Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition

Jan H. Hulstijn  
University of Amsterdam

Batia Laufer  
University of Haifa

EFL learners in two countries participated in two parallel experiments testing whether retention of vocabulary acquired incidentally is contingent on amount of task-induced involvement. Short- and long-term retention of ten unfamiliar words was investigated in three learning tasks (reading comprehension, comprehension plus filling in target words, and composition-writing with target words) with varying “involvement loads”—various combinations of need, search, and evaluation. Time-on-task, regarded as inherent to a task, differed among all three tasks. As predicted, amount of retention was related to amount of task-induced involvement load: Retention was highest in the composition task, lower in reading plus fill-in, and lowest in the reading. These results are discussed in light of the construct of task-induced involvement.

Jan H. Hulstijn, Department of Second Language Acquisition; Batia Laufer, Department of English Language and Literature.

We would like to thank our colleagues Jean-Marc Dewaele (Birbeck College, London) and Alex Housen (Vrije Universiteit, Brussels), the three anonymous reviewers, and Language Learning’s editor, Nick Ellis, for their fruitful comments on earlier versions of this paper.

Correspondence concerning this article may be sent to Batia Laufer, Department of English Language and Literature, University of Haifa, Haifa 31905, Israel. Internet: batialau@research.haifa.ac.il, or to Jan H. Hulstijn, Department of Second Language Acquisition, University of Amsterdam, Spuistraat 134, 1012 VB Amsterdam, The Netherlands. Internet: hulstijn@hum.uva.nl
Learning a foreign or second language to intermediate and advanced levels of proficiency involves the acquisition of many thousands of words. Teachers and learners alike would like to know in which ways instructional programs might foster the acquisition of so many words. One essential question is whether instructional tasks can be classified in terms of their vocabulary-learning effectiveness. This paper aims to provide a partial answer to this question. It reports on a comparison of three tasks in terms of their effectiveness for vocabulary acquisition. The selection of the tasks was based on the notion of task-induced involvement and the Involvement Load Hypothesis of Laufer and Hulstijn (2001). In this introduction, we first turn our attention to the psychological memory literature and then present the Involvement Load Hypothesis on which our research question was based.

Levels of Processing and Elaboration

What do we know about the processes that foster memory of explicit information (facts, events, reasoning, names, words)? In an attempt to answer this question, Craik and Lockhart (1972), in a seminal paper, launched the concept of depth of processing, meaning that the chance of some piece of new information being stored into long-term memory is not determined by the length of time that it is held in short-term memory but rather by the shallowness or depth with which it is initially processed. Craik and Lockhart further postulated several levels of processing depth. For instance, processing the meaning of a new lexical item takes place at a rather deep level, whereas processing the phonological form takes place at a rather shallow level. Craik and Tulving (1975) suggested that what is critical to retention is not simply the presence or absence of semantic encoding (depth of processing; a qualitative construct), but also the richness with which the material is encoded (spread or elaboration of encoding, a quantitative construct). Craik & Lockhart’s levels of processing theory was challenged by Baddeley (1978), Eysenck (1978), and Nelson (1977). The main points of criticism were concerned with
the following two questions: (1) What exactly constitutes a level of processing, and (2) How do we know that one level is deeper than another? In a response to these questions, Lockhart and Craik (1978) acknowledged that the terms depth and spread/elaboration lack operational definitions and independent indices, and thus that circularity is inherent in the levels-of-processing approach. They also acknowledged the importance of investigating the effect of encoding tasks and processes in conjunction with the effect of retrieval tasks and processes. Lockhart and Craik emphasized, however, that their approach should be judged on its heuristic value as a framework for steering new research that will generate data to outlast any theoretical account.

A major obstacle facing subsequent proposals to replace depth of processing with any notion, be it encoding specificity, distinctiveness of encoding, degree of elaboration, cognitive effort, degree of richness, etc., resides in the difficulty of providing unambiguous, operationable definitions (for detailed reviews, see Anderson, 1995, chapter 6; Baddeley, 1997, chapter 7; Eysenck, 1982; Zechmeister & Nyberg, 1982). Yet, it is generally agreed that retention of new information depends on the amount and the quality of attention that individuals pay to various aspects of words. Rich (qualitative) and numerous (quantitative) associations with existing knowledge (e.g., in the form of establishing similarities and contrasts between old and new information) increase the chances that the new information will be retained. The overall term most commonly used in the literature to refer to these processes is elaboration in both its qualitative and quantitative sense.

