

 Open access • Journal Article • DOI:10.1111/J.1467-9922.2009.00502.X

Learning L2 German vocabulary through reading: the effect of three enhancement techniques compared — [Source link](#)

Elke Peters, Jan H. Hulstijn, Lies Sercu, Madeline Lutjeharms

Institutions: Catholic University of Leuven, University of Amsterdam, Free University of Brussels

Published on: 01 Mar 2009 - Language Learning (Wiley-Blackwell)

Topics: Vocabulary, Word recognition, Reading comprehension, Vocabulary development and Reading (process)

Related papers:

- [Some empirical evidence for the involvement load hypothesis in vocabulary acquisition](#)
- [Incidental vocabulary acquisition in a second language: the construct of task-induced involvement](#)
- [Learning Vocabulary in Another Language](#)
- [Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary Use, and Reoccurrence of Unknown Words](#)
- [Review article: instructed second language vocabulary learning](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/learning-l2-german-vocabulary-through-reading-the-effect-of-2ikzk7vsev>

Learning L2 German Vocabulary Through Reading: The Effect of Three Enhancement Techniques Compared

Elke Peters

Lessius University College, Antwerp and Catholic University of Leuven

Jan H. Hulstijn

University of Amsterdam

Lies Sercu

Catholic University of Leuven

Madeline Lutjeharms

Free University of Brussels

This study investigated three techniques designed to increase the chances that second language (L2) readers look up and learn unfamiliar words during and after reading an L2 text. Participants in the study, 137 college students in Belgium (L1 = Dutch, L2 = German), were randomly assigned to one of four conditions, forming combinations of two between-subject factors: \pm prereading test announcement and \pm postreading vocabulary task. Comprehension questions were used to direct participants' attention to half of the 16 target words in this study, creating the within-subject factor (word relevance). Participants accomplished the experimental tasks at computers. They could look up the meaning of unfamiliar words in an online dictionary. The dependent variables are the following: use of online dictionary during reading, performance on a word-form recognition test, and performance on two word-meaning recall tests (immediate and

We would like to thank Benjamin Biebuyck, Walter De Cubber, and Luc De Grauwe (University of Ghent), Henri Bloemen and Nathalie Moss (Lessius University College), Ingrid Bollaerts and Joël Rooms (VLEKHO), Luk Draye (Catholic University of Leuven), and Jan Van Nieuwenbergh (Erasmus University College) for their help in the data collection. We are also very grateful to the anonymous reviewers for their valuable comments and suggestions.

Correspondence concerning this article should be addressed to Elke Peters, Lessius University College, Jozef De Bromstraat 11, B-2018 Antwerp, Belgium. Internet: elke.peters@lessius.eu; elke.peters@arts.kuleuven.be

delayed). Test announcement and word relevance substantially prompted participants to use the online dictionary more. Only test announcement and vocabulary task (not word relevance) affected performance in the word recognition test positively. Both word relevance and postreading vocabulary task substantially affected word retention in the recall posttests. These findings, together with those of the studies reviewed, provide robust evidence that the low incidence of vocabulary acquisition through reading (“input only”) can be substantially boosted by techniques that make students *look up the meaning of unknown words*, *process* their form-meaning relationship *elaborately*, and process them *again* after reading (“input plus”).

Keywords vocabulary acquisition, reading, enhancement technique, (repetition) task, German, dictionary use, word learning, incidental/intentional learning

One of the most discussed themes in second language¹ (L2) vocabulary acquisition is how written input affects word learning. A central question has been whether reading is as beneficial for L2 vocabulary acquisition as it is for L1 vocabulary acquisition. Estimates of L1 speakers’ vocabulary size cannot be accounted for by formal instruction only. Consequently, researchers argued that most words are acquired incidentally through reading (Krashen, 1985; Nagy, 1997; Nagy, Herman, & Anderson, 1985; Saragi, Nation, & Meister, 1978). A meta-analysis of 20 experiments examining incidental L1 word learning during normal reading, conducted by Swanborn and De Glosper (1999), showed that students learn around 15% of the unknown words they encounter.

For L2 learning, researchers have found evidence that incidental acquisition-through-reading is a slow and error-prone process with small vocabulary gains (Horst, Cobb, & Meara, 1998; Hulstijn, Hollander, & Greidanus, 1996; Laufer, 2005; Nation, 2001; Read, 2004). Laufer (2003, 2005) explained these small learning gains in terms of the “fault” in the “Default Hypothesis.” L2 learners do not always notice unfamiliar words when reading a text. If they do, guessing the meaning is not always possible. Moreover, L2 learners do not always guess the meaning successfully. Finally, L2 vocabulary acquisition is an incremental process. Yet, in formal language learning contexts, L2 learners do not receive sufficient input and do not have sufficient encounters with new words to achieve this. That is why one of the recurrent themes in vocabulary research has been how L2 vocabulary acquisition through reading can be enhanced.

The literature on learning and memory and the literature on L2 vocabulary learning, in particular, suggest that successful L2 vocabulary acquisition through reading is contingent on three factors. First, L2 learners should discover the meaning of unfamiliar words. Second, they should process the lexical

information elaborately. Third, the form-meaning connections of these words should be reinforced by means of repetition. When L2 learners want to learn new words through reading, they first need to *discover their meaning*. They can try to infer the meaning on the basis of contextual clues, if possible, or they can consult its meaning in a dictionary. Although discovering a word's meaning is a necessary condition, it is not a sufficient one to learn words. L2 learners need to *process* (the lexical information of) new words thoroughly before acquisition can take place. The "depth-of-processing" hypothesis (Craik, 2002; Craik & Lockhart, 1972; Craik & Tulving, 1975) states that retention of information is directly and strongly determined by the way the information is processed. Basing himself on the work by Craik and his associates, Hulstijn (2001) argued that processing new lexical information elaborately, by paying attention to various features (e.g., the word's meaning[s], semantic relations to other words, the word's grammatical category, pronunciation, and orthography) will enhance word retention. The Involvement Load Hypothesis (Laufer & Hulstijn, 2001) provides a framework for research on L2 word learning. Task-induced involvement consists of one motivational (Need) and two cognitive components (Search and Evaluation). *Need* refers to the need to accomplish a task. *Search* is the attempt to find the meaning of a word or to find the word form for a specific concept. *Evaluation* entails that a given word is compared with other words or that a specific meaning of a word is compared with its other meanings (Laufer & Hulstijn, p.14). Evaluation is part of what Hulstijn (2001) called *elaborate processing*. Finally, findings from memory research (Baddeley, 1997) have demonstrated that learners should be reexposed to new words as soon as possible after the first encounter in order to *reinforce* the form-meaning connections of these words. Immediate repetition of a word after its initial encounter is especially beneficial for word learning (Baddeley; Hulstijn; Nation, 2001).

This study explored the use of three potential enhancement techniques: telling students that the reading task will be followed by a vocabulary test (vocabulary test announcement), forcing students to pay attention to unfamiliar words in the reading text via comprehension questions (task-induced word relevance), and requiring students to pay attention to the unfamiliar words again, after completion of the reading task (vocabulary task). Our specific focus was on how the variables vocabulary test announcement, task-induced word relevance, and vocabulary task, and the interaction among them, would affect students' look-up behavior (i.e., online dictionary use) and students' word retention.

Previous Empirical Research

This section surveys previous studies that were concerned with enhancement techniques that prompted L2 learners to discover and process the meaning of new words.

Test Announcement

Some researchers have focused on how various prelearning instructions will affect what L2 learners do when they come across unfamiliar words when they read a text. Peters (2007a, 2007b) found that when the prereading instructions contained the announcement of a postreading vocabulary test, students were more intensively engaged in looking up words while reading the text. However, the announcement of a postreading vocabulary test did not affect word retention. This finding was corroborated by Sercu and Dewachter (Sercu, Dewachter, Peters, Kuiken, & Vedder, 2006). Thus, in the aforementioned studies, the presence or absence of a postreading vocabulary test announcement created situations for, respectively, intentional and incidental learning. This means that the labels *incidental* and *intentional* were used strictly in their methodological sense, as defined by Hulstijn (2001, 2003). *Incidental* entails that participants are not forewarned of an upcoming test, whereas the label *intentional* refers to the fact that participants are explicitly told that an upcoming test will follow. In this study, we will adhere to this methodological interpretation when using the labels *incidental* and *intentional*.

Word Relevance

Hulstijn (1993) investigated how a word's relevance affected students' look-up behavior when they read an L2 (= English) text. A relevant word was defined as a word that needed to be used receptively to answer a text comprehension question, whereas an irrelevant word was not linked to a comprehension question. Participants answered these questions with the text and an online dictionary available. Thus, the relevance factor was operationalized by the comprehension questions directing students' attention to the relevant words. The findings showed that L2 learners looked up more words when these words were deemed relevant for task completion than when these words were not considered relevant. These findings were corroborated by Laufer and Levitzky-Aviad (2003) and Peters (2007b). Peters also explored whether word relevance had an effect on immediate and delayed recall of word meanings (L2 = German). Her findings demonstrated that word relevance promoted word learning. Learners were more actively engaged in processing the relevant words compared to the

irrelevant ones. Moreover, the effect was durable, as measured by posttests 1 week and 2 weeks later.

Thus, whereas test announcement makes L2 learners more sensitive to looking up words, answering comprehension questions will prompt L2 learners not only to look up (relevant) words (the dimension of *Need* of the Involvement-Load Hypothesis of Laufer and Hulstijn, 2001) but also to process the lexical information more elaborately in order to provide a correct answer to the comprehension questions.

Word-Focused Activities

Research has also investigated whether word-focused activities, productive ones in particular, promote vocabulary learning. Examples of word-focused activities are a composition-writing task, a fill-in-the-blanks task, or a sentence-writing task. In several studies, Laufer (2001, 2003, 2005) compared reading with word-focused activities and found that the word-focused activities always yielded better vocabulary learning gains than Reading Only.

