Canadian Journal of Applied Linguistics
Revue canadienne de linguistique appliquée

Canadian Journal of Applied Linguistics (CJAL)
Revue canadienne de linguistique appliquée (RCLA)

Effectiveness in L2 Vocabulary Study – A Classroom-based Investigation of Deliberate Learning

Paul Pauwels

Explore this journal

Cite this article

Article abstract

Deliberate vocabulary study has mostly been studied within a strictly experimental framework of learning and memorization. More ecologically valid investigations embedded in existing study contexts have been rare. This study fits into the latter paradigm, investigating how students attempted to learn 90 English words over a period of three weeks and tracking their efforts via study logs and intermediate receptive and productive tests, with final testing five weeks after the study period. The results are in line with findings from earlier research. Study logs showed students mainly relied on different kinds of repetition and retrieval. Selective attention for specific items was an important predictor for short-term learning, and sufficient spacing was the most important predictor for longer-term learning. From a pedagogical point of view, a point of attention is that students mostly practised retrieval after first repeating, making retrieval less difficult and creating an impression of knowledge.

Effectiveness in L2 Vocabulary Study – A Classroom-based Investigation of Deliberate Learning

Paul Pauwels
KU Leuven

Abstract

Deliberate vocabulary study has mostly been studied within a strictly experimental framework of learning and memorization. More ecologically valid investigations embedded in existing study contexts have been rare. This study fits into the latter paradigm, investigating how students attempted to learn 90 English words over a period of three weeks and tracking their efforts via study logs and intermediate receptive and productive tests, with final testing five weeks after the study period. The results are in line with findings from earlier research. Study logs showed students mainly relied on different kinds of repetition and retrieval. Selective attention for specific items was an important predictor for short-term learning, and sufficient spacing was the most important predictor for longer-term learning. From a pedagogical point of view, a point of attention is that students mostly practised retrieval after first repeating, making retrieval less difficult and creating an impression of knowledge.

Résumé

L'apprentissage intentionnel du vocabulaire a été principalement étudié dans un cadre strictement expérimental d'apprentissage et de mémorisation. Des recherches avec une validité écologique plus ancrées dans des contextes authentiques d'étude sont rares. Cette étude s'inscrit dans ce dernier paradigme; elle analyse la manière dont les étudiants ont tenté d'apprendre 90 mots anglais sur trois semaines en suivant leurs efforts répertoriés dans les journaux d'étude et en effectuant des tests intermédiaires réceptifs et productifs, avec un test final après cinq semaines. Les journaux d'étude ont montré que les étudiants se sont principalement appuyés sur différents types de répétition et de récupération. Une attention sélective pour des éléments spécifiques était un prédicteur important pour l'apprentissage à court terme, et un espacement suffisant était le prédicteur le plus important pour l'apprentissage à plus long terme. D'un point de vue pédagogique, il convient de noter que les étudiants pratiquaient principalement la récupération après la première répétition, ce qui crée une impression de connaissance.

Effectiveness in L2 Vocabulary Study – A Classroom-based Investigation of Deliberate Learning

Deliberate vocabulary study is an important part of language learning (Nation 2013; Schmitt 2008) and its effectiveness has long been demonstrated (Thorndike 1908; Webb 1962). Investigation into effective ways of deliberate study has followed different paths. On the one hand, there has been questionnaire- and interview-based research into the variety of strategies students use and find helpful (Sanaoui 1995; Schmitt 1997). Correlating strategy preferences with (vocabulary) proficiency tests, Gu and Johnson (1996) and Fan (2003) also attempted to identify which were the productive strategies or strategy profiles. The major limitation of this kind of research is that it considers strategy use as a user characteristic and does not relate it to specific learning tasks.