Cognitive psychologists appear to agree that processing new lexical information more elaborately will lead to a better retention than if it had been processed less elaborately. This position can be summarized with the following words of William James (1890, p. 662), quoted by Baddeley (1997, p. 119): “All improvement of the memory lies in the line of elaborating the associates.” In practice, this means that if learners pay careful attention to the word’s pronunciation, orthography, grammatical category, meaning, and semantic
relations to other words, they are more likely to retain the word (i.e., the link between at least one representation of the word’s form and at least one of its meanings) than if they pay attention to only one or two of the above word properties. Processing can be more or less elaborate, irrespective of whether vocabulary is learnt incidentally or intentionally. In other words, careful attention can be paid to the properties of a certain word during intentional learning (e.g., preparation for a vocabulary test) just as well as during incidental learning (e.g., when the word occurs in a text and successful completion of the reading task requires such attention).

Empirical research on tasks and incidental acquisition points to a variety of factors conducive to successful retention of words. Words whose meanings were correctly inferred during a reading task were remembered better than words explained by a synonym (Hulstijn, 1992). Words that were looked up in a dictionary during a reading task were remembered better than words that were not looked up (Cho & Krashen, 1994; Knight, 1994; Luppesku & Day, 1993), or than words that were glossed in text margin (Hulstijn, Hollander, & Greidanus 1996). However, the presence of marginal glosses was found to enhance vocabulary retention, when compared to the absence of marginal glosses (Hulstijn, 1992; Hulstijn, Hollander & Greidanus, 1996; Jacobs, Dufon, & Fong, 1994; Watanabe, 1997). Words that were negotiated during communicative activities were retained better than nonnegotiated words (Ellis, Tanaka, & Yamazaki, 1994; Newton, 1995). Words that were used in productive tasks, particularly those used in original context, were remembered better than words practiced in nonproductive tasks (Ellis & He, 1999; Hulstijn & Trompetter, 1998; Joe, 1995, 1998). Words practiced in a series of vocabulary-focused exercises following a reading task led to a better retention than words that received additional exposure in texts (Paribakht & Wesche, 1997). In explaining the superiority of one task over another, most of these authors suggested that the more effective task required a deeper level of processing of the new words than the other task.
Yet the concepts of deep processing or elaboration are hard to formalize and operationalize. Let us take, for example, three simple tasks learners can do with the word “assess”: (a) look up its meaning in a dictionary and write a sentence with the word; (b) look up its meaning and explain the difference between “assess,” “evaluate,” and “estimate;” and (c) try to infer its meaning from a sentence context when four alternatives are presented by the teacher. There are no definite criteria by which we could grade the three tasks in terms of the depth of processing they require and know which task is more effective than another. However, research on task effectiveness would require the identification of criteria that could be observed, manipulated, and measured.

The Involvement Load Hypothesis

Acknowledging the importance of the notions of depth of processing (Craik & Lockhart, 1972) and elaboration (Craik & Tulving, 1975), but feeling the need to translate and operationalize such general cognitive notions in terms of L2 vocabulary learning tasks, we developed our Involvement Load Hypothesis for L2 vocabulary learning (Laufer & Hulstijn, 2001), which we will describe briefly here. We proposed a motivational–cognitive construct of involvement, consisting of three basic components: need, search, and evaluation. The need component is the motivational, noncognitive dimension of involvement. Two degrees of prominence were suggested for need: moderate and strong. Need is moderate when it is imposed by an external agent. An example is the need to use a word in a sentence that the teacher has asked for. Need is strong when it is intrinsically motivated, that is, self-imposed by the learners, for instance, by the decision to look up a word in an L1–L2 dictionary when writing a composition. Search and evaluation are the two cognitive dimensions of involvement, contingent upon allocating attention to form–meaning relationships (Schmidt, 1994, and forthcoming). Search is the attempt to find the meaning of an unknown L2 word or the attempt to find the L2 word form expressing a concept (e.g., trying to find the L2
translation of an L1 word) by consulting a dictionary or another
authority (e.g., a teacher). Evaluation entails a comparison of a
given word with other words, a specific meaning of a word with its
other meanings, or comparing the word with other words in order
to assess whether a word does or does not fit its context. For
example, when a word looked up in a dictionary is a homonym (e.g.,
bank of a river, or bank as a financial institution), a decision has
to be made about its meaning by comparing all its meanings
against the specific context and choosing the one that fits best. The
kind of evaluation that entails recognizing differences between
words (as in a fill-in task with words provided), or differences
between several senses of a word in a given context, is referred
to as moderate. Evaluation that requires a decision as to how
additional words will combine with the new word in an original
(as opposed to given) sentence or text is referred to as strong
evaluation.

