Elaborated keyword strategies for foreign language vocabulary acquisition

Donald Hogben and Michael J. Lawson
Flinders University

The keyword method (Atkinson, 1975) has proved to be an effective means of enhancing the deliberate acquisition of foreign language vocabulary items. This has been the general finding of many studies in which the subjects have been university undergraduate students with no knowledge of the foreign language from which the words-to-be-learned were taken. The original experiments conducted at Stanford employed Russian (Atkinson, 1975; Atkinson and Raugh, 1975) and Spanish (Raugh and Atkinson, 1975). Subsequent experiments with naive undergraduates have utilised Spanish (Kasper and Glass, 1988), Latin (Pressley, Levin, Nakamura, Hope, Bispo, and Toye, 1980; Pressley and Ahmad, 1986; Pressley, Levin, and Ghatala, 1988), French (Wang, Thomas, and Oullette, 1992), and German (Desrochers, Gélinas, and Wieland, 1989; Desrochers, Wieland, and Coté, 1991). In addition to these experiments with university undergraduates, Pressley, Levin, Hall, Miller, and Berry (1980) and Miller, Levin, and Pressley (1980) demonstrated keyword effectiveness with elementary school pupils learning Spanish words.

The keyword method involves two distinct steps or stages for the learning of a new word: an acoustic stage and a visual image stage. If the method were to be used, for example, to learn the Italian word COLTELLO, meaning knife, the first step would be to find an English word that sounds like (and possibly looks like) part, or all, of the Italian word. This English word is referred to as the "keyword", and an obvious keyword candidate in this example is the English word colt (a young horse). In the second (visual image) stage, one searches for a visual image that clearly and unambiguously links the keyword and the target word: in our example, perhaps a young horse with a large knife in its mouth, or on its back. On subsequently encountering the word coltello, the keyword is generated and used to search for the image; the image being used to generate the English meaning. There is the further possibility that, having established these links, one can produce coltello in response to knife; that is, the keyword method may facilitate what has been referred to as backward recall (Pressley, Levin, Hall, Miller, and Berry, 1980).

The two stages of the method can be seen as establishing an interaction between two forms of elaborative processing (Anderson, 1990). Both the acoustic and image stages require the learner to embed features of the word in a complex associative network. As Anderson (1990) argues, we might expect, in general terms, that appropriate elaborated encoding will benefit recall
by providing a degree of redundancy that can be exploited at time of retrieval. If this argument is sound we might expect positive results from use of the method for deliberate vocabulary acquisition and, as indicated above, this has, in broad terms, proved to be the case. Nevertheless, two areas of doubt that are of specific relevance to a teaching context have been raised in discussions of the method. Atkinson (1975) raised the possibility of the keyword method not proving to be so effective for vocabulary learning if the learners were in actual foreign language learning situations. "It may prove useful only in the early stages of learning a language and more so for some languages than others" (Atkinson, 1975, p.828). This caution also appears in the Atkinson and Raugh (1975) report on Russian vocabulary learning by naive undergraduates. In his doctoral dissertation, Fuentes (1976) reported no advantage for the keyword method for second-year high school students studying Spanish. However, in their keyword method review, Pressley, Levin, and Delaney (1982) raised doubts about Fuentes' study. Pressley et al (1982) suggested that the inclusion by Fuentes of a wide variety of vocabulary items, of all grammatical forms, may have reduced the overall keyword effect. They also point out that Fuentes found that keyword subjects did in fact tend to learn concrete nouns better than did his control subjects. Willerman and Melvin (1979) investigated the learning of French vocabulary (nouns, verbs, adjectives, and adverbs or function words) by university students who were enrolled in an introductory French course. They found no advantage for the keyword method. Unfortunately, only limited instructions were given to the students, and only eight seconds were allowed each for the learning and recall stages. The results may have been otherwise had more learning and recall time been allowed. Levin, Pressley, McCormick, Miller, and Shriberg (1979) conducted four experiments to check on the findings of Fuentes (1976). Their general conclusion from these experiments with high school students studying Spanish was: "The keyword method, as conceptualised to date, cannot be translated directly into practice in actual high school foreign language classes" (p.590). This conclusion might have served to discourage further research. No reports of studies with foreign language learners appear to have been published since that time. This is perhaps somewhat surprising, since Levin et al (1979) were not entirely pessimistic about the possibility of achieving positive advantage from use of the keyword approach. They suggested that the keyword method might well be successfully implemented in high school classrooms "though it likely depends on a number of implementation and classroom climate variables" (p.593). They do not pursue this possibility at any length. However, the lack of
improvement noted in these studies has given rise to a second area of doubt about the keyword method: that the method adds little to the strategic resources of experienced learners.

One obvious suggestion for the lack of effect associated with use of the method in classroom situations is that language students develop vocabulary-learning strategies similar to the keyword method, or learn other strategies that are just as effective. Unfortunately, the evidence available on this question is rather inconsistent. Fuentes (1976) reported that 55 per cent of his control group subjects used keyword-like techniques. "Apparently, successful second-year foreign language students spontaneously use mnemonic techniques closely akin to the keyword as a matter of course. In addition their learning repertoire includes other approaches such as the use of root words and occasionally rote" (Fuentes, quoted in Nation, 1982, p.26).