The other major line of research finds its origins in psychological experiments into learning and memorization (Ellis & Beaton 1993; Papagno et al. 1991; Thomas and Dieter 1987; see Baddeley et al. 1998 for a survey). Such experiments usually present a limited number of words to learners in highly controlled one-off interventions, comparing two or three types of study behaviour, with mostly immediate and delayed post-tests after only one week. The participants are usually not language students, and foreign language vocabulary is taken as a test-case for learning. The number of items is determined by the format of the experiment, and the vocabulary is mostly presented to the learners for a fixed number of seconds on slides or on PC. The strength of this type of research lies in its ability to demonstrate the differential effect of very specific study techniques on the test results. Still, such experiments tend to have limited ecological validity, since students' study behaviour is not usually restricted to one technique or to one specific intervention, and exposure times in normal study contexts vary. There are also possible motivational issues with one-off studies that are not embedded in regular studying contexts. Experimental research into deliberate studying has also been conducted in language learning contexts (Barcroft, 2006, 2009; Elgort, 2011; Nakata, 2015, 2016; Webb, 2005, 2009; Webb and Piasecki, 2018). Such experiments still struggle with ecological validity because they use short exposure times to present words on screen or use nonsense words to exclude previous knowledge effects. More recent studies (Elgort, 2011; Nakata, 2016; Webb & Piasecki, 2018) indicate that they are aware of this and attempt to get closer to realistic studying behaviour, for example by relaxing time constraints (Webb & Piasecki, 2018).

Classroom-based research into vocabulary learning strategies, such as Griffin and Harley's (1996) and Mondria and Wiersma's (2004) investigations into receptive and productive learning, Schuetze and Weimer-Stuckmann's (2011) and Schuetze's (2015) investigations into spacing effects over several weeks or Laufer and Schmueli's (1997), Mondria's (2003) and (Pauwels 2012, 2018) exploring the effect of different types of input, is comparatively rare. Such studies cannot control study behaviour to the same extent as the experimental studies but gain in ecological validity because they are embedded in regular class activity or mirror regular study assignments.

This study aims to trace the effects of different kinds of vocabulary learning strategies on immediate and longer-term learning in the context of a regular study assignment and strives to maximize ecological validity. The assignment consisted of learning a 90-word vocabulary which was part of the standard syllabus. Students were invited to volunteer for the study and were promised feedback on their results in order to increase motivation. Learning took place during one supervised one-hour study session and

three unsupervised one-hour study sessions during which students kept logbooks. Learning was tested immediately after the first study session and after the three study sessions and in a five-week delayed post-test.

Literature Review

Popular Study Strategies

Studies of general strategy use in vocabulary learning have yielded divergent results with respect to the popularity of strategies and their (believed) usefulness. Schmitt's (1997) survey of Japanese EFL students showed that memorization strategies like repetition (oral or written) and paying attention to spoken and written form were among the most frequently used strategies (usage scores around 70%) that were also judged most useful (> 80%), while deeper strategies like the keyword method or using semantic maps were much less popular and judged as less useful. Flashcards were among the less popular strategies (25%). Gu and Johnson's (1996) survey of Chinese EFL students shows that students do not really believe in memorization strategies or encoding strategies linking the word to previous knowledge of some kind, that they do believe in the importance of metacognitive regulation and that they attach more importance to strategies that help them discover the meaning of words. Fan (2003) surveying Hong Kong Chinese EFL learners seems to confirm Gu and Johnson's (2006) results, and especially students who scored highest on a vocabulary proficiency test considered memorization strategies as less important (o.c. 231). There are several possible reasons for the discrepancy between Schmitt (1997) and Gu and Johnson (1996): the existence of different learning cultures, students giving socially acceptable responses reflecting the different learning cultures, or, in the case of Gu and Johnson (1996) also the difficulty of making decontextualized comparative judgments about strategies that are used in different stages of the learning process (see also Nation, 2013, p. 334).

A general survey on learning strategies among US undergraduates seems to confirm the popularity of repetition over deeper strategies. Karpicke et al. (2009) found that in answer to an open question about study strategies, 84% of their students said they used rereading – it was the number one strategy for 55%; 43% did practice problems and 40% used flashcards. More elaborative strategies like rewriting, using mnemonics, or making outlines were much less popular. Only 11% indicated that they would self-test. A follow-up question forcing students to choose between rereading and testing themselves showed that only 18% would self-test. In a review on self-regulated learning, Bjork et al. (2013) contrast this latter finding with other survey research that indicates self-testing is popular; however, students seem to consider it as a way of evaluating their learning rather than as a study strategy. They list a number of misunderstandings that impair effective studying behaviour; a preference for passive learning strategies like repeated reading, the avoidance of errors, and the mistaken belief that fluency (ease of learning, ease of short-term recall) equals better knowledge leads to unproductive studying behaviour such as massed learning and avoidance of self-testing.