Each of the above three factors can be absent or present when
processing a word in a natural or artificially designed task. The
combination of factors with their degrees of prominence consti-
tutes involvement load. Consider an example of two tasks that
vary in involvement load. In task one, the learner is asked to write
original sentences with some new words and these words are
translated or explained by the teacher. The task induces a moder-
ate need (imposed by the teacher), no search (the words are
glossed) and strong evaluation because the new words are evalu-
ated against suitable collocations in learner-generated context. If
we want to describe the task in terms of an involvement index,
where absence of a factor is marked as 0, a moderate presence of
a factor as 1, and a strong presence as 2, then the involvement
index of the task is 3 (1 + 0 + 2). In task two, the student has to
read a text and to answer comprehension questions. New words,
which are relevant to the questions, are glossed. The task will
induce a moderate need to look at the glosses (moderate because
it is imposed by the task), but it will induce neither search nor
evaluation. Its involvement index is 1. Hence, task one induces a
greater involvement load than task two.
The basic contention of the Involvement Load Hypothesis is that retention of unfamiliar words is, generally, conditional upon the degree of involvement in processing these words. In other words, it is conditional upon who has set the task, whether the new word has to be searched, and whether it has to be compared, or combined with other words. The greater the involvement load, the better the retention.4

To test this contention, we conducted two parallel experiments in two countries. We consider them parallel because we used identical methodologies and comparable participants, as described below. We designed tasks with different involvement loads and compared their effect on incidental vocabulary learning.

Method

Aim

The aim of the two experiments was to investigate the effect of involvement load on the retention of ten English words by young adult EFL learners in Israel and the Netherlands in an incidental learning setting. We wanted to test the assumption that tasks with a higher involvement load will be more effective for vocabulary retention than tasks with a lower involvement load.

Target Words

Ten low-frequency words and expressions were selected for investigation: rigmarole, wrath, grist, not one whit, sanitise, privy to, morally derelict, curb, inflammatory, deeply ingrained. The words were taken from a text we decided to use for the study, and they were expected to be unfamiliar to our participants, as will be described below. We investigated learners’ retention of the meaning of these words, that is, receptive knowledge only.
The Tasks

Three tasks with different involvement loads were selected for investigation and incorporated in three experimental conditions. Each task was administered to different groups of students.

Task 1: reading comprehension with marginal glosses. Students in group 1 were provided with a text and a set of ten multiple-choice comprehension questions. The text and its questions were taken from the national reading comprehension exam in the highest stream of the Dutch high school system, produced by CITO (the national educational testing institute of the Netherlands). The text was a 621-word letter to the editor of the British paper The Listener (March 1, 1984) on a proposal of some Members of Parliament in the UK to amend the Obscene Publications Act. We highlighted the ten target words, whose understanding was relevant to the task, by putting them in bold print, and we glossed them in L1 in the margin of the text. The task of the students was to read the text and answer the ten comprehension questions. In terms of involvement load, Task 1 induced moderate need, but neither search nor evaluation. Its involvement index was 1.

Task 2: reading comprehension plus “fill in.” Students in group 2 were given the same text and the same questions as those in group 1. For group 2, however, the ten target words were deleted from the text, leaving ten gaps numbered 1–10. The ten target words, along with five words that had not appeared in the original text, were printed in random order as a list on a separate page, with their L1 translations and L2 explanations (see below). The task was to read the text, fill in the ten gaps with the missing words from the list of 15 words, and answer the comprehension questions. In terms of involvement load, Task 2 induced moderate need, no search, and moderate evaluation, because the context was provided. Its involvement index was 2.

Task 3: writing a composition and incorporating the target words. Students in group 3 were asked to write a composition, in the form of a letter to the editor of a British newspaper, on the occasion of the proposal of some Members of Parliament to amend
the Obscene Publications Act. In this letter they had to argue for or against censorship of video films. The instructions to the learners ended as follows: “In your letter, YOU MUST USE THE FOLLOWING TEN WORDS. You may decide yourself in which order you will use them. Explanations of the words and examples of usage are given below.” The same ten words were given as in the Reading and Reading plus Fill-in tasks. In terms of involvement load, Task 3 induces a moderate need, no search, and a strong evaluation, because words had to be used in an original context. Its involvement index was 3.