However, Levin et al (1979) reported that less than 10 per cent of their control group students claimed to have used a keyword strategy, although about half of the high school Spanish students in their control groups used "strategies involving cognates, phoneme correspondences, and some other mnemonic tricks" (Levin et al, 1979, p.587). It is not clear from this latter description how closely these procedures were related to elements of the keyword method.

Willerman and Melvin (1979) did not specifically enquire of the non-keyword method students in their experiment as to which strategies were used for acquisition and recall of items. However, they conjecture that "Students who have been studying a foreign language, even if only for a month or so, have most likely developed conscious or unconscious learning strategies to master the material" (p.452). They further suggest that the students who have already developed strategies may retain these, and disregard the experimental instructions, if they believe they can learn the given vocabulary by doing so. And, where delayed post-testing is featured (one month in the Willerman and Melvin study), the students' daily assimilation of vocabulary and new language structures during the intervening period may interfere with the retention of the vocabulary items learned during an experiment.

Given the relatively meagre evidential basis for this view of the state of sophistication of language learners, there is a need to seek more detailed information from learners about procedures used for vocabulary acquisition. This became one of the purposes of the present study.

This suggestion, that the lack of effect in use of the keyword method with experienced learners arises from the development by these learners of methods for vocabulary acquisition that are of similar power, can be examined in another way: through assessment of the effects of instruction. If these learners have developed
procedures whose use has an effect similar to that of the keyword method, then the effect of instructing students in use of the method should have minimal effect when compared with use of the methods these experienced learners typically use for this task. On the other hand, if these language learners do not possess as wide a repertoire of strategies as suggested above, instruction in use of the keyword and the additional strategies might produce higher levels of recall than the keyword method alone. Thus, the second major purpose of this study was to examine the relative performance on a vocabulary acquisition task of foreign language students trained in use of keyword methods and that of students encouraged to use their usual procedures.

Two forms of the keyword procedure were developed for this instruction. The decision to include a form of keyword instruction additional to that of the standard keyword approach was prompted by our conceptualisation of the keyword method as a form of elaboration. In a recent informal observation of an experienced language learner’s attempts to acquire new vocabulary in an Italian university course, we employed a concurrent think-aloud procedure (Ericsson & Simon, 1984). What became apparent was that, although the learner was attempting to use the standard keyword two-stage method of acoustic and imaginal elaboration, he was, in addition, drawing on other features of his knowledge base in establishing links for the target word. The learner’s knowledge of grammatical features in both English and Italian provided one such basis of further elaboration. Though this observation was of a single subject, it suggested that use of the process of multiple elaboration, or types of elaboration additional to those of the standard keyword method, would be worthy of study, and may also be of relevance in consideration of the apparent threshold effect noted above. It seemed possible that one of the reasons for the lack of recall advantage noted in the studies with older/experienced learners might reflect the use by these learners of additional forms of elaboration of the kind we had observed in our case study. If this were so, it could be that the strategies argued to be employed spontaneously by experienced learners actually involve such additional forms of elaboration. To assess the power of such additional elaborative procedures, we set out to compare the word acquisition performance of experienced high school language learners using their usual approaches with that associated with a multiple elaboration condition, and a standard keyword method.

In the majority of keyword method experiments conducted to date, the experimenters have supplied keywords to students in the keyword method groups. This means that, in the main, investigations have employed only words which suggest keywords more-or-less readily. What of the effectiveness of the method for the learning and recall of words for which keywords cannot readily be generated? Michael Pressley, who has been involved in a large number of keyword studies, has recently expressed a long-
held concern on this issue: "I have argued since 1982...that keyword method gains are unlikely for items that do not have obvious keywords" (Pressley, 1991, p.166). In a study by Hall (1988), subjects reported that they were able to generate keywords for only 36 per cent of a set of words referred to by Hall as "typical words" (words not obviously suggesting keywords). If the keyword method only "works" for words that fairly readily suggest keywords, and if there is serious doubt about the value of the method for foreign language students, even when such words are used and keywords are supplied, there seems little value in the method for foreign language students in their day-to-day encounters with new words that they may wish to commit to memory. However, a recently completed experiment by Hogben and Lawson(1993) suggests otherwise, and largely prompted the lines of enquiry pursued in the experiment reported in the present paper. Working with third-year high school students of Italian, who were not provided with keywords, but were required to generate their own, Hogben and Lawson showed that, while their "elaborated keyword method" was generally superior to the standard keyword strategy for acquiring new Italian vocabulary items (nouns), students using the standard strategy nevertheless acquired a not insignificant number of words. And, although students found the acquisition of Italian nouns without obvious keywords more difficult than those nouns which suggest keywords fairly readily, they did manage to learn and recall a reasonable number of these words .

The major interest of the present research was the investigation of the overall value of the keyword method for vocabulary acquisition by experienced foreign language students, together with a further examination of the elaborated strategies developed by Hogben and Lawson (1993). To this end, the relative success of students employing standard and "elaborated keyword" strategies was examined, and the performance of these students compared with that of students untrained in keyword methods. Further, exploration of the issue of keyword efficacy with words not having obvious keywords was undertaken. Information was gathered on the pre-experiment vocabulary-learning strategies typically employed by the students in their foreign language studies. The students also reported on the learning and recall strategies that they used in the course of the present research.