Two studies investigated students' preferences in the context of specific learning assignments (Barcroft, 2009 and Lawson and Hogben, 1996). In both cases, though, the limited study time imposed limits on the types of activity students could perform. This is certainly the case in Barcroft (2009), where students studied 24 Spanish word-picture pairs

that were each presented for 6 seconds twice – a little more than 6 minutes in total – and answered questions about their strategy use afterwards. The most frequently used strategies involved establishing some kind of form-meaning link (L2-picture, L2-L1 translation), L2-L1 association (form-form link) and repetition. Results on immediate productive (picture-L2) and receptive (L2-L1) post-tests showed that less frequently used strategies (mnemonics, visualizing, silent repetition and productive self-testing) were more effective. In Lawson and Hogben (1996) participants studied 12 Italian nouns from flashcards providing a sentence context and an L1 translation while performing a think-aloud protocol. The time spent studying was not controlled and ranged from 17 to 38 minutes. The most frequently used strategies were repeated reading and oral repetition. Elaboration techniques like mnemonics or paraphrasing were used less often. Oral repetition and elaboration techniques correlated positively with results on an immediate post-test. Self-testing was used by only one student – the top scorer.

Effective Study Strategies

Direct investigations of the effectiveness of study activities have, as mentioned above, mostly been conducted in the framework of learning and memory studies. One of the most important findings of learning research has been the so-called testing effect, i.e. that tests can actually promote memory for the materials tested (see Roediger & Butler, 2011, for a review). This effect has also been demonstrated in learning foreign language words (Kang et al., 2013; Karpicke & Roediger, 2008; Pyc & Rawson, 2012; Soderstrom et al., 2016). In Karpicke and Roediger (2008) participants studied 40 Swahili-English word pairs under four conditions in which they dropped words from either study and/or testing or not. Results on a 1-week delayed receptive test showed that dropping words from testing led to much lower retention rates (30% vs 80%) irrespective of whether words were restudied or not while dropping words from studying had no differential effect. The conclusion is clear: repeated retrieval is much more effective than mere repetition. Pyc and Rawson (2012) reach a similar conclusion when participants study Swahili-English word pairs using the keyword method, and Kang et al (2013) demonstrate a similar effect for studying Hebrew-English word pairs through oral repetition. Soderstrom et al. (2016) provide a nuance with Karpicke and Roediger's (2008) findings. Using, again, Swahili-English word pairs, they conducted a replication yielding similar results, and a variation on that experiment manipulating the spacing between encounters. In this second experiment, repeated study did enhance learning when encounters were properly spaced, but its effect does not appear to be stronger than the retrieval effect.

Spacing, or more specifically, the difference between massed and spaced study has also received considerable attention in learning research. On the basis of a quantitative meta-analysis, Cepeda et al. (2006: 371) conclude that "spaced (vs. massed) learning of items consistently shows benefits, regardless of retention interval, and learning benefits increase with increased time lags between learning presentations". Several recent experiments have investigated the effect of retrievals in different spacing schedules. Again, some of these involved foreign language vocabulary as a test case (Kornell and Bjork, 2009; Nakata, 2015; Pyc and Rawson, 2007, 2009). Pyc and Rawson (2007) report on three experiments involving the study of 24 Swahili-English word pairs comparing different degrees of spacing of test-restudy trials (test and restudy after 5 words, after 23 words, at varying intervals). In some conditions, correct recalls were dropped from further study to

mirror the behaviour of students using flashcards. Their result show that different spacing schedules do not affect results and that dropping words after correct recall is more efficient in terms of time spent and maybe slightly more effective if there remains enough space between encounters of the same word. In another study on the use of flashcards, Kornell and Bjork (2009) focused on the effect of dropping known words from study and found that dropping a word from study after a single correct recall had a negative effect on a receptive test and that students tended to drop words too early (in some cases even without having recalled them correctly). Allowing students to drop flashcards had a slight but non-significant negative effect on results. Pyc and Rawson (2009) manipulated the spacing between correct retrievals and the number of correct retrievals students needed to perform before dropping an item from study, and found that more difficult retrievals (i.e., with longer intervals) contributed more to learning; the effect of multiple correct retrievals was less strong and levelled off. Like Pyc and Rawson (2007), Nakata (2015) focused on the effect of different types of spacing and compared equal and expanding spacing using different numbers of intervening trials. The study found a significant advantage of expanding over equal spacing on an immediate and a 1-week delayed receptive test, but effect sizes and mean gains were small. There were clear advantages for spaced learning over massed learning, except when spacing was short – i.e., from one to nine intervening trials. The study found no differential effects of spacing on productive learning except for an advantage of short spacing on the immediate post-test; productive learning decayed significantly from the immediate to the delayed post-test. Students' learning phase performance provided no indication for their longer-term retention.