In a similar line of research, Paribakht and Wesche (1997, 1998, 1999) investigated how students in a Reading Only condition differed from students in a Reading Plus condition. Both groups were exposed to reading materials. The Reading Only group read more texts about the same topic, being exposed to the target words several times (different texts, reading comprehension questions). The Reading Plus group had to complete receptive as well as productive vocabulary activities instead of reading more texts. Both treatments yielded vocabulary gains, but students in the Reading Plus condition learned more words than students in the Reading Only condition. The differences were both qualitative and quantitative. In the Reading Only treatment, knowledge remained at the recognition level, whereas students in the Reading Plus group were also able to use words productively (writing a sentence). In a study conducted by Laufer (2001), students were assigned to one of three conditions: (a) reading with provision of a dictionary, (b) reading with provision of a dictionary, and a sentence-writing task, and (c) a sentence-writing task (with provision of definitions of target words). On an immediate vocabulary test, participants in the second condition performed significantly better than participants in the two other groups. However, on a delayed vocabulary test, the reading and sentence-writing group no longer differed significantly from the sentence-writing group, but both groups still performed better than the Reading Only group.

In short, the studies surveyed in this subsection suggest that word-focused activities tend to promote vocabulary learning. Unlike most studies, the present study did not focus on a productive word-focused activity but on a *receptive*

vocabulary *consolidation* task because too little research on consolidation tasks has been carried out so far. As Bogaards and Laufer (2004, p. XII) pointed out, “one of the most important phases in vocabulary learning which has not been researched sufficiently is consolidation of knowledge after initial presentation.” This study is an attempt to accommodate this gap in vocabulary research.

Aim and Research Questions

Taking the previous studies into account, we wanted to explore the effect of three enhancement techniques, telling students that the reading task will be followed by a vocabulary test (vocabulary test announcement), forcing students to pay attention to unfamiliar words in the reading text via comprehension questions (task-induced word relevance), and requiring students to pay attention to the unfamiliar words again, after completion of the reading task (vocabulary task). (Note that the labels *vocabulary test* and *vocabulary task* do not refer to the same thing. This will become clear in the Method section.) We investigated whether test announcement, word relevance, and their interaction would have an effect on L2 learners’ look-up behavior. In addition, we examined whether test announcement, word relevance, vocabulary task, and the interaction among them would promote word retention. Although previous studies looked at these factors either in isolation (Hulstijn, 1993; Laufer & Levitzky-Aviad, 2003; Peters, 2007a) or as a combination of two factors (Peters, 2007b), this study is, to our knowledge, the first one to investigate the effect of all three techniques, their interactions, and any cumulative effects in a single experiment.²

With respect to learners’ look-up behavior, this study addressed the following four research questions:

1. What is the effect of vocabulary test announcement on students’ look-up behavior?
2. What is the effect of word relevance on students’ look-up behavior?
3. Is there a significant interaction effect between vocabulary test announcement and word relevance on students’ look-up behavior while they read a text?
4. Is there a significant interaction effect between vocabulary test announcement and word relevance on students’ look-up behavior while they carry out a vocabulary task?

With regard to word retention, we wanted to find an answer to the following research questions:

5. What is the effect of vocabulary test announcement on word retention?
6. What is the effect of vocabulary task on word retention?
7. What is the effect of word relevance on word retention?
8. Is there a significant interaction effect among vocabulary test announcement, word relevance, and vocabulary task on word retention?

Method

Design

The study adopted a factorial design with two between-subject factors and one within-subject factor. Vocabulary test announcement (presence or absence) and vocabulary task (presence or absence) constituted the between-subject variables; word relevance (plus-relevant or minus-relevant target words) was the within-subject variable. The design is presented schematically in Table 1.

Participants in all four groups first read a text in German (reading *task*), participants in groups 2 and 4 additionally performed a postreading vocabulary *task*, whereas participants in all groups finally performed several vocabulary *tests* (note the difference between *task* and *test*). The German text, which participants had to read, contained 16 pseudowords, constituting the target words of this study. The target words appeared only once in the text. Students could look up the meaning of the 16 target words and of all low-frequency words (those not belonging to the 2,000 most frequent words in German). In the Results section, we refer to these words as “annotated words.” By clicking on them, a window popped up rendering the word’s translation in L1 as well as a meaning paraphrase in L2. Participants in the first and second groups were not forewarned that after having answered comprehension questions, a vocabulary test would be administered (no prereading test announcement; see Table 1). Students in the third and fourth groups, on the other hand, were explicitly forewarned that after having answered the comprehension questions, they would be tested on the vocabulary used in the text (prereading test announcement). The former

Table 1 Design: Test announcement, word relevance, and vocabulary task by group

	Vocabulary test announcement	Word relevance (= comprehension questions)	Vocabulary task
Group 1 (INCID ONLY)	No	Yes/No	No
Group 2 (INCID PLUS)	No	Yes/No	Yes
Group 3 (INTENT ONLY)	Yes	Yes/No	No
Group 4 (INTENT PLUS)	Yes	Yes/No	Yes

groups constituted the incidental groups, whereas the latter groups formed the intentional groups (see the Empirical Research section for a definition of *incidental* and *intentional*).

Participants who were assigned to the second and fourth groups had to carry out a vocabulary task, in which the 16 target words were presented (see Table 1). The vocabulary task was only announced after students had finished the comprehension questions in order not to compromise the incidental learning character of the second group (= INCID PLUS). Students in the fourth group (= INTENT PLUS), on the other hand, were reminded of the upcoming vocabulary test prior to the vocabulary task. In other words, they accomplished the vocabulary task under an intentional vocabulary learning condition.

Task-induced word relevance constituted the within-subject factor with two levels: plus-relevant and minus-relevant target words. Task-induced word relevance was operationalized by asking text comprehension questions, which formed part of the reading task. Half of the target words appeared in text passages whose content was focused on by the comprehension questions; the other target words appeared in passages whose content was not targeted by the comprehension questions. Thus, in order to answer the comprehension questions, students needed to process half of the target words (*viz.* the plus-relevant target words) receptively. In this way, the comprehension questions implicitly led students to look up the meaning of eight plus-relevant words. Next, they had to incorporate these words in their answer to the text comprehension questions. The minus-relevant target words were not related to the comprehension questions. Thus, although the comprehension questions led students to elaborately process the plus-relevant target words, an elaborate processing of the minus-relevant words was not enforced by the task.

The dependent variables were students' look-up behavior and word retention. Students' look-up behavior refers to whether a target word was looked up (first coding), frequency of clicks on target words (second coding), and clicks on target words during the vocabulary task. Word retention was measured in three immediate and two delayed receptive vocabulary tests (see also Data Collection Instruments subsection).

Participants

Two Belgian universities and three Belgian university colleges provided the participants for this study, in which 137 students (L1 = Dutch, L2 = German) took part. Students studied either German Linguistics and Literature at a university or German Translator and Interpreter Studies at a university college. They were of an upper intermediate to advanced proficiency level.

Materials

Computer Program

The data were gathered via computers. To keep track of participants' look-up behavior, a computer program was developed using the software QuestionMark Perception.³ Each click on a word was registered in log files in a database. These files enabled us to collect data on students' look-up behavior in an unobtrusive way. The online look-up facility itself contained the L1 translation as well as the L2 synonym or definition of a word, as can be seen in Figure 1. In the remainder of this text, we sometimes refer to this facility as an online dictionary. It should be borne in mind, however, that, as Figure 1 illustrates, the look-up device did not constitute a full-fledged dictionary. A mouse click only produced an L2 paraphrase and an L1 translation of the word in context.

Text

We selected an authentic text from the German newspaper *Die Zeit*, entitled *Ossis sind Türken* (East Germans are Turks) (Staud, 2003), which focuses on the situation of the West and East Germans in the reunified Germany. The

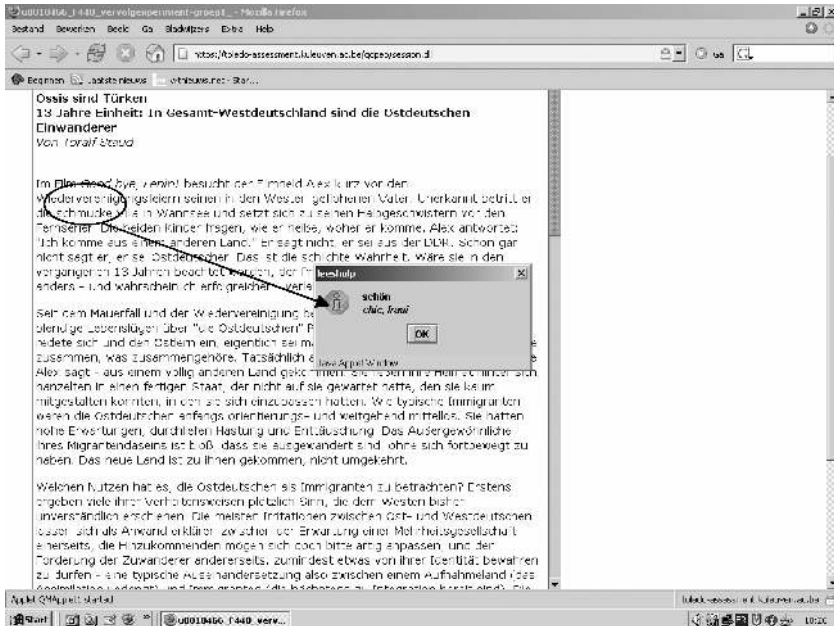


Figure 1 Screenshot of text with pop-up window of dictionary information of the word *schmucke*.

author elaborates on why the reunification has not been very successful by comparing the situation of East Germans with the situation of immigrants. The text,⁴ shortened to 1,096 words, was only available on the computer screen (see Figure 1). Both the text and the computer program were identical to the ones used in the study by Peters (2007b). The text is the type of text these high-intermediate to advanced students would be expected to (be able to) read.⁵

Text Comprehension

Text comprehension questions were developed to operationalize the relevance factor. Participants had to answer nine questions with the text and the look-up facility at their disposal. The questions were of the so-called “textually explicit” type, meaning that the question-answering process “involves search-and-match strategies rather than actual comprehension” (Alderson, 2000, p.107). The questions (in Dutch) had to be answered in students’ L1, Dutch, to ensure that students would consult the online dictionary and would use the plus-relevant target words receptively in their answer. Otherwise, they could copy the sentences without knowing the meaning of the plus-relevant target words. For one question, two plus-relevant target words needed to be used. The relevance factor is illustrated in Figure 2. Note that a correct answer to the comprehension question requires that the students do not only look up the meaning of the plus-relevant target word but that they integrate the word’s meaning in a proposition meaningfully related to the context and the question (elaborate processing).