Method
Subjects
Thirty-nine male students comprising the two year-10 (third year) Italian classes at an Adelaide, South Australia metropolitan boys secondary school served as subjects for this study. The average age of the students was 15.06 years, with a standard deviation of
Materials and Procedure
The target words (30 Italian nouns) used in the study were the same as those previously employed by Hogben and Lawson (1993) with year-10 students at another school. In that study, it was established that none of the words was known to any of the students beforehand. Two teachers of Italian at the school used in the current study checked this list and agreed that it was highly unlikely that any of the words on the list would be known to the students. The list of 30 nouns comprised 10 with obvious keywords (Obvious), 10 without obvious keywords (Non-obvious), and a group of 10 nouns containing 5 words with the suffix ezza and 5 with the suffix astro (Suffix). The suffix words were included because discussion about suffixes formed a part of the multiple elaboration training for the Multiple Elaboration group. Only the 5 nouns exhibiting the ezza suffix were abstract words; the remaining 25 nouns all had concrete referents. All of the words' English equivalents would be known to English speakers at the education level of the subjects used in the experiment. The 30 Italian nouns are given in Appendix 1.

Prior to their assignment to the three groups used in the experiment, the students completed a vocabulary test (Australian Council for Educational Research Word Knowledge Test; 1989). The scores obtained on this test were used to create matched triads of students, members of which were then assigned at random to either the control group (C), the standard keyword group (SK), or the multiple elaboration group (ME). In the course of the study, three of the thirty-nine students were absent for at least one training session and one student spoiled a test paper. The analyses presented below were thus based on a reduced total of thirty-five subjects: C=12, SK=12, ME=11. The two researchers were responsible for the two training and initial testing sessions with the keyword groups. Each conducted one session with one group and then swapped to the other group for the subsequent session. The two training and initial testing sessions for the control group were conducted by a research assistant who is a fluent speaker of Italian. The subsequent short testing sessions that were carried out were conducted jointly by the two researchers.

The first session lasted 90 minutes (a double school class period), and the three groups were dealt with simultaneously. The students in each group spent the first part of this session completing a short questionnaire in which details were sought on their usual vocabulary-learning methods. In addition, the questionnaire booklet asked the students to identify any
particular kinds of Italian words that they found either difficult or easy to learn and to remember, and to give reasons for this if they were able. From the information provided on this questionnaire, a detailed analysis of students’ usual vocabulary acquisition strategies was carried out. The two questions of interest in the questionnaire were:

1. What could you do when you come across an Italian word whose meaning you do not know, and you decide that this is a word whose meaning you want to learn and remember? Describe in as much detail as you can what you could do to learn and remember that word. If you have more than one method you could use describe each method.

2. Suppose you have looked up the meaning of a new Italian word in your dictionary. What could you do then to help you learn and remember that word? Describe in detail what you could do.

For the training phase that followed, three sets of training manuals were prepared: one for each group. Each student received a copy of the appropriate manual and worked through the manual under the direction of the experimenter. Students were encouraged to ask questions if any points were not clear. This initial training phase lasted approximately 50 minutes. The first three pages of the manuals for the experimental groups (SK and ME) were identical and consisted of an explanation of the keyword method. The initial explanation used the Italian noun coltello and the keyword colt in a manner similar to that presented earlier in this paper. Immediately after this explanation stage, the subjects in each experimental group practised the method on two Italian nouns with obvious keywords: catena and fango. The standard keyword group then practised on a further eleven words, only three of which had fairly obvious keywords. In the manual, the introduction to the vocabulary items without obvious keywords for the SK group, described as "hard words", was as follows:

"With some Italian words it is quite easy to get a keyword; either in English or in Italian. But with others it is much more difficult. Sometimes you just can't think of a keyword, or you can't develop a picture that links the keyword and the meaning.

Do not despair! Many people who use the keyword method find that just the attempt to use Steps 1 and 2 [this referred to the steps we called 'sounds like' and 'develop a picture' in the original explanation] somehow helps them to remember many of these difficult words. The struggle they go through seems to help in remembering words and their meanings. So the advice is: Do not abandon the search for keywords and pictures too soon; try saying the word out loud, try different syllable and letter splits."

This section also appeared in the manual for the ME group, although their practice here was limited to two words.
Preceding the "hard words" section, the manual for the ME group included a section headed "Add On Strategies". The students in this group practised four additional ways of linking words to help them learn and remember vocabulary: Looking for English-Italian word links, using knowledge of grammar to form links, making use of "related words" when consulting a dictionary, and employing known language "rules". The students in this group were specifically taught two "rules" applying to the Italian suffixes ezza and astro. All words containing these suffixes that were used as examples with the ME group during training were also included as examples in the manual for the SK group. The students in this group, however, did not have their attention drawn to the suffix "rules".

The control group students, after they had supplied information on their current vocabulary-learning methods, read through excerpts from an article dealing with certain issues of vocabulary acquisition. The article made no reference to any actual vocabulary-learning techniques. Issues dealt with in the article were discussed with the students for approximately ten minutes, care being taken to avoid any discussion of strategies of vocabulary acquisition. The students then practised individually, learning the same eleven nouns as were worked on by the SK group. The amount of time spent to this point was the same as that for the two experimental groups.

