang, A. Y. (1991). Assessing developmental differences in retention. *Journal of Experimental Child Psychology, 51,* 348–363.

Wang, A. Y., & Thomas, M. H. (1992). The effect of imagery-based mnemonics on the long-term retention of Chinese characters. *Language Learning, 42,* 359–376.

Wang, A. Y., & Thomas, M. H. (1995). The effect of keywords on long-term retention: Help or hindrance? *Journal of Educational Psychology, 87,* 468–475.

Wang, A. Y., Thomas, M. H., Inzana, C. M., & Primacerio, L. J. (1993). Long-term retention under conditions of intentional learning and the keyword mnemonic. *Bulletin of the Psychonomic Society, 31,* 545–547.

Wang, A. Y., Thomas, M. H., & Ouellette, J. A. (1992). The keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology, 84,* 520–528.

Wood, L. E., & Pratt, J. D. (1987). Pegword mnemonic as an aid to memory in the elderly: A comparison of four age groups. *Educational Gerontology, 13,* 325–337.

Received December 18, 1995
Revision received March 25, 1996
Accepted March 28, 1996 ■