In order not to draw too much attention to unknown words, two of the nine questions did not require a plus-relevant target word. These two questions were, as were the other seven, textually explicit questions, the answering of which required participants to have grasped the meaning of two nontarget words.

Text excerpt with plus-relevant target word in bold⁶:

Fahnung für die zurückgelassene Kultur ist eine Voraussetzung für eine angemessene Verarbeitung des Kulturschocks.

Comprehension question related to this excerpt:

Wat is een belangrijke voorwaarde voor een geschikte verwerking van de “cultuurschok”? (*What is an important condition for coping adequately with culture shock?*)

Answer:

Verdriet om de achtergelaten cultuur (*sadness for the culture that is left behind*)

"Sadness" (*Fahnung*) is one of the eight plus-relevant target words, which is used to answer the question.

Figure 2 One of the text comprehension questions.

We assumed that the participants would be familiar with these two nontarget words.

Target Words

Sixteen target words were selected. They were all replaced by pseudowords in order to ensure that no student knew the target words. German morphology was respected when creating the pseudowords so that they looked like normal, existing German words to the participants. The target words appeared only once in the text. It was not possible to infer the precise meaning of the target words on the basis of contextual clues. Consequently, students simply had to consult the online dictionary to know a target word's meaning and, in the case of plus-relevant target words, be able to answer the comprehension questions. This allowed us to gain reliable data on when and how often participants tried to discover the meaning of the target words.

Vocabulary Task

The postreading vocabulary task was available on a computer screen, which was divided into two frames (see Figure 3). The vocabulary task appeared in the

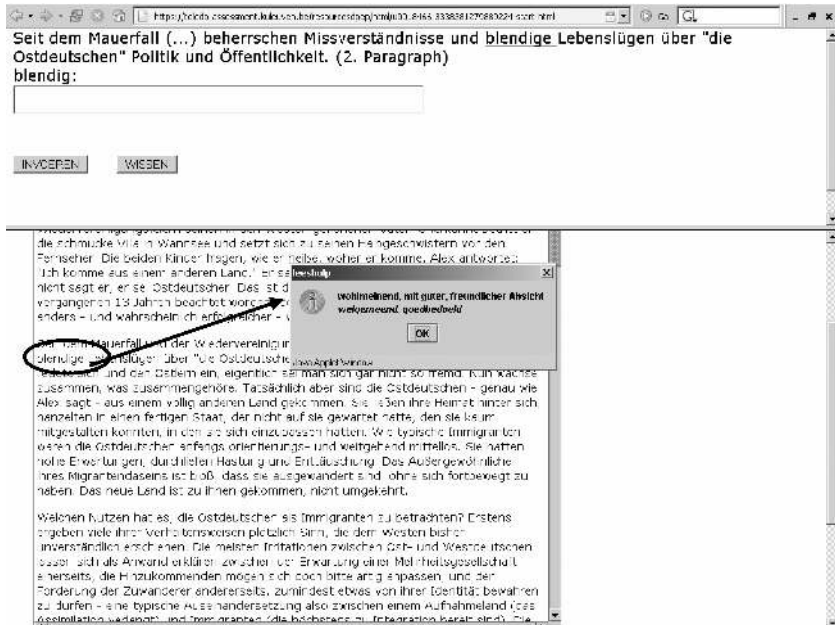


Figure 3 Screenshot of post-reading vocabulary task (upper frame) and text with dictionary (bottom frame).

upper frame, and the text with the look-up facility, still available, appeared in the bottom frame of the screen. The target words, for which participants needed to provide an L2 synonym/definition or an L1 translation, were presented in the sentences as they had occurred in the text. After each sentence, the paragraph, in which the sentence occurred, was mentioned to help participants locate the word in the text. They could supply their answer immediately, but they could also scroll back and forth through the text and reread the sentence and/or paragraph before they submitted an answer. They were free to look up the word's meaning. Irrespective of the answer students submitted, the correct translation and synonym were provided, but without a score so as to focus the learners' attention again on the target words. The target words appeared one by one.

Data Collection Instruments

The data collection instruments, apart from the software registering participants' online word look-ups, consisted of a vocabulary size test, three vocabulary posttests, think-aloud protocols, and retrospective questions.

We used a self-designed vocabulary size test based on the German frequency list of the Institut für deutsche Sprache (Mannheim). It consisted of 50 multiple-choice items covering the frequency bands from 4,000 to 10,000 words.⁷ This test was administered because research has shown that word retention correlates positively with students' vocabulary size (Horst et al., 1998). Moreover, students' vocabulary size would function as a covariate in the statistical analyses.

In order to test word retention and to give learners credit for partial vocabulary knowledge, three receptive vocabulary posttests were administered (see Figure 4): one recognition test and two recall tests. They appeared online, one by one and in the following sequence: (a) Word Form Recognition Test, (b) Isolation Test, and (c) Context Test. Participants could only see one test at a time, and they could not go back to the previous tests.

In each test, items were presented in a randomized way to control for order effects. The first test, the Word Form Recognition Test, contained 40 stimulus words, of which 16 were the target words, and 24 words did not appear in the text (distracter words). Participants were required to tick off whether the word had appeared in the text. With this test we wanted to find out whether students would recognize the form of the target words at the most superficial level. This test was labeled "Word Form Recognition Test," hereafter referred to as recognition test.

In the first recall test, students had to provide a German definition or Dutch translation for the 16 target words, which were presented in isolation

-
- Word Form Recognition Test: Did the word appear in the text?
 verzetteln yes no
 - Isolation Test (target words in isolation):
 verzetteln:
 - Context Test (target words in context):
 Immigranten neigen dazu, sich **verzettelt**, ausgebeutet und [...] zu fühlen.
 verzetteln: _____

Figure 4 Target word “verzetteln” tested in three vocabulary posttests.

(= isolation test). In the second recall test, the target words were offered in the sentences as they had occurred in the text (= context test). The context test was identical in format to the vocabulary task except that the text, the online dictionary, and the reference to the paragraph, in which the target word had occurred, were not available. We assumed that the isolation test would be the most demanding test. The isolation test and context test were also used to test word retention 2 weeks after the first session (see also the *Procedure* section).

In addition to the collection of quantitative data, qualitative data were collected to refine our understanding of the processes underlying the learning activity. Students filled in a questionnaire with retrospective questions on their strategy use and on how they had processed the dictionary information. Furthermore, 12 students provided us with think-aloud protocols.

Procedure

Data were collected at two universities and three university colleges in Belgium. The experiment took place in a computer lab. Students in each class were randomly assigned to the four experimental groups. They worked simultaneously but individually at their own work station. Before starting with the experiment, participants were familiarized with the online look-up facility in a 5-min introduction. However, they were not informed that their look-up behavior would be registered. All participants started by reading the instructions on the computer screen. Participants in the incidental groups (INCID ONLY and INCID PLUS) received the instruction that they had to read a text on a computer screen and that they had to answer nine comprehension questions while reading the text. They were also told that, having read the text, they would have

to answer some questions about how they had experienced the experiment. The intentional groups (INTENT ONLY and INTENT PLUS) received the same instruction, but, in contrast to the incidental groups, they were forewarned of an upcoming receptive vocabulary test. At this moment, participants in the PLUS groups were not yet informed that they would have to accomplish a postreading vocabulary task.

Next, participants started reading the text with the comprehension questions at their disposal. They provided the answers to the nine questions in their L1 (Dutch) on a separate sheet of paper. Although no time limit was set for answering the comprehension questions, participants were asked to write down when they had finished their task.

Having completed the comprehension task, students who were not set a vocabulary task started taking the vocabulary posttests (a surprise for the INCID ONLY group), whereas students in the INCID PLUS and INTENT PLUS groups were informed that they were to carry out a vocabulary task and that the text with the online dictionary would remain available in the bottom frame of the screen. In addition, they were told that if they wanted to, they could look up a word's meaning in the online dictionary. Unlike students in the INCID PLUS group (– test announcement), students in the INTENT PLUS group (+ Test announcement) were reminded of the upcoming vocabulary test. In other words, the vocabulary task was presented as a preparation for the vocabulary test. No time limit was set for the vocabulary task, but the time-on-task was registered by the computer program. Having finished the vocabulary task, participants in these two groups moved on to the vocabulary posttests (a surprise for the INCID PLUS group).

All four groups took all vocabulary posttests. They started with the recognition test before continuing first with the isolation test and then the context test. As mentioned already, the next test only appeared after the previous one had been completed. Next, students answered retrospective questions about how they had experienced the experiment. Finally, all students took the vocabulary size test. Except for the comprehension questions task, all other tasks and tests were administered on the computer. We also administered the isolation test and context test 2 weeks after the first session took place (= delayed vocabulary posttests).

Except for six students who participated on a voluntary basis, the experiment was organized during students' normal class hours. After the delayed posttests, participants were debriefed about the aim and procedure of the experiment.