Immediately following the initial phases described above, the students in all three groups learned a practice set of nine Italian nouns (three with obvious keywords, three without obvious keywords, and three "suffix rule" words). An eleven-page learning booklet was prepared for use by the students in each group during this learning practice session. The first page of the booklet contained instructions to be followed and the second page was an example page on which the Italian word panca was detailed. For the two keyword groups only, the keyword "pan" was offered. Following the example page were nine pages each containing a different noun to be learned. In addition to the target Italian words, each page contained a dictionary-like definition and two defined "related words". No keywords were provided. This page layout was adopted as an approximation of conditions likely to prevail either in a classroom with a teacher dealing with methods of deliberate vocabulary acquisition, or in situations in which students encounter unknown words during their private study. Such situations would typically involve use of a dictionary. The order in which the words appeared in the booklet was randomly determined. The example below shows the page layout that was employed.

```
RAMPINO
Mean hook, prong, grapnel.
Also used figuratively to mean excuse,
pretext.
```
Related words:
Rampinate: blow from a hook.
Rampone: hook, harpoon.

Students were allowed 30 seconds for each booklet page.

Immediately following the learning, the students were tested; the task being for them to write down the English equivalents for the Italian nouns. On the test paper the Italian words appeared in a different random order from that in the learning booklet. Following the testing, the students corrected their own test papers, and the results were discussed with them. In each group the students were praised for their generally high level of learning and recall. In the two keyword groups, the experimenters emphasised the power of the keyword method. At the conclusion of this session the students in each group were asked not to discuss events with students from other groups during the period before the next session.

One week later the groups re-assembled for a single class period (45 minutes). With the two keyword groups, the first 15 minutes were spent reviewing the methods learned in the first training session and practising use of these on an additional six new words. In the control group the students discussed the results of their attempts to learn and remember the meanings of the words in the nine word test they had done the previous week. They also discussed which words they found easy or hard to learn and remember. They then practised individually learning the same six new words as the students in the two keyword groups. Following this 15 minute review period, the students in each group were issued with a thirty-two page learning booklet containing the 30 target words. The cover page, the example page, and the 30 pages containing the words to be learned followed exactly the same layout as that employed with the practice word learning booklet described above. In each of the three groups the experimenters read through the instructions and the example page. The students then worked silently through each page, with 30 seconds being allowed for each noun. The pace of learning was strictly controlled by the experimenters. A short time after the completion of this learning phase, the students' immediate recall was tested. Each student was issued with a sheet consisting of instructions followed by the 30 Italian nouns in a random order different from that of the 32-page learning booklet. Ten minutes were allowed for students to recall and write down as many English meanings as they could. The hypotheses to be tested here were: that the two keyword groups would significantly outscore the control group on all measures: the total 30 words, the 10 Italian nouns with obvious keywords, the 10 words without obvious keywords, and the 10 suffix words;
and that the multiple elaboration group would significantly outscore the standard keyword group on the total 30 words, and the 10 suffix words. For the 10 words with obvious keywords, and the 10 with no obvious keywords, no significant difference in level of meaning recall was expected between the two keyword groups.

After a further two weeks the students were re-tested on the 30 target words. This testing was conducted in the students' regular classroom groups. Although it was expected that the overall performance of the students would significantly decline on this re-test (Fuentes, 1976; Willerman and Melvin, 1979; Wang et al, 1992), the hypotheses set up to be tested were identical to those set up for the results of the immediate testing conducted two weeks earlier. In order to determine the extent to which the students in the two keyword groups actually did apply the techniques they had been taught, and to check on the strategies employed by the control group subjects, measures were obtained immediately following the re-testing. On a sheet containing 18 of the 30 target words, the students were asked to write down what they actually did to learn and remember each of the words. The instructions to the students read:

"Listed below are some of the nouns you have just been tested on. Think back to the lesson when we asked you to learn these words in the 30 item booklet. Beside each word say what you did at that time to help you learn and remember that word."

The assumption made here was that the strategy recalled from the initial learning was that used in the subsequent recall task. Time constraint was the major reason for not gathering information from the students on all 30 target words. Nevertheless, reducing the number lessened the likelihood of the students becoming bored with the task. The 18 words used comprised six words chosen randomly from each of the Obvious, Non obvious, and Suffix groups.

Results
Questionnaire Information
In the first session, prior to strategy instruction, students in all groups were asked to answer questions about their usual methods of learning new Italian words. Responses to these questions were first examined to identify distinct methods reported as being used to learn and remember new words. A strategy was recorded whenever the student reported use of a procedure that differed from those reported by other students. Strategy identification was initially undertaken independently by two raters who achieved 83 per cent agreement in this task. Following examination of instances where identifications differed, and specification of the final strategy types, the analysis produced a list of 47 strategies, many of which differed...
only slightly from others in this list. Examination of this list indicated that strategies could be formed into the 13 strategy types shown in Table 1.
While the range of strategy types was of interest here, the major focus of our analysis was on the degree of elaboration evident in the students’ reported methods for vocabulary acquisition, and in particular the frequency of procedures that could be related to the keyword methods used in designing the instructional interventions in this study. For this reason the strategy types were further classified on the basis of the degree of elaboration evident. This classification produced four distinct major categories. Many of the strategies reported involved passive non-elaborative treatment of the new word whereby the word and its meaning were noted, perhaps more than once, but no attempt was made to undertake any further transformation of the word-meaning complex. The second major class of strategies, identified here as active non-elaborative, did involve explicit consideration of the learning and remembering process and use of special procedures that were designed to make that process more effective. However, these did not involve transformation of the word and its meaning in a manner that showed evidence of activity designed to establish interaction of the new word/meaning with existing knowledge. The other two classes of strategy types can be argued to provide evidence of this elaborative processing, though they differ in the degree of that activity. Strategies included in the minor elaboration category did involve analysis of, and allocation of attention to, features of the word for the purpose of establishing interaction between the new word and either the text or existing knowledge. The final strategies showed evidence of substantial elaboration involving transformation of either features of the word or of its meaning or of both.
The results of the analysis of student reports on the strategies used for acquisition of new vocabulary are shown in Table 1. The percentage frequency of use of the four classes of strategy types and of each strategy is shown in the middle column. The rightmost column indicates the numbers of students who reported at least one instance of the strategy class.