Here is an example of the information provided for the word wrath:

WRATH (noun, uncountable)
Strong fierce anger
Example: The wrath of the opponents to the proposed bill.
Dutch = gramschap Hebrew = צע (za’am)
The last line of the entry contained the L1 translation, Dutch in The Netherlands and Hebrew in Israel.

Because of the differences in the involvement loads of the investigated tasks, we formulated the following hypothesis.

Hypothesis

The retention scores of the ten target words will be highest in Task 3 (composition with incorporated target words); lower in Task 2 (reading comprehension plus filling in target words); and lowest in Task 1 (reading comprehension with marginal glossing for target words).

Participants

The participants were six intact groups of advanced university learners of English as a foreign language in the Netherlands and Israel, three groups in each country, one group randomly
assigned to each task in each country. The three groups in the Netherlands were parallel groups of first-year English majors. The three groups in Israel were parallel classes taking the most advanced-level course in English for academic purposes. Even if the precise level of English may not have been identical between students in the two countries, the groups in both countries were competent in performing the experimental tasks. The reading groups could answer the comprehension questions, and the writing groups could write a composition. Most importantly, the target words were new to all the subjects.5

The overall number of students taking part in the study was 97 in the Netherlands and 128 in Israel. Fewer students, however, were available for the delayed post-test: 87 in the Netherlands and 99 in Israel. Because the tests were not announced in advance (owing to the incidental-learning nature of the study), we could not ensure all the students who took part in the study would be present at the delayed test session.

Procedure and Tests

Before the experiment proper, we assessed the likelihood of target-word familiarity among our students. A group of high-proficiency university EFL majors in each country, who did not take part in the experiment proper, were given the list of ten words and asked to translate or explain them. The average mean of knowledge was lower than 1 out of 10. As the students in the experimental groups were of similar or lower language proficiency, we assumed that most of the target words would indeed be unfamiliar to them. However, as will be described in the next paragraph, preknowledge of the experimental groups was also checked at the end of the experiment.

All tasks were administered during normal class time. The instructors who introduced the tasks were blind to the purpose of the experiment. None of the tasks was presented as a vocabulary learning task. Time on task was different for all three tasks. Tasks 1, 2, and 3 took about 40–45, 50–55, and 70–80 minutes,
respectively. Task 2 required more time than Task 1, because it included an additional fill-in subtask. Task 3 (composition) was the most demanding one and therefore the most time consuming. It is often argued that time on task should be kept identical in research on task effectiveness. Yet we regard time on task as an inherent property of a task, not as a separate variable. Since our study aimed to investigate retention effects of tasks, we made no attempt to control for time on task, thereby accepting the likelihood that time on targets would covary with time on task.

After the completion of the task, the work sheets were collected, and the students were unexpectedly given a list of the ten target words (word forms). They were asked to provide the L1 equivalents or English explanations for these words. In addition, they were asked to indicate whether they had known the words prior to the task. This was our additional check for preknowledge. The test forms were collected and not returned to the students. One week later in the Netherlands and two weeks later in Israel, the students received the same test again.6

The tests were scored by researchers and student assistants. A word that was not translated or was wrongly translated received a score of zero. A correct response received a full point. A “semantically approximate” response received half a point. The researcher and the assistant negotiated the degree to which the response was semantically approximate or not. If the learner had a correct response but had also marked the target word as known prior to the experiment, the response was scored as zero. Such preknowledge happened rarely, and most target words were unfamiliar to most students.

Results

Retention scores of the 87 and 99 students who had been present at both the immediate and the delayed post-tests in the Dutch–English and Hebrew–English experiments are displayed in Table 1.
The figures in Table 1 show that in both experiments, performance in the writing groups was higher than that in the reading plus fill-in groups, which, in turn, was higher than that in the reading groups.

Retention scores were submitted to 3 × 2 analysis of variance (ANOVA), with task as the between-subject factor (reading, reading plus fill-in, writing) and time (immediate and delayed post-tests) as the within-subjects factor. The measure of effect size is $\eta^2$, expressing explained variance.\(^7\)

In the case of the Dutch–English experiment, a significant task effect \(F(2, 84) = 11.50; p < .001; \eta^2 = .22\] was obtained and, obviously, a significant time effect \(F(1, 84) = 73.58; p < .001; \eta^2 = .47\], as well as a significant task × time interaction \(F(2, 84) = 3.89; p < .05; \eta^2 = .084\]. The significant interaction of task and time may have resulted from floor effects in the delayed test for the reading and reading plus fill-in groups, but not in the writing group. The first two tasks resulted in little retention; the composition task fared significantly better.