Scoring and Data Analysis

Information on participants' look-up behavior came from the computer log files. It pertains to the number of target words looked up (first coding) and to the frequency of clicks of target words; that is, subsequent clicks were also counted (second coding). We also analyzed the look-up behavior of the two PLUS groups during the vocabulary task.

The vocabulary posttests were scored dichotomously, with 1 assigned to a correct answer and 0 to an incorrect one. In all statistical analyses, an alpha level of .05 was taken as the level of statistical significance.⁸ As a measure of effect size, we used η^2 , which is appropriate when more than two groups are involved (e.g., ANOVA). It refers to the proportion of the total variance that is attributable to an effect ($\eta^2 = SS_{\text{effect}}/SS_{(\text{corrected})\text{total}}$). Its values were interpreted as follows: $\eta^2 > .0099 =$ small, $\eta^2 > .0588 =$ moderate, and $\eta^2 > .1379 =$ large (Cohen, 1992). For instance, η^2 with a value of .40 would indicate that the effect size is large. The advantage of reporting effect sizes is that the values are independent of the sample size, in contrast to the p -values.

Results

We first present students' scores on the comprehension questions, vocabulary task, and the vocabulary size test. Next, we focus on the results with respect to students' look-up behavior and word retention.

Text Comprehension

Participants completed the comprehension task in an average time of 37 min. In addition, they scored generally well on the comprehension task, which we took as an indication that they took the experiment seriously ($M = 7.30$, Max. = 9, with a range from 3 to 9; $SD = 1.35$). A two-sample t -test indicated that the incidental groups (– test announcement) did not differ significantly from the intentional groups (+ test announcement)⁹ ($t = -0.11$, $DF = 135$, $p = .91$) in their comprehension of the text nor in the time they devoted to the comprehension task ($t = 0.23$, $DF = 130$,¹⁰ $p = .81$).

Vocabulary Task

Participants in the PLUS (+ vocabulary task) groups performed very well on the vocabulary task, which took 10 min on average. Because of the small difference between the INCID PLUS and the INTENT PLUS group with mean scores of 14.4 and 14.7, respectively (Max. = 16), it is no surprise that the ANOVA did not reveal a statistically significant difference, $F(1, 69) = 0.18$,

$p = .67$, $\eta^2 = .00$. Although students could look up all target words, they obtained higher scores on the plus-relevant than on the minus-relevant target words (7.5 vs. 6.9; Max. = 8). The difference was statistically significant, $F(1, 69) = 12.02$, $p = .001$, $\eta^2 = .15$. The ANOVA did not reveal an interaction effect between word relevance and test announcement ($p > .05$).

Vocabulary Size Test

The mean vocabulary size score was 24.5 ($SD = 7.6$; Max. = 50; $N = 136$). Although there were differences between the four experimental groups, an ANOVA showed that they did not differ significantly from each other in their vocabulary size, $F(3, 135) = 2.25$, $p = .09$. We found a good reliability index ($N = 136$, Cronbach's alpha = .80). Students' vocabulary size was later used as covariate in the inferential statistics on look-up behavior and word retention, reported below.

Look-up Behavior

First, we present the results of participants' look-up behavior of the target words according to the first coding (looked up or not). After a description of students' look-up behavior of the target words according to the second coding (frequency of clicks), we report on the analysis of students' frequency of clicks during the vocabulary task.

Looked Up or Not

As can be seen from Table 2, the groups that had received a prereading test announcement (= intentional groups) looked up the meaning of more target words than the groups that had not (= incidental groups). An ANCOVA (see Table 3) showed a significant difference between the incidental and intentional groups with regard to the number of looked-up target words, $F(1, 135) = 16.77$, $p < .0001$. Its effect size was moderate ($\eta^2 = .11$).

Table 2 Number of look-ups (first coding) by condition and word type

	<i>N</i>	All target words (Max. = 16)		Plus-relevant (Max. = 8)		Minus-relevant (Max. = 8)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Incidental (– test announcement)	69	10.7	2.8	7.5	0.7	3.3	2.3
Intentional (+ test announcement)	68	12.7	2.9	7.4	1.0	5.3	2.3
All students	137	11.7	3.0	7.5	0.8	4.3	2.5

Table 3 One-way ANCOVA of look-up behavior according to the first coding: Two groups

	<i>DF</i>	<i>F</i>	<i>p</i>	η^2
Relevance	1	24.81	<.0001	.13
Relevance \times test announcement	1	35.34	.001	.18
Relevance \times vocabulary size	1	0.15	.67	.00
Error (within-subject)	133			
Test announcement	1	16.77	<.0001	.11
Vocabulary size (covariate)	1	0.301	.58	.00
Error (between-subject)	133			

As mentioned in the Materials subsection, participants could look up more words than only the 16 target words. An analysis of all annotated words revealed that participants in the intentional group clicked considerably more frequently on words than participants in the incidental group (55 vs. 37). This difference was significant, as revealed by an ANCOVA, $F(1, 134) = 12.33$, $p < .0001$, $\eta^2 = .09$.

Students looked up more plus-relevant than minus-relevant words. In addition, students consulted almost all plus-relevant target words, as is apparent from Table 2. The difference was statistically significant (see Table 3; $p < .0001$). In addition, the effect size ($\eta^2 = .13$), which was moderate, was slightly larger than the one of test announcement ($\eta^2 = .11$).

Let us now consider how the relevance factor interacted with vocabulary test announcement. As can be seen from Table 2, students who were forewarned of an upcoming vocabulary test looked up an average of two more minus-relevant words than students who were not forewarned. The ANCOVA (see Table 3) revealed a significant interaction effect ($p = .001$) and the highest effect size ($\eta^2 = .18$); that is, the groups only differed in the number of minus-relevant target words they looked up.

Frequency of Look-ups

The question we want to address now is whether the analyses of students' look-up behavior according to the second coding corroborate the above findings. As can be seen from Table 4, the intentional group produced on average four more clicks than the incidental group. An ANCOVA¹¹ (see Table 5) indicated that the incidental and intentional groups differed significantly from each other in their frequency of clicks on target words ($p < .0001$, $\eta^2 = .11$). As for the first coding, the effect was moderate.

Table 4 Frequency of clicks on target words by condition and word type

Condition	N	All target words		Plus-relevant		Minus-relevant	
		M	SD	M	SD	M	SD
Incidental (– test announcement)	69	16	6	12	4	4	3
Intentional (+ announcement)	67	21	8	14	5	7	4
All students	136	18	7	13	4	5	4

Table 5 ANCOVA of look-up behavior according to the second coding: Two groups

	DF	F	p	η^2
Relevance	1	92.62	<.0001	.40
Relevance \times test announcement	1	3.22	.08	.01
Relevance \times vocabulary size	1	1.62	.21	.01
Error (within-subject)	132			
Test announcement	1	16.19	<.0001	.11
Vocabulary size (covariate)	1	0.59	.45	.00
Error (between-subject)	132			

Participants clicked more often on plus-relevant words than on minus-relevant words. As can be seen from Table 4, students produced, on average, eight more clicks on the plus-relevant than on the minus-relevant target words. The ANCOVA (see Table 5) indicated that this difference was statistically significant ($p < .0001$). Moreover, we found a large η^2 of .40.

Even though students in the intentional condition clicked more often on minus-relevant target words compared to students in the incidental condition, we did not find a significant interaction between vocabulary test announcement and word relevance (see Table 5; $p = .08$).

Look-up Behavior During Vocabulary Task

As can be seen from Table 6, students in the INTENT PLUS group looked up slightly more target words while accomplishing the vocabulary task than students in the INCID PLUS group. Not surprisingly, the ANOVA¹² (see Table 7) did not reveal a significant group effect ($p > .05$).

As is apparent from Table 6, students who carried out the vocabulary task looked up more minus-relevant than plus-relevant target words while doing this task. The average number of look-ups per word category was subjected to an

Table 6 Number of clicks on target words during vocabulary task per condition

Condition	<i>N</i>	All target words		Plus-relevant		Minus-relevant	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
INCID PLUS (– test announcement)	35	9.8	4.3	3.4	2.5	6.5	2.9
INTENT PLUS (+ test announcement)	35	10.6	4.7	3.4	2.7	7.1	2.6
Both groups	70	10.2	4.5	3.4	2.6	6.8	2.7

Table 7 ANOVA of look-up behavior during vocabulary task

	<i>DF</i>	<i>F</i>	<i>p</i>	η^2
Relevance	1	94.56	<.0001	.58
Relevance × test announcement	1	0.81	.37	.00
Test announcement	1	0.48	.49	.01
Error	68			

ANOVA. This analysis showed that the difference was statistically significant (see Table 7); its effect size was also very large. It also indicated that there was no significant interaction effect between test announcement and word relevance.

To summarize, both test announcement and word relevance had a moderate to large effect on learners' look-up behavior. As expected, when forewarned of a vocabulary test, students clicked more frequently on the target words than when not forewarned. In addition, plus-relevant target words were looked up more often than minus-relevant ones. However, exactly the opposite occurred during the vocabulary task. Students clicked more often on the minus-relevant words than on the plus-relevant words. We found only one significant interaction between vocabulary test announcement and word relevance: Students forewarned of the vocabulary test looked up more minus-relevant words than those not forewarned. We will return to these findings in the Discussion section.

Word Retention

In this subsection, we provide the results and statistical analyses of learners' retention of the target words. The results are presented by test type.

Word Form Recognition Test

As can be seen from Table 8, students in the INTENT PLUS group remembered having seen more target words in the text than students in the other three groups.