Table 1 about here
A first feature of interest in this analysis is the extent and range of the procedures reported by students for use in this task of vocabulary acquisition and use. It is apparent that these students, all of whom have been learning the language at high-school level for at least three years, have developed a wide variety of learning and remembering procedures. As a group, these learners do have access to a number of different strategic
procedures, and 20 of the 39 students show evidence of some strategic activity that moves them beyond the level of simple recirculation or maintenance of information. However, it is also apparent that the majority of strategic activity reported by these students does not involve substantial elaboration of the word-meaning complex. About 75 per cent of the strategic activity reported by students does not appear to involve any systematic elaborative activity that might be expected to establish strong links between the new vocabulary item and existing knowledge in a relatively short period of time. As such it can be predicted that use of these activities would be associated with the need to undertake repeated attempts to acquire the meaning of the new word. Over half of the strategic activity reported was of a passive, non-elaborative nature involving inspection and recording of the word and various forms of rehearsal. For ten of the students this was the only form of activity reported, with their usual approach involving reading of the dictionary entry and repetition of the word and meaning. One student noted that when other strategies could not be used “the only other option is to learn by brute memorisation.” Another reported that “I just will my self to remember.” Although a further nine students reported that they were active in arranging testing of memory for words and meanings, this activity did not appear to involve any explicit attempt to transform the word-meaning complex in a way that sought either to establish links with existing knowledge or interaction between the word and the meaning. For these 19 students there was no evidence of explicit elaborative activity.

Five students did report use of procedures that can be seen to involve formation of relationships between the new word and existing knowledge; either through use of surrounding material in a text or by classification of the new word upon some basis, such as: “grouping words of various types”, or “putting words into types, such as grouping irregular verbs.” This type of elaboration accounted for 11 per cent of the total strategic activity.

Only 14.6 per cent of the reported strategic activity involved substantial elaboration, although this was evident in the reports of 15 of the 39 students. Most of this elaboration involved the seeking of relationships between the new word and known English or Italian words. Eight students reported use of techniques that attempted to establish physical or semantic relationships with known words. Five students indicated that they develop sentences in which the new word might be embedded. Included in this category of procedures were those that related directly to one of the components of the keyword method. These were of three different types. Six students reported that they focussed on the sound of the new word and tried to relate this sound to that of a known English or Italian word. “I try and find any similarities to an English word, so when I’m shown the
meaning I immediately think of the English word the Italian word sounds like.” One student reported an extension of this strategy to include a link between the meaning and sound of the new and generated (key) word, so that he could “try to find any English words within the Italian word which could give me a hint about the meaning; for example, ‘braccio’ sounds like ‘branch’ and an arm is like a branch.” Only one student reported use of a procedure that involved picturing of the word. Of the 39 students, only eight reported use of strategies that could be seen as directly related to components of the standard keyword method: none employed the full (two-step) keyword method.

Cued recall
As previously explained, students completed two cued recall tests in which they were given the target Italian words and required to provide definitions. The first testing occurred immediately after the completion of training, the second after a two-week delay. For total 30-word recall, and for recall of each of the three item sub-types (Obvious, Non-obvious, Suffix), two scores were derived: a strict score and a lenient score. The strict score total indicates the number of words for which correct definitions were supplied. A score of 1 was awarded where at least one correct definition was provided for a given word; otherwise a score of 0 was recorded. For this scoring, approximations or closely related definitions were not accepted. In the lenient scoring, credit was given for approximations and close relations, and for providing more than one definition where this was possible (not all the Italian nouns had more than one definition). Three points were assigned if two or more correct definitions were recalled. Students received two points for a single correct definition, and one point for a reasonable approximation of the definition supplied for a given word in the learning booklet. A student was awarded one point, for example, if the word sdraia was defined as "chair" instead of "deck chair". Adjectives, such as "soft" rather than the noun "softness" (for mollezza), were also awarded one point. Lenient assessment, it can be argued, is a more sensitive assessment of acquisition and thus perhaps more appropriate in the study of classroom vocabulary acquisition. All scoring was carried out by one of the experimenters, care being taken to ensure that it was completed in ignorance of student group membership.

The details of the performance on these two occasions are shown in Table 2, which includes both total recall score(strict and lenient scores) and the scores for recall of items in the three different item sub-types(lenient score). Differences in recall for the three groups were examined using a series of two-way (Group x Occasion) repeated measures analyses of variance, with
the occasion of the test (Immediate vs Delayed) constituting the repeated measure. Separate analyses were carried out for total score, and for each of the three item subtypes. The analyses discussed here are those using the lenient scoring procedure. The results of these analyses are set out in Table 3. The strict score, which correlated \( r = .98 \) with the lenient score on both test occasions, is included in Table 2 to provide a clear indication of the total number of words recalled in the two conditions.