A Student-Newman-Keuls posthoc multiple-range test was selected as an option in the ANOVA just described. The test

<table>
<thead>
<tr>
<th></th>
<th>Immediate</th>
<th></th>
<th></th>
<th>Delayed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch–English experiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>20</td>
<td>2.7 1.5</td>
<td>1.3 1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading + fill-in</td>
<td>33</td>
<td>2.9 1.8</td>
<td>1.6 1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>34</td>
<td>4.9 2.3</td>
<td>2.6 1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hebrew–English experiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>31</td>
<td>2.0 2.1</td>
<td>0.4 0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading + fill-in</td>
<td>27</td>
<td>4.0 1.9</td>
<td>1.7 1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>41</td>
<td>6.9 2.5</td>
<td>3.7 2.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
revealed that the mean score of the writing group was significantly different from the mean scores of the reading and reading plus fill-in groups, but that the means of the latter two groups did not significantly differ from each other. Subsequently, the scores of the reading and the reading plus fill-in groups were combined into a single reading (plus fill-in) group. The means of the writing group (N = 34) and the combined reading (plus fill-in) group (N = 53) in both the immediate and delayed post-test were then compared. In both analyses, there was a significant group difference, $F(1, 85) = 25.13; p < .001; \eta^2 = .23$ in the immediate post-test, and $F(1, 85) = 9.89; p < .01; \eta^2 = .10$ in the delayed post-test.

The same analyses were conducted on the data of the Hebrew–English experiment, yielding a similar pattern of results. In the ANOVA, a significant task effect [$F(2, 96) = 63.12; p < .001; \eta^2 = .57$] was found and a significant time effect [$F(1, 96) = 80.85; p < .001; \eta^2 = .46$], as well as a significant task × time interaction [$F(2, 96) = 3.36; p < .05; \eta^2 = .065$]. The significant interaction of task and time may have resulted from a floor effect in the delayed test for the reading group, but not in the reading plus fill-in and writing groups. A post hoc Student-Newman-Keuls test revealed that the mean scores of all three groups were significantly different from each other. Subsequently, the means of groups 1 and 2 were compared with each other in both the immediate and the delayed post-test. In both analyses there was a significant effect of condition [$F(1, 56) = 13.01; p < .01; \eta^2 = .20$ in the immediate post-test, and $F(1, 56) = 19.35; p < .000; \eta^2 = .26$ in the delayed post-test]. Next, the means of groups 2 and 3 were compared with each other in both the immediate and delayed post-test. In both analyses there was a significant effect of condition [$F(1, 66) = 27.42; p < .000; \eta^2 = .29$ in the immediate post-test, and $F(1, 66) = 14.55; p < .000; \eta^2 = .18$ in the delayed post-test].

The results in both experiments provide support for the first part of our hypothesis, which claimed that the composition task would produce the best results. The results in the Israeli, but not the Dutch, setting provide support for the second part of our
hypothesis, which stated that Task 2 (reading + fill-in) will be more effective than Task 1 (reading).

Discussion

The results of the Hebrew–English experiment fully support the hypothesis that words that are processed with higher involvement load will be retained better than words that are processed with lower involvement load. The results of the Dutch–English experiment partially support the involvement hypothesis, in that Task 3 yielded higher retention than Tasks 1 and 2 but that Task 2 did not produce retention significantly higher than Task 1.

The superiority of the composition task may seem to support Swain’s Output Hypothesis (Swain 1985, 1995), given that the composition task required the learners to stretch their linguistic resources. However, the Involvement Load Hypothesis does not predict that any output task will lead to better results than any input task. It predicts that higher involvement in a word induced by the task will result in better retention, regardless of whether it is an input or an output task. This prediction should be empirically tested.

Laufer and Hulstijn (2001) analyzed several studies that compared task effect on learning; it was shown that the more effective tasks had a higher involvement load than the less effective tasks. “Involving tasks” have been practiced with learners of different language proficiency levels. While learners in our study were advanced enough to write a composition, less proficient learners can perform an involving task, which requires search and evaluation in contexts provided by the teacher. Further research should be conducted to corroborate these findings. Because the construct of involvement can be operationalized and investigated empirically, researchers could devise tasks with different involvement loads and compare them with regard to their effect on incidental vocabulary learning.