Table 8 Recognition test: Word retention by condition and word type (Max. = 16)

	<i>N</i>	All target words		Plus-relevant		Minus-relevant	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
INCID ONLY	34	10.41	3.12	6.53	1.60	3.88	2.11
INCID PLUS	35	14.97	1.22	7.43	0.78	7.54	0.74
INTENT ONLY	33	11.97	2.78	6.55	1.28	5.42	1.95
INTENT PLUS	35	15.31	0.93	7.69	0.72	7.63	0.55
All students	137	13.20	3.01	7.06	1.25	6.15	2.16

Note: INCID ONLY (– test announcement, – vocabulary task), INCID PLUS (– test announcement, + vocabulary task), INTENT ONLY (+ test announcement, – vocabulary task), INTENT PLUS (+ test announcement, + vocabulary task).

Table 9 Two-way ANCOVA of recognition test

	<i>DF</i>	<i>F</i>	<i>p</i>	η^2
Within-subject				
Relevance	1	0.66	.42	.00
Relevance \times test announcement	1	8.16	.005	.04
Relevance \times vocabulary task	1	56.59	<.0001	.27
Relevance \times vocabulary size	1	1.86	.18	.01
Relevance \times test announcement \times vocabulary task	1	11.37	.001	.05
Error (within-subject)	131			
Between-subject				
Test announcement	1	5.48	.02	.02
Vocabulary task	1	101.28	<.0001	.42
Test announcement \times vocabulary task	1	1.92	.17	.01
Vocabulary size (covariate)	1	3.84	.052	.02
Error (between-subject)	131			

Students in the INCID ONLY group on the other hand obtained the lowest score on the Recognition Test. A two-way ANCOVA (see Table 9) indicated that vocabulary test announcement had a significant effect on the recognition test ($p < .05$), although the effect size was small ($\eta^2 = .02$).

It is apparent from Table 8 that participants who had done the vocabulary task performed better than participants who had not. The former recognized, on average, four words more than the latter ones. The ANCOVA (see Table 9) indicated that this difference was statistically significant, $F(1, 135) = 101.28$,

$p < .0001$, $\eta^2 = .42$. The effect size of vocabulary task was considerably larger than the one of test announcement.

Students remembered slightly more plus-relevant than minus-relevant words. However, this difference was not statistically significant, as can be seen from Table 9.¹³ Table 9 also provides the interaction effects. The ANCOVA revealed three significant interactions, between test announcement and relevance ($p = .005$), between vocabulary task and relevance ($p < .0001$), and among the three independent variables ($p = .001$). We conducted post hoc Tukey-Kramer analyses to check which groups differed significantly on which type of words (plus- or minus-relevant words). With regard to the plus-relevant target words, there was only a significant difference between the groups that had carried out the vocabulary task and those that had not. The former groups recalled, on average, the meaning of one more plus-relevant target word than the latter ones. With regard to the minus-relevant target words, we found that the PLUS groups (+ vocabulary task) performed better than the ONLY groups (- vocabulary task), but also that the INTENT ONLY group performed significantly better than the INCID ONLY group. However, only the interaction between word relevance and vocabulary task had a large effect size; the other effect sizes were small.

*Word Retention in Isolation*¹⁴

To investigate word retention as measured in the isolation test and the context test, we used repeated-measures ANCOVAs. These analyses allowed us to investigate how word retention had changed over time.

As can be seen from Table 10, the differences in word retention measured in the first isolation test are to be found predominantly between the groups that had carried out a vocabulary task and those that had not. Participants in the incidental (- test announcement) and intentional (+ test announcement) condition differed less. Retention scores in all four groups dropped on the second Isolation Test. Although the differences among the four groups became smaller on the second Isolation Test, the groups that had carried out the vocabulary task (INCID PLUS and the INTENT PLUS) still obtained higher scores than the two groups that had not (see Table 10). A two-way ANCOVA (see Table 11) revealed that only vocabulary task yielded a significant effect on word retention ($p < .0001$). In addition, its effect size was large ($\eta^2 = .21$). There was no significant difference between the incidental and intentional groups (see Table 11).

On average, participants recalled the meaning of twice as many plus-relevant as minus-relevant words on the first Isolation Test (see Table 10).

Table 10 Word retention in two Isolation Tests (Immediate Isolation Test in upper panel, Delayed Isolation Test in lower panel), by condition and word type (Max. = 16)

	<i>N</i>	All target words		Plus-relevant		Minus-relevant	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Isolation Test 1							
INCID ONLY	34	4.15	2.24	3.47	1.89	0.68	0.88
INCID PLUS	35	8.31	2.92	5.51	1.44	2.80	1.86
INTENT ONLY	33	4.91	3.58	3.85	2.45	1.06	1.48
INTENT PLUS	35	9.20	3.53	5.71	1.69	3.49	2.19
All students	137	6.69	3.76	4.66	2.12	2.03	2.04
Isolation Test 2							
INCID ONLY	27	2.41	1.62	2.07	1.24	0.33	0.68
INCID PLUS	27	3.93	1.90	3.22	1.42	0.70	0.99
INTENT ONLY	23	3.52	2.21	2.78	1.65	0.74	0.81
INTENT PLUS	24	3.83	2.58	2.71	1.49	1.13	1.51
All students	101	3.41	2.15	2.69	1.49	0.71	1.06

Note: INCID ONLY (– test announcement, – vocabulary task), INCID PLUS (– test announcement, + vocabulary task), INTENT ONLY (+ test announcement, – vocabulary task), INTENT PLUS (+ test announcement, + vocabulary task).

On the delayed test, retention scores were also significantly higher for the plus-relevant than for the minus-relevant words. The ANCOVA (see Table 11) showed that the difference was statistically significant at the .05 level. In terms of effect size, word relevance had a moderate effect on word retention.

As can be gleaned from Table 10, the ANCOVA revealed no significant interactions between test announcement and word relevance. However, this analysis revealed a significant interaction between vocabulary task and Time, which means that vocabulary loss was not the same for the groups that had carried out the vocabulary task (INCID PLUS and INTENT PLUS) and those that had not (INCID ONLY and INTENT ONLY). Its effect size was also large. The forgetting curve of both groups is presented graphically in Figure 5. Although students in the PLUS groups (+ vocabulary task) forgot more words than students in the ONLY groups, they still remembered more target words on the delayed isolation test ($p < .0001$). No other interactions were found. In Table 11, only the results related to the Time factor are reported. Nonsignificant interactions are not provided in the table.

Table 11 ANCOVA with repeated measures of isolation test

	<i>DF</i>	<i>F</i>	<i>p</i>	η^2
Within-subject				
Relevance	1	11.47	.001	.10
Time	1	1.44	.23	.01
Time × vocabulary size	1	6.13	.02*	.04
Relevance × Time	1	0.11	.74	.00
Error (within-subject)	96			
Between-subject				
Test announcement	1	2.42	.12	.02
Test announcement × Time	1	0.77	.38	.01
Vocabulary task	1	27.37	<.0001	.21
Vocabulary task × Time	1	35.99	<.0001	.25
Relevance × test announcement × Vocabulary task × Time	1	2.22	.14	.02
Error (between-subject)	96			

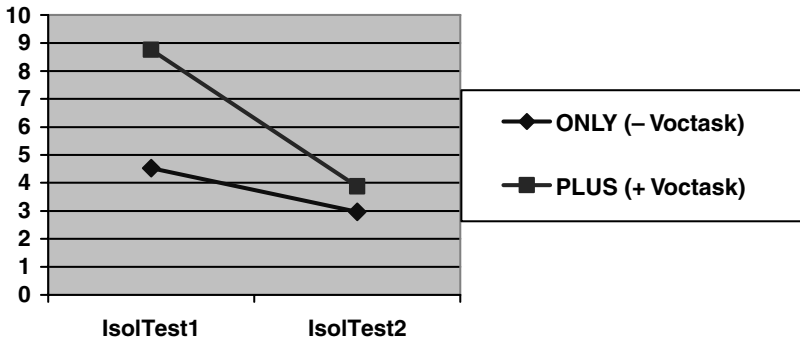


Figure 5 Line chart of interaction between time and vocabulary task on the isolation test.

*Word Retention in Context*¹⁵

All four groups scored much better on the first context test than on the first Isolation Test (see Table 12) by recalling on average four words more. A paired *t*-test indicated that the difference in retention score (between the isolation test and the context test) was statistically significant ($t = 17.62$, $DF = 136$, $p < .0001$ for the immediate tests; $t = 19.19$, $DF = 100$, $p < .001$ for the delayed posttests).

Table 12 Word retention per condition and word type on two context tests (context test 1 in upper panel, context test 2 in lower panel)

	<i>N</i>	All target words		Plus-relevant		Minus-relevant	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Context test 1							
INCID ONLY	34	8.12	1.77	6.38	1.35	1.74	0.99
INCID PLUS	35	11.77	2.18	6.77	1.21	5.09	1.60
INTENT ONLY	33	8.64	2.36	6.09	1.53	2.55	1.33
INTENT PLUS	35	12.17	2.48	7.26	0.95	5.00	1.93
All students	137	10.21	2.85	6.64	1.33	3.62	2.10
Context test 2							
INCID ONLY	27	6.04	1.60	4.63	1.24	1.41	1.05
INCID PLUS	30	8.27	1.95	5.40	1.30	2.55	1.33
INTENT ONLY	27	6.33	2.63	4.44	1.48	1.89	1.37
INTENT PLUS	25	8.72	2.69	5.48	1.19	3.24	2.05
All students	109	7.34	2.50	4.99	1.37	2.32	1.54

Note: INCID ONLY (– test announcement, – vocabulary task), INCID PLUS (– test announcement, + vocabulary task), INTENT ONLY (+ test announcement, – vocabulary task), INTENT PLUS (+ test announcement, + vocabulary task).

As can be seen from Table 12, students in the INTENT PLUS group obtained the highest retention score, whereas the INCID ONLY group obtained the lowest one. Although students' scores dropped on the second context test, it was still the INTENT PLUS group that performed best. An ANCOVA¹⁶ (see Table 13) indicated that the four groups differed significantly from each other. Of the two between-subject variables, it was solely vocabulary task that yielded a significant and large effect ($p < .0001$, $\eta^2 = .34$). Vocabulary test announcement did not affect students' retention scores when tested in context.