From the mean scores displayed in Table 2 it can be seen that on all four measures of immediate post-learning performance the ME group outscored the SK group, which in turn outscored the control group. Repeated measures analysis of variance and simple effects analyses revealed that for the Total score and for Suffix items the superiority of both the SK and the ME groups over the control group was statistically significant (see Table 3). For Obvious and Non-obvious items while the difference in level of recall between the ME and C groups was again significant, the difference between the SK and C groups was not significant. Simple effects analyses indicated that, although both keyword groups outscored the control group on all measures, the superiority of the ME group over the SK group was not sufficiently large to be statistically significant.

The higher level of recall of the ME group compared to that of both the SK and control groups holds also for the results of the delayed testing. The SK group also outscored the control group on all delayed tests except for the Non-obvious words; here the mean scores were the same. None of these differences at time of the delayed test was, however, statistically significant. The obtained significant Occasion main effects and the significant Group x Occasion interaction effects, indicated in Table 3, are clearly suggested by an examination of the means for immediate and delayed recall. Two weeks following the initial learning and testing the definition recall of all groups had dropped very noticeably, and simple effects analysis revealed that no significant differences were present among the groups. So, the very substantial effect evident immediately after training - representing an effect size of \( 0.87 \) - was considerably reduced after the two-week delay.

The statistically significant superiority on Total score and Suffix score of the ME and SK groups over the control group on immediate test results supported two the hypothesis set up at the outset. However, although the level of performance on
Obvious and Non-obvious was in the predicted direction the differences between SK and C groups on these measures was not significant. In addition the expected superiority of the ME group over the SK group on Total score and Suffix score did not occur. The expected statistically significant superiority of the keyword groups over the control group on the delayed tests also did not eventuate.

The design of this study allowed us to compare the level of performance of the groups on Obvious and Non-obvious words. Analyses of these performances showed a superiority for Obvious items over Non-obvious items for all three groups. An examination of Table 2 shows that this superiority held both for the Immediate and the Delayed testing. Two-way repeated measures analyses of variance (Group x Type), produced statistically significant word type effects (p<.01) on each of the testing occasions: Immediate, F (1,32)=40.68, MSe=9.03; Delayed, F (1,32)=12.69, MSe=4.59. These Type effects were significant for each of the three groups on the Immediate test, but only for the SK group on the Delayed test.

Debriefing following testing
The written reports from students in each of the groups on methods used to learn 18 of the target words, given following the delayed testing, were examined for evidence of keyword strategy use. Reports provided by control students were closely aligned to those given in their descriptions of procedures used to learn new words gathered prior to training. While some of these subjects reported use of acoustic similarity between target words and known English or Italian words, none reported use of either imagery alone or the full keyword procedure. The most frequently reported strategies were again based on some form of repetition. In contrast, 11 of the 12 SK students, and all of the ME students, reported use of the full keyword procedure though not for all words. Two students in the ME group reported use of 'add on strategies'.

Discussion
While the present study does not replicate any one of the three previous studies reporting on the value of the keyword method for vocabulary acquisition by students studying foreign languages (Fuentes, 1976; Levin et al, 1979; Willerman and Melvin, 1979), our findings certainly suggest that there may be room for some considerable optimism for the method in this sphere. In the first place, students using keyword methods significantly outsored students using their customary vocabulary-learning strategies when the task was to learn the meanings of Italian nouns and to recall the meaning definitions when cued with the words immediately following the learning. This superiority
prevailed for the total 30 nouns, and for Suffix words. The findings that the SK and C groups did not differ significantly on Obvious and Non-obvious items were not as predicted. While the pattern of performance was in the expected direction the effect of training on these words was not as strong as on the Suffix items.

When tested after a delay of two weeks, the superiority remained with the keyword groups over the control group, though the differences were not significant. In fact the decline in number of words recalled across all categories was large, and reflects the findings reported by Willerman and Melvin (1979). The somewhat sharper decline for the keyword groups was reflected in the significant Group x Occasion interactions. The fact remains, though, that in the present study significant advantages have been obtained for keyword methods when the criterion is immediate recall of word definitions. Despite overall score decline after a two-week delay, the advantage still remained with the keyword groups; although the superiority was not marked enough to be significantly significant.

The decline in recall over the period of the delay is disappointing. Two possible causes of this effect, found now in several studies, seem worthy of further investigation. First, it is possible that students do not readily incorporate the new method for learning vocabulary into their normal classroom activities. It is possible that the method, although interesting and obviously useful, nevertheless is seen by students as associated mainly with some special class project; something associated with research on foreign language learning and not clearly tied to normal language lessons. Students in our study may have found it unusual, for example, that no practice or method revision occurred in the two weeks between the two 30-word testing sessions. It would be rather strange for this not to occur under normal classroom teaching conditions. In addition, it is possible that the full power of the trained method is not effectively exploited by students, especially at the time of retrieval of the word meanings. While the instructions provided to the SK and ME groups in the present study did indicate that the two components of the procedure should be used at the time of retrieval, no focussed practice was given to reinforce this feature. Again such practice would probably occur in regular classroom routines. In future studies involving keyword methodology, we plan to incorporate instruction and practice on word retrieval into our instruction manuals.