In research with a language learning focus, the effectiveness of different types of repetition has been a major point of attention (Abbs et al., 2008; Barcroft, 2006; Ellis and Beaton, 1993; Kaushanskaya and Yoo, 2011; Thomas and Dieter, 1987; Van Hell and Candia-Mahn, 1997; Webb and Piasecki, 2018). Thomas and Dieter (1987) conducted three experiments on learning FrenchL2-EnglishL1 word pairs through either oral or written repetition. They found that oral repetition did not result in better learning than the control condition (no repetition), and that written repetition did yield better results, but only on productive learning – they concluded that written rehearsal improves form knowledge, but not knowledge of the form-meaning link. In his experiments on learning Spanish words via word-picture pairs with or without writing Barcroft (2006) found that the writing condition had a negative effect on learning; however, the short learning times and the limited number of repetitions (only two) may have influenced his results. In a partial replication using pseudowords, Webb and Piasecki (2018) found that a third condition which allowed more time for repeated writing had a positive effect on both form knowledge and the learning of the form-meaning connection; the difference between the non-writing and writing with limited time conditions was not significant. Abbs et al. (2008) and Kaushanskaya and Yoo (2011) both focused on the learning of spoken word form. In a series of four experiments, Abbs et al. studied the possible effects of oral repetition on the learning of the spoken form – which could be expected given Thomas and Dieter's (1987) results for written form – but they found no advantage. Kaushanskaya and Yoo found that subvocal repetition was more effective than overt repetition for learning forms that used non-native phonemes on an immediate post-test, but not on a 1-week delayed post-test. Ellis and Beaton (1993) compared the effectiveness of the keyword method to that of repetition. Subjects studied 12 German words in three conditions: repetition, keyword link or own strategy (control group) and took receptive and productive tests. Repetition - the shallower strategy – yielded

superior results for immediate and delayed receptive tests. Productive tests yielded no significant differences, but students using the keyword method did outperform the control group. Van Hell and Candia-Mahn (1997) report on two experiments comparing the keyword method to subvocal repetition; in the first, experienced Dutch foreign language learners studied 60 Spanish words, in the second inexperienced US students studied 56 Dutch words following the same procedure. In both experiments, subvocal repetition yielded better results on immediate and delayed tests, although the differences between the groups of inexperienced FL learners were not significant.

A second topic in language learning investigations has been the difference between productive (L1-L2) and receptive (L2-L1) learning (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Steinel et al., 2007; Webb, 2005, 2009). Griffin and Harley investigate how two classes of 13-year-old English-speaking pupils learn 20 French words in an 8-minute session, manipulating the direction of learning and the direction of testing. They conclude that forward association (where direction of testing matches direction of learning) is stronger, but that direction of learning has no significant effect in se. Productive and receptive knowledge have a similar pattern of decay, with forgetting taking place mainly from the immediate test to the first delayed test one day later, and then flattening out. Overall, productive (E-F) learning is the better all-purpose direction. Mondria and Wiersma critically review five earlier experiments on word list learning (including Griffin and Harley, 1996). They conclude: (1) that receptive learning is easier and that students score higher on receptive tests; (2) that test results are higher if direction of learning is similar to direction of testing and that productive learning results in a considerable amount of receptive knowledge, while the reverse effect is less strong; (3) that productive knowledge is less resistant to decay. Their own experiment in which Dutch youths study 16 French words under three conditions (receptive, productive, receptive + productive) confirms these results, and also shows that a combination of receptive and productive learning does not lead to better results on either immediate or two-week delayed tests. In both of his studies, Webb (2005, 2009) investigated the learning of FL words by Japanese students using ten different tests in an attempt to isolate different types of productive and receptive knowledge. The first study reports on two experiments comparing receptive learning of an English word with an L1 gloss plus three example sentences with productive learning of the glossed English word by writing a sentence with it. The results were contradictory, with students learning receptively scoring higher on all ten tests in experiment 1, and students learning productively scoring higher on all ten tests in experiment 2. The main differences between the experiments were number of words (10 vs 20) and time on task. The receptive task took less time, which meant the receptive group had had more time to repeat in the first experiment. Webb (2009) used 10 nonsense words in two learning conditions – receptive L2-L1 order and productive L1-L2 order, and a within-subjects design. The results show that productive learning yielded significantly better results on productive tests and a receptive test of orthography, and that receptive tests yielded better results on receptive tests, but that these differences were not significant, which may be due to a ceiling effect on the receptive tests. The lowest gains (both receptively and productively) were attested to syntactic (patterning) knowledge. Steinel et al. (2007) focus on the learning of 20 English idioms by Dutch university students. They conclude that "direction of learning affected immediate performance, and its effect was particularly large on the productive test" (o.c. 468). The differences persisted on a three-week delayed test.