It is apparent from Table 12 that students recalled the meaning of almost twice as many plus-relevant words as minus-relevant words on both the immediate and the delayed context test. An ANCOVA (see Table 13) revealed that students performed significantly better on the plus-relevant than on the minus-relevant target words on the first as well as on the second context test ($p < .0001$). Moreover, the effect size was larger when the target words were tested in context ($\eta^2 = .26$; large) than when tested in isolation ($\eta^2 = .10$; moderate).

With regard to possible interactions, we found a significant interaction between relevance and vocabulary task ($p < .0001$) on the one hand and

Table 13 ANCOVA with repeated measures of context test

	<i>DF</i>	<i>F</i>	<i>p</i>	η^2
Within-subject				
Relevance	1	47.93	<.0001	.26
Time	1	9.99	.002	.08
Relevance × Voctask	1	22.35	<.0001	.12
Relevance × TestAnnounc × Voctask	1	5.24	.02	.03
Relevance × Time	1	3.39	.07	.02
Error	103			
Between-subject				
TestAnnounc	1	0.89	.35	.01
TestAnnounc × Time	1	0.07	.79	.00
Voctask	1	54.64	<.0001	.34
Voctask × Time	1	10.73	.001	.09
Relevance × TestAnnounc × Voctask × Time	1	8.22	.005	.06
Error	103			

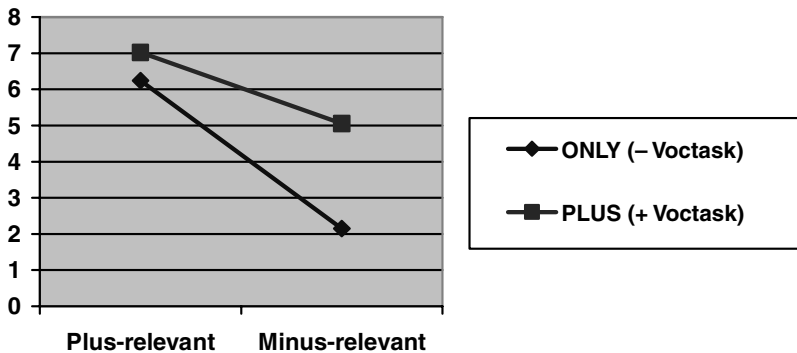


Figure 6 Line chart of interaction between vocabulary task and word relevance in context test 1.

among relevance, test announcement, and vocabulary task ($p < .05$) on the other hand (see Table 13). Concerning the former interaction (see Figure 6), students who had carried out a vocabulary task performed slightly better on the plus-relevant words but much better on the minus-relevant target words than students who had not, especially in the immediate context test. Concerning the latter interaction among the three variables (see Figure 7), students in the INTENT PLUS group obtained a significantly higher score on the plus-relevant words than the INTENT ONLY group. In addition, the groups that

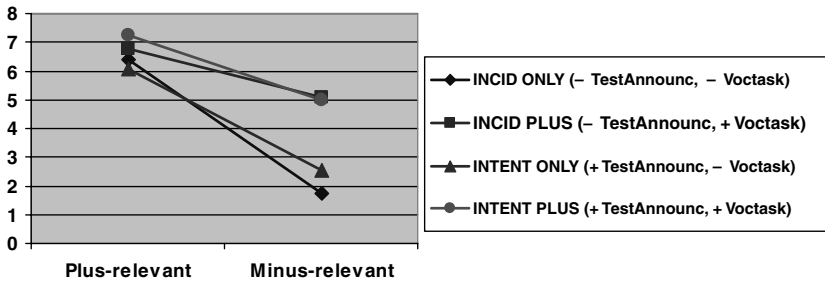


Figure 7 Line chart of interaction among test announcement, vocabulary task, and word relevance in context test 1.

had accomplished the vocabulary task (INCID PLUS and INTENT PLUS) remembered significantly more minus-relevant words than the groups that had not (INCID ONLY and INCID PLUS).

A last issue we want to address is how word retention changed over time and how the independent variables test announcement, relevance, and vocabulary task interacted with the Time factor. Table 13 shows a significant effect of Time; that is, scores on the delayed context test were significantly lower than on the immediate context test ($p < .05$). We found no interaction between Time and word relevance on the one hand and Time and test announcement on the other hand. However, the Time factor interacted significantly with vocabulary task ($p < .0001$) on the one hand and with the other three independent variables together ($p < .05$) on the other hand. Retention scores of participants who had carried out a vocabulary task dropped more drastically than those of students who had not. However, the effect was not as strong as in the Isolation Test ($\eta^2 = .09$ [moderate] vs. $\eta^2 = .25$ [large]). Concerning the second interaction, participants who had carried out the vocabulary task forgot more minus-relevant target words compared to plus-relevant ones. In addition, the INCID PLUS group forgot more minus-relevant words compared to the INTENT PLUS group (2.50 vs. 1.75). This interaction is presented graphically in Figure 8.

To sum up, we found that participants who were forewarned of an upcoming vocabulary test recognized the word forms of both plus-relevant and minus-relevant target words (in the recognition test) better than participants not forewarned. Participants who accomplished a vocabulary task also performed better on this test compared to participants who did not do this task. Word retention in the immediate and delayed isolation test and context test was affected positively by word relevance as well as the vocabulary task. Participants’

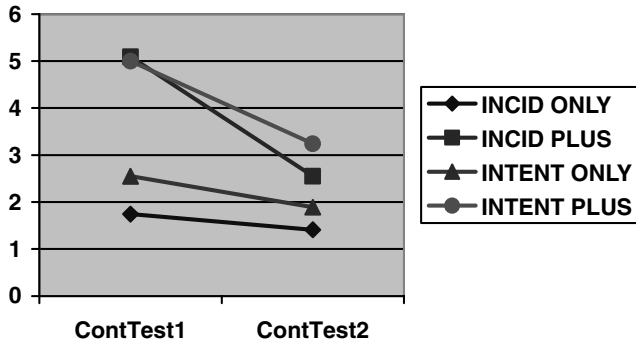


Figure 8 Line chart of interaction between the three independent variables and Time in context test.

retention scores were lower on the delayed than on the immediate posttests, but the difference was only significant in the case of the context test. In addition, participants who had carried out the vocabulary task had a steeper forgetting curve than participants who had not. This was the case in the isolation test as well as the context test.

Discussion

In this section, we discuss our findings, first with respect to word look-up behavior and then with respect to word retention.

Look-up Behavior

A first aim of this study was to investigate whether two enhancement techniques—telling students that the reading task will be followed by a vocabulary test (vocabulary test announcement) and forcing students to pay attention to unfamiliar words in the reading text via comprehension questions (task-induced word relevance)—would have an effect on students' look-up behavior. This study provides evidence that L2 learners look up more target words when they know that a vocabulary test will follow, which is in line with the findings of Peters (2007a, 2007b). In addition, they look up more plus-relevant than minus-relevant words. Hence, this study corroborates the findings of other studies (Hulstijn, 1993; Laufer & Levitzky-Aviad, 2003; Peters, 2007b). With regard to the interaction between vocabulary test announcement and task-induced word relevance, participants forewarned of an upcoming vocabulary test were more inclined to consult the meaning of minus-relevant target words

compared to students not forewarned. In short, we found evidence that these two techniques prompt L2 learners to find the meaning of unknown words. The effects sizes were moderate to large.

In addition to the finding that students in the intentional groups (+ test announcement) looked up target words more often than students in the incidental groups (– test announcement), we found that individual students differed greatly in their frequency of clicks on annotated words, ranging from 9 to 177 clicks. Hulstijn (1993) obtained a similar look-up variability, with clicks ranging from 1 to 103. However, more frequent clicking did not result in better word retention because the correlations between participants' frequency of clicks on target words and word retention were low for all posttests. We also observed that some students did not consult the meaning of all eight plus-relevant target words while reading the text. It is therefore no coincidence that these students, in that case, did not provide a correct answer to the question to which the plus-relevant target word pertained.

Compared to test announcement, word relevance had a larger effect on students' dictionary use, which was reflected in the effect sizes. Its effect was particularly large when all clicks were counted. Moreover, the strong effect of task-induced word relevance became very obvious during the postreading vocabulary task. Students looked up the minus-relevant target words more often than the plus-relevant target words. Even though students could verify the meaning of the plus-relevant words, they seemed to be quite sure about their knowledge of these target words, which was indeed corroborated by their scores on the vocabulary task. Consequently, they felt no need to verify their answer.

Previous research has shown that the provision of an online dictionary can facilitate word retention (Chun & Plass, 1996; De Ridder, 2002; Knight, 1994; Laufer & Hill, 2000; Peters, 2007b; Yoshii, 2006). This study and the studies reviewed above demonstrate that telling students that the reading task will be followed by a vocabulary test (vocabulary test announcement) and forcing students to pay attention to unfamiliar words in the reading text via comprehension questions (task-induced word relevance) have the potential of prompting L2 learners to actually use the look-up facility. However, word relevance is more effective because it directs L2 learners' attention to specific words (i.e., the plus-relevant words). Test announcement is rather an indirect way of drawing students' attention to vocabulary when they read a text, whereas word relevance produces a more direct effect because students were required to use the plus-relevant words in their answer to the comprehension questions.

Word Retention

Our second aim was to explore whether three enhancement techniques—telling students that the reading task will be followed by a vocabulary test (vocabulary test announcement), forcing students to pay attention to unfamiliar words in the reading text via comprehension questions (task-induced word relevance), and requiring students to pay attention to the unfamiliar words again, after completion of the reading task (vocabulary task)—would promote word learning when students read an L2 text. This study demonstrates that target words that were looked up, processed when answering the comprehension questions, and repeated in the postreading vocabulary task (as was the case for the plus-relevant target words in the groups that had carried out the vocabulary task [PLUS groups]) were better retained than target words that were only looked up and processed for either the comprehension questions or the vocabulary task (as was the case for the plus-relevant target words in the ONLY groups [–vocabulary task] and for the minus-relevant words in the PLUS groups [+vocabulary task]). The lowest retention scores were found for the minus-relevant target words among the participants in the ONLY groups (–vocabulary task), who had the opportunity to look the words up but did not have a task-induced incentive to do so.