As previously noted, Fuentes (1976) and Levin et al (1979) have suggested that the lack of clear keyword superiority observed in their studies was probably due, in large part, to the use by control subjects of techniques closely resembling the keyword method; or other effective vocabulary acquisition strategies. While the students in the present study reported that they employed a considerable variety of vocabulary-learning
procedures, no students reported using techniques closely resembling the standard keyword method. In fact, approximately 75 per cent of the reported strategies appeared not to involve any significant elaboration that might lead to strong links between existing knowledge and words-to-be-learned. Indeed, only 14.6 per cent of the reported strategic activity could be said to involve substantial elaboration. Seven students reported some focusing on the sound of new words and attempts at relating this to sounds of known English or Italian words. None, however, employed any visual imaging. Only one student reported such imaging and he did not link this with word sound. So, only 8 students (20.5 per cent) reported strategy use that related directly to some component of the keyword method: no student reported using the fully-developed technique. No group differences could be detected among the students assigned to the control and the two keyword groups.

Taken together, the information provided by students on their usual strategies, and the patterns of recall performance, indicate the need for modification of the suggestions made on this point by earlier researchers. Clearly these experienced foreign language learners do have access to a number of varied strategies during their attempts to acquire new word meanings. However, it is also clear that having a variety of strategies does not mean that these strategies are what Fuentes(1976) termed effective. The range of strategies available to all three groups in this study prior to training was similar. Yet when provided with relatively brief training in the use of an elaborative strategy, the level of performance of the two experimental groups was elevated significantly compared with that of the uninstructed control group. This suggests a need to reconsider the level of strategic sophistication of these language learners.

Another possible factor contributing to the difference in the pattern of results between the present study and those reported from the United States, could be general acquaintance with the keyword technique. The regular classroom teacher of Italian responsible for the students used in our study was not familiar with the keyword method, and the same might be assumed of the students' previous teachers. Perhaps the keyword method is better known in the United States than in Australia, and foreign language teachers may acquaint their students with it; or the students may learn it elsewhere and transfer it to their vocabulary acquisition tasks.

It is also apparent that the different studies related to this area have differed in their procedures and in materials chosen for students. In the Fuentes (1976) study, students were allowed an average of two minutes to learn each vocabulary item (five new Spanish words were presented to be learned in each of a series of 10-minute daily learning sessions). Levin et al (1979) point out
that, because of this generous learning time allocation, it is likely that both control group and keyword group subjects were actually very close to ceiling after the initial learning and, hence, the lack of difference between them on the subsequent tests might have been expected. In their subsequent study, Levin et al. required their high school student subjects to learn Spanish words at an average rate of 12 seconds per word. In two of the four experiments the students were required to learn 50 words; in the other two, 32 words. No overall keyword superiority was revealed in these experiments: perhaps the 12 seconds per word may not have been sufficient for the students fully to exercise/implement the keyword method, even though keywords were supplied to the learners. In the other keyword study with foreign language learners, to which we have previously referred (Willerman and Melvin, 1979), students were allowed only 8 seconds per word both for learning and for later recall. Our feeling is that this amount of time is far too short. In our previous study (Hogben and Lawson, 1993) we allowed 30 seconds per word for learning, and the students reported this as being sufficient; at least with words suggesting obvious keywords. As previously noted, 30 seconds per word were allowed for learning in the present study.

In their comprehensive review of the keyword method, Pressley et al. (1982) were severely critical of the sketchy instructions given to their students by Willerman and Melvin (1979) in their study. Pressley et al. were also critical of the fact that no provision was made for any student practice, nor any opportunity for the students to ask questions during the learning phase. These criticisms appear valid, particularly that directed at the instructions. By our count, the instructions, which were simply read to the subjects, consisted of just 140 words. In comparison, the directions given to the keyword subjects in the Levin et al. (1979) experiments with high school subjects appear generous: approximately 650 words. However, in the Levin et al. experiments only 10 minutes were allowed for the students to read through the instructions. It seemed to us that considerably more attention may need to be paid to explaining and clarifying the method to students, allowing them time to practise, and giving the opportunity to ask questions and offer suggestions. And so, in the present study about 50 minutes were devoted to this explanation, practice, and discussion prior to any list learning. In the following session, a period of approximately 15 minutes was set aside for method revision before the students learned the 30-word list. Further practice occurred in the first session with the learning and testing of the trial 9-word list. The training manual used in the first session by our SK group (the group most comparable to the keyword groups in the above studies) contained 1460 words. This total included practice instructions,
practice words and definitions.
The time allocated for learning the methods and practising, and
the manner in which the sessions were conducted, did, in our
view, approximate reasonably closely to regular classroom
practice. And the students in the present study were required to
generate their own keywords for the 30 target words. This too,
it seems, more closely parallels the "real life" situation than
does supplying keywords to students for words to be learned.
The information provided to us by the students immediately
following the delayed test about the word-learning strategies
they employed, showed quite clearly that the majority of students
in the two keyword groups did in fact use the keyword method.
This is certainly some support for our claim that our
instructional procedures had made an impact. The responses from
the control group students showed no evidence of any keyword use.
This is consistent with the information obtained from all
students at the outset about their vocabulary-learning
techniques, and, further, it suggests that little or no method
contamination occurred in the course of the study.
We noted above that a quite unexpected result occurred in the
study; namely, all three groups recalled significantly more word
definitions for Italian nouns containing obvious keywords than
they did for words without obvious keywords at the time of the
Immediate test. While this Obvious keyword superiority was
expected for the two keyword groups, it is not easy to explain
for the control group (who reported no focusing on keywords). In
developing the word list, an effort was made to "equate" the
words in the two sub-type groups; that is, to produce lists of
near equal difficulty. Apart from selecting words that would be
unknown to the students, the lists were balanced for object
familiarity and for number of syllables. While no word length
(number of letters per word) balancing was specifically carried
out, a letter count shows that the average word length of the
Non-obvious keyword vocabulary items was slightly less than that
of the Obvious items (5.5 versus 5.9 letters). The lists were
not balanced for beginning letter (vowel or consonant), and this
could have some significance for ease of learning. An
examination of the lists shows that no word in the Obvious group
commences with a vowel, whereas four words in the Non-obvious
list do. It may also be significant that one word in the Non-
obvious list (sdraia) begins with the letters "sd": a beginning
combination not found in English. Another word (ghirba) begins
with the fairly rare letter combination "gh". This may have
contributed to list difficulty.