Explaining Effective Study Strategies

Several hypotheses have been proposed to account for different aspects of the learning effects described above. The lack of effect or the negative effect of the concurrent performance of different tasks shown in Abbs et al. (2008), Barcroft (2006), Kauskanskaya and Yoo (2011), Thomas and Dieter (1987), and Webb and Piasecki (2018), where studying a stimulus is combined with another activity like saying or writing the word, can be explained by Barcroft's (2015) TOPRA model, which suggests that such study situations under time restrictions lead to a depletion of processing resources, which in turn negatively affects word form learning. The superiority of repetition over deeper learning found in Ellis and Beaton (1993) and Van Hell and Candia-Mahn (1997) can be explained by the reliance on short-term phonological memory in learning foreign words (Papagno et al 1991). Mondria and Wiersma (2005), Steinel et al. (2007) and Webb (2005, 2009) have explained their results by referring to the transfer-appropriate learning hypothesis (Bransford et al., 1979; Morris et al., 1977; but see Rowland (2014) for a critique). Although these accounts do indeed refer to the fact that a specific focus in studying has a positive effect when subsequent tests have a similar focus, the actual hypothesis has a broader remit, suggesting that some types of studying may lead to transfer, i.e. the use of this knowledge in other contexts, while others only result in memory traces. Finally, the testing effect and the spacing effect have been explained alternatively by the elaborative retrieval hypothesis (Carpenter and Delosh, 2006) and the retrieval effort theory (Pyc and Rawson, 2009); both rely on the assumption that more difficult access to an item during learning creates stronger memory traces.

This Study

Aims

The present study was conducted within the framework of the participants' classroom work. It aims to investigate which study strategies are most effective with a view to long-term learning. It is characterized by minimal control in an attempt to elicit studying behaviour that is normal for the students involved. The topic, the nature of the study materials (thematically organized bilingual list and concordance materials) and the planning (four one-hour study sessions with pre- and post-tests) are the aspects that are controlled. The specific research questions are:

1. Are there significant short and long-term effects of continued learning on receptive and productive knowledge?
2. Which strategies or strategy combinations do experienced learners use when faced with a specific vocabulary studying assignment?
3. Which strategies or strategy combinations are most effective for receptive and productive learning within different time frames?

Method

Participants

Participants for this study were recruited amongst first-year students of English taking a degree in Applied Language Studies in Flanders. These students are native speakers of Dutch and have attained an upper-intermediate level in English at the start of higher education. After a pilot in 2016, the study was run in three consecutive years and 21 students (7, 10 and 4) volunteered in exchange for a 10 EUR book token. Data of 2 participants were incomplete, leaving 19 useable datasets. All participants signed an informed consent form.

The participants' score on a receptive Vocablab levels test (Peters et al., 2019) at the beginning of the academic year shows that most of them master the 5000 most frequent words. Four participants do not completely master the 5000 and 4000-word levels.

Materials

The study materials consisted of a thematic vocabulary of 90 English items (61 nouns, 30 of which are compounds, 5 adjectives, 16 verbs and 8 multi-word units) on the topic of work relations (see appendix A). The vocabulary was part of students' normal coursework, and the selection was topic-based. The words were presented in an organized list on two pages containing five thematic sublists (workers and management, pay, engaging and dismissing, unions and industrial action, working conditions). In the list, English items were presented in one column accompanied by necessary grammatical information (wordclass, countability, transitivity …) and a Dutch translation in a second column. In addition, students were given a second set of materials organized per thematic subset, where each list was followed by approx. 5 example sentences from the BNC per item.