Schematically, we have the following:

Discovering meaning + elaborate processing + repeating
 > discovering meaning + elaborate processing
 > discovering meaning

We will first discuss the effect of test announcement and why it was less effective than word relevance before explaining the findings with regard to the vocabulary task.

Although the presence of a *vocabulary test announcement* positively affected students' ability to recognize whether a word had occurred in the text, effective learning additionally requires focused attention and elaborate processing. A critical feature of human memory appears to be how one processes information, not why one processes information (Baddeley, 1997; Craik, 2002) (e.g., because one knows that a vocabulary test will follow). Thus, we conclude that although vocabulary test announcement had an effect on students' *recognizing* the word form, it did not induce the kind of processing that resulted in participants being able to *recall* the target words' meaning.

As in Peters (2007), *task-induced word relevance* had a moderate to large effect on word retention as measured in the immediate and delayed (= 2 weeks

later) recall posttests. First, we discuss why word relevance did not have an overall significant effect on the recognition test; then we interpret the positive findings of word relevance in the two recall tests.

Word relevance affected performance on the recognition test only in the case of participants who had not carried out the vocabulary task. These participants were made aware of the target words via test announcement and the comprehension questions, which made them pay attention predominantly to the plus-relevant words. In contrast, performance on the recognition test of the PLUS groups (+ vocabulary task) was not affected by word relevance only. The postreading vocabulary task made them pay attention to both the minus-relevant and plus-relevant words.

We provide two explanations why the technique word relevance had a larger effect than test announcement, especially in the recall tests. First, in line with the Involvement Load Hypothesis (Laufer & Hulstijn, 2001; see Introduction), the text comprehension questions created a *need* to discover the meaning of the plus-relevant target words. The value of this need was moderate because it was externally driven by the questions. The minus-relevant target words, on the other hand, were not targeted by the text comprehension questions. As a consequence, there was no externally driven need to discover their meaning. Log files revealed that participants clicked more frequently on the plus-relevant target words than on the minus-relevant ones (= *search*). The second reason for the better results of word relevance compared to test announcement can be found in the fact that the comprehension questions (Word relevance) forced students to *do* something with the information looked up. In order to answer these questions, students had to pay attention to the semantic role of the target words in their context and decide on their semantic relevance to answer the comprehension questions. These activities fall in the realm of elaborate processing. In a study conducted by de la Fuente (2006), the effect of three tasks, differing in their degree of guidance, on productive use of target words was investigated. Participants were randomly assigned to one of three conditions. The results indicate that when involved in free interactive activities (= unguided) students tended not to incorporate target words, whereas they did so in the case of a preplanned, guided, and well-focused task. This resembles the distinction between test announcement and word relevance. The open, free nature of test announcement has the potential to increase students' number of look-ups. However, test announcement is less compelling than word relevance, because, over the course of a 35-min reading activity, students are more likely to neglect or forget the rather implicit instruction that they will be tested on their knowledge of words used in the text than that they will refuse to

answer text comprehension questions when requested to do so. Furthermore, vocabulary test announcement did not tell students on *which* words to focus their attention on. Students are cooperative when it comes to answering questions, but they probably simply forget to remain focused on word learning while reading a text when only told that a vocabulary test will follow the reading task. This was confirmed by participants' answers to one of the retrospective questions. Thus, a task feature that *specifically* enforces learners to process new words (asking students to answer comprehension questions pertaining to new words) produces more effect than a task feature that does so only *generally* (announcing that a vocabulary test will follow). In other words, students allocate their attentional resources in function of the specificity of the task they have to perform. Test announcement at the beginning of the reading task produced rather little "intentional" learning behavior. In the retrospective questions, only 15 of the 68 students in the intentional groups said that they had tried to memorize the dictionary information while reading the text. They reported hardly any other learning strategy, except for using the context as a retrieval cue. This illustrates the view among most cognitive psychologists that "memory performance is determined far more by the nature of the processing activities engaged in by the learner than it is by the intention to learn per se" (Eysenck, 1982, p. 203).

A last issue with respect to word relevance that we want to draw attention to is the fact that participants performed better on the context test than on the isolation test, on the immediate as well as delayed tests. They recalled the meaning of two more plus-relevant target words in the context test than in the isolation test. This can be explained by the similarity between learning task and retention test (Bransford, Franks, Morris, & Stein, 1979; Lotto & de Groot, 1998). The context test and the comprehension questions were similar with regard to the sentences in which the plus-relevant target words occurred. By seeing the sentences again, students could recall the meaning of more plus-relevant words than when no context was provided. These sentences acted as retrieval cue. This explains, we believe, among others why the effect size of word relevance was larger when the target words were tested in context than when tested in isolation.

Participants who had carried out a *postreading vocabulary task* (INCID PLUS and INTENT PLUS groups) outperformed the other two groups in the number of words recalled both in the immediate and delayed posttests. In these groups, word retention was not only affected by word relevance but also by the vocabulary task. The effect of the vocabulary task was so strong that it, among others, may have undone the effect of relevance in the recognition test

(see above). Its effect size was large in all posttests. We will first interpret the effect of vocabulary task on retention of the minus-relevant target words before discussing the findings with respect to the plus-relevant target words.

Participants in the PLUS groups, who had carried out the vocabulary task, recalled the meaning of more minus-relevant target words than participants in the ONLY groups. This is supporting evidence for the findings obtained by Paribakht and Wesche (1997, 1998) and Laufer (2003) that a word-focused activity after having read a text yields better results than only reading a text. If students had not seen or looked up one of the minus-relevant target words during the reading task, they were prompted to look it up and process it (for the first time) in the vocabulary task. Thus, one could say that the minus-relevant words turned into “relevant” words, so to say, in the vocabulary task. The difference in retention between the ONLY (– vocabulary task) and PLUS groups (+ vocabulary task) was particularly large on the immediate vocabulary posttests. Although the PLUS groups forgot more minus-relevant target words than the ONLY groups on the delayed tests, they still performed better. It seems that the more one learns, the more one can forget (Hulstijn, 2003). Yet, surprisingly, this only applies to the minus-relevant and not to the plus-relevant words.

Participants who had carried out the vocabulary task (INCID PLUS and INTENT PLUS groups) also recalled the meaning of more plus-relevant target words compared to participants who had not. The positive findings with regard to the PLUS groups can be explained in terms of repetition and consolidation of words previously encountered and processed. The vocabulary task allowed the learners to consolidate their knowledge of the form-meaning connections of the plus-relevant words. Repetition is important because it has the potential to lead to consolidation of word knowledge (Hulstijn, 2001; Laufer, 2003, 2005; Nation, 2001). In the present study, participants who had performed the vocabulary task processed the plus-relevant target words first when answering the comprehension questions. Next, they processed them again in the vocabulary task, immediately after the text comprehension task. This immediate additional encounter with the words consolidated the form-meaning connections of the plus-relevant target words, resulting in better word retention (see Hulstijn, pp. 277–280, with references to work on rehearsal regimes). This also explains why the groups that had carried out the vocabulary task still performed better than the other two groups on the delayed posttests. In addition, the former groups had forgotten fewer plus-relevant words than the latter when tested in context.

Better performance on the plus-relevant words can additionally be explained by a retrieval practice effect (Baddeley, 1997). Although many students did not consult the online look-up facility for the plus-relevant target words while

doing the vocabulary task, we surmise that they still retrieved these word meanings covertly (and successfully). Each successful retrieval strengthens the retrieval route and hence the form-meaning connection of the word. This retrieval occurred rarely or not at all for the minus-relevant words because many students looked up their meaning for the first time while completing the vocabulary task.

Participants who had carried out the vocabulary task also performed better on the context test than on the isolation test. As was the case for the plus-relevant words in the context test (see above), this finding can be explained in terms of the similarity between the learning task (= the vocabulary task) and the test (= Context Test) (Bransford et al., 1979; Lotto & De Groot, 1998). The vocabulary task and the context test were almost identical because the target words were presented as they had occurred in the text. The form-meaning connection of some words was reactivated when participants saw again the context in which the target words had occurred. In the questionnaires, several participants referred to the context as a retrieval technique by memorizing the sentences in which the target words had appeared. This could explain the higher effect size of vocabulary task in the context test than in the isolation test.

Implications for Teaching

Our findings, together with those of the studies reviewed, provide robust evidence that the low incidence of vocabulary acquisition through reading (“input only”) can be substantially boosted by techniques that make students *look up* unknown words, process their form-meaning relationship *elaborately*, and process them *again* after reading (“input plus”). The preceding paragraphs have dealt extensively with the grounds for this conclusion. We deem our conclusion relevant for L2 instruction. Simply telling students that they will be tested later apparently does not bring about the three forms of behavior required for word learning (finding out what the word means, elaborate processing of the word’s form together with its meaning, and repetition). If we want L2 learners to enlarge their vocabulary size through reading, we need techniques that require students to perform concrete word-directed actions. Asking text comprehension questions that *force* students, as it were, to look up an unfamiliar word, as well as assigning a repetition task that *forces* students, as it were, to reprocess the form-meaning link that was processed a while ago during reading were each shown to be more effective than announcing that unfamiliar words will be tested after reading because the former two techniques lead to concrete, word-specific action, whereas the latter does not or does so only among some students for some words. Furthermore, previous studies and the present one have shown

that word learning is not a matter of either elaborate processing or repetition, but both. A one-time elaborate processing of a new form-meaning link runs the risk of falling into oblivion. We therefore recommend the assignment of a reprocessing task immediately after completion of a reading comprehension task.