Although a number of issues yet remain to be resolved, the
results from the present study into deliberate vocabulary
acquisition by foreign language students suggest that further
research into the efficacy of elaboration strategies - including
the keyword method - is warranted. Experienced foreign language
students can benefit from instruction in use of elaborative strategies. The results of this investigation suggest that the instruction did provide the students with ways of combining strategies that had not previously been used. However the short-lived nature of the effect suggests that further investigation into the ways students use the strategy at the time of retrieval is required.

References


Journal of educational psychology, 84, 520-528.


Table 1. Students’ reported vocabulary acquisition strategies.

<table>
<thead>
<tr>
<th>Strategy type</th>
<th>Frequency of strategy(%)</th>
<th>Number of students with at least 1 instance of strategy type(n=39)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive</td>
<td>53.4</td>
<td>10</td>
</tr>
<tr>
<td>Read and look-up</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>Write</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Repetition</td>
<td>17.4</td>
<td></td>
</tr>
<tr>
<td>Assistance</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Memorise</td>
<td>7.6</td>
<td></td>
</tr>
<tr>
<td>Active non-elaborative</td>
<td>20.9</td>
<td>9</td>
</tr>
<tr>
<td>Test</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>Special attention</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>Segment</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>Active - minor elaboration</td>
<td>11.1</td>
<td>5</td>
</tr>
<tr>
<td>Context</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>Classification</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>Active - substantial elaboration</td>
<td>14.6</td>
<td>15</td>
</tr>
<tr>
<td>Association</td>
<td>11.1</td>
<td></td>
</tr>
<tr>
<td>(Keyword-related)</td>
<td>(5.6)</td>
<td>(8)</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>3.5</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. ANOVA results for Total Score and Word Subtype analyses.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Effect</th>
<th>df</th>
<th>F1</th>
<th>MSe</th>
<th>Group comparison (Immediate test)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group</td>
<td>Degrees of Freedom</td>
<td>Mean Squared</td>
<td>F</td>
<td>Group Significance</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>--------------------</td>
<td>--------------</td>
<td>---</td>
<td>--------------------</td>
</tr>
<tr>
<td><strong>Total score</strong></td>
<td>2,32</td>
<td>1,32</td>
<td>11.44</td>
<td>86.15</td>
<td>ME, SK &gt; C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occasion</strong></td>
<td>1,32</td>
<td></td>
<td>215.07</td>
<td>40.16</td>
<td></td>
</tr>
<tr>
<td><strong>Group x Occasion</strong></td>
<td>2,32</td>
<td></td>
<td>9.85</td>
<td>40.16</td>
<td></td>
</tr>
<tr>
<td><strong>Obvious</strong></td>
<td>2,32</td>
<td>1,32</td>
<td>6.54</td>
<td>18.29</td>
<td>ME &gt; C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occasion</strong></td>
<td>1,32</td>
<td></td>
<td>168.23</td>
<td>8.56</td>
<td></td>
</tr>
<tr>
<td><strong>Group x Occasion</strong></td>
<td>2,32</td>
<td></td>
<td>5.37</td>
<td>8.56</td>
<td></td>
</tr>
<tr>
<td><strong>Non-obvious</strong></td>
<td>2,32</td>
<td>1,32</td>
<td>6.44</td>
<td>15.68</td>
<td>ME &gt; C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occasion</strong></td>
<td>1,32</td>
<td></td>
<td>161.47</td>
<td>4.32</td>
<td></td>
</tr>
<tr>
<td><strong>Group x Occasion</strong></td>
<td>2,32</td>
<td></td>
<td>8.91</td>
<td>4.32</td>
<td></td>
</tr>
<tr>
<td><strong>Suffix</strong></td>
<td>2,32</td>
<td>1,32</td>
<td>7.93</td>
<td>15.47</td>
<td>ME, SK &gt; C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occasion</strong></td>
<td>1,32</td>
<td></td>
<td>102.31</td>
<td>7.97</td>
<td></td>
</tr>
<tr>
<td><strong>Group x Occasion</strong></td>
<td>2,32</td>
<td></td>
<td>7.88</td>
<td>7.97</td>
<td></td>
</tr>
</tbody>
</table>

1. For all F values p values < .01

Appendix 1. Items in target list grouped by item type.

**Obvious items**

Cinta
Draga
Fante
Gazza
Lapide
Lastrico
Matassa
Palude
Ramazza
Tasto

**Non-obvious items**

Argine
Elica
Erta
Fabbro
Fascio
Foschia
Ghirba
Gorgo
Orma
Sdraia
Suffix items
  Furbastro
  Giovinastro
  Grettezza

  Medicastro
  Mitezza
  Mollezza
  Rozzezza
  Salmastro
  Scaltrezza
  Sordastro