If further research were to corroborate the findings of this study, the implications for L2 pedagogy would be obvious. Teachers
could design tasks varying in involvement load for different words depending on the type of reinforcement they want to provide. Tasks inducing high involvement load (and, admittedly, requiring more time) would be suitable for words that are particularly important for learners, such as academic words for university students, or words that create special learning difficulties, such as false cognates, synforms, idioms, and other problematic words (for a survey of criteria of lexical difficulties, see Laufer, 1990, 1997). Tasks with a lower involvement load, on the other hand, may be sufficient for easy words, which do not require as much learning effort. Similarly, autonomous learners could be informed about effectiveness of different tasks so that they could make strategic decisions concerning a potential trade-off of the tasks they select.

Teachers often argue that only a limited number of words can undergo “rich” instruction. To our knowledge, there is no evidence regarding the proportion of vocabulary that can be practiced in involving tasks. Further research could determine whether a regular “dose of involvement” promotes vocabulary learning.

An additional pedagogical question is whether task type is just as important, or more so or less so, than the number of times the learner is exposed to the new word. It is almost indisputable that multiple exposures to new words are desirable. Hulstijn (forthcoming-a) argues that the low retention rates in the empirical literature on incidental vocabulary learning indicate that only seldom will a single exposure lead to a sufficiently deep imprint in memory such that, without further reinforcement, the word remains available for retrieval in the long term. According to Nation and Wang (1999), at least ten exposures are necessary for a word to be a good candidate for acquisition, and even then acquisition cannot be guaranteed. Exposure alone may not be enough as many words may go unnoticed. The question then is whether several tasks with low involvement that take less class time are as effective as one task with high involvement. Ellis (1995) argues that a simpler task may lead to more words acquired per minute of input. Further research should reveal whether a few involving tasks can compensate for the relatively limited amount
of exposure that is characteristic of L2 learning in a foreign language environment and whether there is an optimal degree of involvement in terms of teaching time and acquisition results. Studies could compare vocabulary retention along two dimensions: varying task involvement loads and the number of exposures to the investigated words.

Revised version accepted 26 March 2001

Notes

1 Conscious elaboration does not affect implicit memory. The notion of implicit memory refers to the effects of past experiences with single events or objects, such as the priming effects of words encountered previously on reaction times in a lexical decision task (Jacoby, 1983; Jacoby & Dallas, 1981; Roediger, 1990; Stadler & Frensch, 1998; Underwood, 1996). Implicit memory need not manifest itself in performance on tasks of explicit memory, such as (conscious) recognition or recall of previously shown information. As this article is concerned with availability of lexical information in tasks requiring the recognition or recall of explicit lexical information (in reading, listening, writing, and speaking), we will not pursue matters of implicit memory.

2 An incidental learning situation, in contrast to an intentional learning situation, is a situation in which individuals process new information without the intention to commit this information to memory. Incidental vocabulary learning, therefore, refers to the learning of vocabulary as a by-product of any activity not explicitly geared to lexical learning. Because the learners are unaware of the forthcoming vocabulary test, any learning that might occur during the task is considered to be incidental. It is important to note that incidental learning does not mean unattended learning. Intentional vocabulary learning, on the other hand, refers to an activity aimed at committing lexical information to memory. (See Hulstijn, forthcoming-b, for an extensive treatment of this distinction.)

3 A task is defined here as “an activity or action which is carried out as the result of processing or understanding language (i.e., as a response)” (Richards, Platt, & Weber, 1985).

4 The Involvement Load Hypothesis assumes that word complexity factors are held constant.

5 It should be borne in mind that the main purpose of the experiment was to compare three tasks to one another in a homogeneous population. Hence, even if the Dutch and the Israeli learners are not fully identical in their knowledge of English, this does not invalidate the comparison. One may even argue that if similar results are obtained with slightly different populations, this only reinforces the claim that more involving tasks are more effective.
As the target words are of low frequency and English is a foreign language in our countries, it is highly unlikely that learners were exposed to the words between the immediate and the delayed tests. The different time gaps had to do with administrative matters in each university and were irrelevant to our research objective, which was to compare word retention in three tasks. In each experiment, the same time delay was introduced for the three tasks.

SPSS10 (computer software): GLM-procedure with the inclusion of $\eta^2$ and Student-Newman-Keuls posthoc test.

SPSS10 (computer software): Compare Means, F-test plus $\eta^2$.

References