Future Research

We would like to formulate some suggestions for further research. This study was not specifically set up to examine the differential effects of vocabulary task on the one hand and word relevance on the other hand. It would be interesting to investigate the effect of either variable on word retention empirically because the drop in retention scores on the delayed vocabulary posttests was not the same for the plus-relevant and the minus-relevant target words. In other words, the question of whether both enhancement techniques induce the same type of processing needs to be investigated. Another area that merits further research is productive vocabulary knowledge. We found encouraging findings for receptive word learning. Yet, we need to address the question of whether these enhancement techniques would also benefit productive word learning. In this respect, future studies might also investigate the effect of enhancement techniques not only on individual lexical items but also on collocations. Future studies using a similar experimental setup would help us to establish a greater understanding of how to foster vocabulary learning.

Conclusion

This study is the first one investigating the effects of three enhancement techniques and their interactions in a single experiment: (a) announcement of a postreading vocabulary test (vocabulary test announcement), (b) a specific task-induced technique to highlight the relevance of unfamiliar words for text comprehension by having students answer text comprehension questions forcing them to elaborately process unknown words (word relevance), and (c) a postreading vocabulary task, which forced students to process the target words once again (vocabulary task).

Concerning students' look-up behavior, we found evidence for the effect of vocabulary test announcement, which affected students' willingness to consult the meaning of an unknown word in an online dictionary more often. In addition, this study corroborated the findings of previous studies when it comes to task-induced word relevance. Plus-relevant words were looked up more than minus-relevant words. Vocabulary test announcement had an effect on students'

recognition of the word form. Otherwise, we did not find evidence that test announcement resulted in better word retention. Word relevance, but especially the combination of vocabulary task and Word relevance, promoted vocabulary learning.

In conclusion, successful L2 vocabulary acquisition is contingent on three factors: finding the meaning of an unfamiliar word (which may produce only a fleeting form of processing), subsequent elaborate processing, and repetition (Hulstijn, 2001). Both vocabulary test announcement and word relevance prompted students to *look up* words. The word-directing comprehension questions (word relevance) induced *elaborate processing*. The vocabulary task enforced the *repetition* of the form-meaning connections of the (plus-relevant) target words, and this turned out to be a successful reinforcement of the previous elaborate processing.

Revised version accepted 7 March 2008

Notes

- 1 We will use the label *second language* (L2) for second as well as foreign language.
- 2 We acknowledge that, as one reviewer of this article remarked, the combination of all three techniques may not have much value in the real world of L2 instruction, but the combination of any two techniques might well be relevant to L2 instruction.
- 3 A computer scientist programmed a Java application into the QuestionMark Perception software (<http://www.questionmark.com>) to register students' look-up behavior.
- 4 The text was pretested among first-year students of German to ensure that the text would not be too difficult.
- 5 One reviewer objected that the type of text we selected was not conducive to vocabulary acquisition because the target words occurred only once, in contrast to, for instance, a graded reader, which would offer more occurrences with the target words. This was pointed out as a limitation. However, we do not agree with this because the participants we recruited were high-intermediate to advanced learners of German being familiar with this type of authentic texts. A feature of these texts is that they contain words that are not frequent and that tend not to occur more than once in the text.
- 6 In the presentation of the text on the computer screen, none of the target words were visually enhanced, such as in bold typeface.
- 7 We only included items from the frequency bands of 4,000 to 10,000 words. No items from the first 4,000 words were included because the participants were high-intermediate to advanced students. Vocabulary size could be a learner variable interfering in the vocabulary acquisition process. It correlated with participants'

word retention in the posttests. Therefore, we wanted a test that would allow us to discriminate among participants.

- 8 One reviewer suggested using Bonferroni tests for multiple testing. These tests did not change our findings (= significant results remained significant after correcting for multiple testing). However, we do not report the figures of the Bonferroni tests because they only take into account the main effects and not the interactions between variables.
- 9 We conducted a two-sample *t*-test on the basis of Test announcement. Since the vocabulary task was only announced after students had finished the text comprehension questions, this variable was not taken into consideration in the analysis.
- 10 The $DF = 130$ and not 135 because some students had forgotten to write down the time when they had finished the comprehension task.
- 11 The ANCOVA is based on data of 135 instead of 137 participants. One student did not take the vocabulary size test. Of another student, we only had information on which target words she looked up and not the frequency of clicks on target words. This was also the case for the time devoted to the target words.
- 12 We did not conduct an ANCOVA because we did not consider vocabulary size important while carrying out the vocabulary task.
- 13 In contrast to the ANCOVA, a paired *t*-test, which does not take into account the other variables or any possible interaction between the variables, did reveal a significant effect for Word relevance ($t = -5.79$, $DF = 136$, $p < .0001$).
- 14 We also computed an ANCOVA per test session. Our results are the following for Isolation Test 1: for Test announcement, $F(1, 135) = 1.88$, $p = .17$, $\eta^2 = .01$; for Vocabulary task, $F(1, 135) = 58.32$, $p < .0001$, $\eta^2 = .30$; and for Word relevance, $F(1, 135) = 10.87$, $p = .001$, $\eta^2 = .07$. The results for Isolation Test 2 are as follows: for Test announcement $F(1, 100) = 1.48$, $p = .23$, $\eta^2 = .01$; for Vocabulary task, $F(1, 100) = 4.70$, $p = .03$, $\eta^2 = .05$; and for Word relevance, $F(1, 100) = 10.30$, $p = .002$, $\eta^2 = .09$.
- 15 We also computed an ANCOVA per test session. Our results are the following for Context Test 1: for Test announcement, $F(1, 135) = 0.89$, $p = .35$, $\eta^2 = .00$; for Word relevance $F(1, 135) = 43.15$, $p < .0001$, $\eta^2 = .18$; and for Vocabulary task, $F(1, 135) = 86.93$, $p < .0001$, $\eta^2 = .40$. The results for Context Test 2 are as follows: for Test announcement, $F(1, 107) = 0.54$, $p = .46$, $\eta^2 = .00$; for Word relevance, $F(1, 107) = 24.26$, $p < .0001$, $\eta^2 = .19$; and for Vocabulary task, $F(1, 107) = 27.20$, $p < .0001$, $\eta^2 = .21$.
- 16 The ANCOVA of Context Test 1 is based on the data of 136 students because 1 student did not take the Vocabulary Size Test; the ANCOVA of the Context Test 2 is based on the data of 108 participants for the same reason. In one school, not all students took Isolation Test 2, which explains why we have data of more participants in Context Test 2.

References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Baddeley, A. (1997). *Human memory: Theory and practice* (Rev. ed.). Hove, UK: Psychology Press.
- Bogaards, P., & Laufer, B. (2004). Introduction. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language. Selection, acquisition, and testing* (pp. vii–xiv). Amsterdam: Benjamins.
- Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 331–354). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chun, D., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal, 80*, 183–198.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future? *Memory, 10*, 305–318.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.
- de la Fuente, M. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research, 10*, 263–295.
- De Ridder, I. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension and the reading process? *Language Learning & Technology, 6*, 123–146. Retrieved January 21, 2003, from <http://lilt.msu.edu/vol6num1/DERIDDER/default.html>
- Eysenck, M. W. (1982). Incidental learning and orienting tasks. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 197–228). New York: Academic Press.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*, 207–223.
- Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal, 77*, 139–147.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.

- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, *80*, 327–339.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal*, *78*, 285–299.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Harlow, UK: Longman.
- Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language. *Prospect*, *16*, 44–54.
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*, 567–587.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. In S. H. Foster-Cohen, M. Garcia-Mayo, & J. Cenoz (Eds.), *Eurosla yearbook* (Vol. 5, pp. 223–250). Amsterdam: Benjamins.
- Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, *3*, 58–76. Retrieved October 29, 2002, from <http://lt.msu.edu/vol3num2/laufer-hill/index.html>
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*, 1–26.
- Laufer, B., & Levitzky-Aviad, T. (2003). *Look-up behavior and word retention as a function of task type and word relevance*. Paper presented at the biennial ASIALEX conference, Tokyo, Japan.
- Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69.
- Nagy, W. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 64–83). Cambridge: Cambridge University Press.
- Nagy, W., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly*, *20*, 233–253.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–200). Cambridge: Cambridge University Press.

- Paribakht, T. S., & Wesche, M. (1998). "Incidental" and instructed L2 vocabulary acquisition: Different contexts, common processes. In D. Albrechtsen, B. Henriksen, I. Mees, & E. Poulsen (Eds.), *Perspectives on foreign and second language pedagogy: Essays presented to Kirsten Haastrup on the occasion of her sixtieth birthday* (pp. 203–220). Odense, Denmark: Odense University Press.
- Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21, 195–224.
- Peters, E. (2007a). The influence of task instruction on vocabulary acquisition and reading comprehension. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 178–198). Clevedon, UK: Multilingual Matters.
- Peters, E. (2007b). Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning & Technology*, 11(2), 45–67. Retrieved July 1, 2007, from <http://llt.msu.edu/vol11num2/peters/default.html>
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146–161.
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6, 72–78.
- Sercu, L., Dewachter, L., Peters, E., Kuiken, F., & Vedder, I. (2006). The effect of task complexity and task conditions on foreign language development and performance. Three empirical studies. *ITL, International Journal of Applied Linguistics*, 152, 55–84.
- Staud, T. (2003, October). Osis sind Türken. 13 Jahre Einheit: In Gesamt-Westdeutschland sind die Ostdeutschen Einwanderer. *Die Zeit*, 41. Retrieved October 2, 2003, from <http://www.zeit.de/2003/41/Einwanderer>
- Swanborn, M. S. L., & De Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69, 261–285.
- Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10, 85–101. Retrieved October 27, 2006, from <http://llt.msu.edu/vol10num3/yoshii/default.html>