Procedure

The study consisted of four stages: a three-hour classroom session with a pre-test, a one-hour study period and immediate post-tests; three one-hour study sessions at home spread over two weeks; announced post-tests; and announced delayed post-tests five weeks later. Table 1 provides an overview of the procedure.

Table 1
Procedure

Time	Day 1 – Pre-test	Day 1 – Study	Day 1 – Immediate post-test	study at home 2 wks	Day 2 – Post-test	5 wk interval	Day 3 – Delayed post-test
Activity	Productive Receptive (VKS)	1 session of 1 hour	Essay Translation 1 Productive Receptive	3 sessions of max. 1 hour	Essay Translation 2 Productive Receptive	no activity	Translation 1 Translation 2 Productive Receptive

Study Sessions and Logging

During the first session, students were briefed about the format and the aim of the study. They then signed the informed consent form. Following the pre-test, they were given one hour to study the vocabulary. The final goal set for studying was that they should be able to use the vocabulary receptively and productively; they were informed that they would be tested again at the end of the session. In order to keep track of their activity, each page of the study materials contained a header and a footer for them to note start- and finish-times for each time they worked with the page; they also had a couple of note pages which they could use as they saw fit. All materials were collected after the session; the researcher took additional notes on each student's activity as the session proceeded. Then followed a short break, after which students took the immediate post-tests.

For the three study at home sessions, students were instructed to space the sessions at least 3 days apart, and to study no more than one hour per session, using the materials and the time as they saw fit. They were told not to study on the day of the post-test. They had to complete a pre-formatted logpage for each session (see appendix B), providing information about date, vocabulary covered, and time spent on different activities and using different materials, and send this in. Starting year two, students were invited to audio-record themselves while studying and explain what they were going to do at each stage of the study session (instruction in appendix C) and send these files in – 12 students provided such additional data.

Test Formats and Scoring

Knowledge of the words was tested receptively and productively. The receptive pre-test (L2-L1) was inspired by the VKS-format (Paribakht & Wesche, 1997) – students indicated whether the L2 word was unfamiliar, familiar or whether they knew its meaning, and in that case, they had to provide a definition or a translation. The format had been trialled in previous studies (Pauwels 2012, 2018), and aimed to provide the participants with an option to indicate partial knowledge and to limit guessing. The productive pre-test (L1-L2) simply asked students to provide English equivalents with 90 Dutch words. For the post-tests, the receptive test was the same format as the productive test. In both tests, words were presented in alphabetical order, i.e., a different order from each other and from the study materials to counter list effects. Tests were scored out of 90, with a fully correct answer scored as 1. Partial knowledge of form (like spelling errors or wrong prepositions in multiword units) and partial knowledge of meaning (vagueness) was credited with 0.5. On the receptive pre-test, only demonstrated knowledge was credited.

Apart from the list-based tests, translation tasks (Dutch to English) were used as additional post-tests in order to simulate the course target of (cued) productive use. We used two different tests, each targeting 22 items, in order to be able to test more items. Correctly used items were scored as 1, with partial credit 0.5 for usage errors (fixed prepositions, count/uncount, collocation …).

During the immediate post-test and the post-test, after the at-home study, students also wrote a short 250-word essay simulating spontaneous productive use. Two different titles were used: 'the pros and cons of trade unions' and 'the pros and cons of equal pay for everyone'. Data from the essays will not be reported at this stage.

The order of the subtests was: essay, translation task, productive post-test, receptive post-test. To avoid test fatigue, it was decided to limit the test battery to four tests at each stage and to vary the translation tests and the essays.

Students were instructed not to work with the materials in the period between the post-test and the delayed post-test; it was explained to them that this would be informative about the long-term effect of their study effort, ad that this would help them plan their coursework. At the end of the final session, students received their 10EUR book token. After all, tests were corrected, students received feedback via e-mail in the form of their test results with additional explanation.

Results

Overview

Scores on all tests (see table 2 below) are normally distributed (Skewness and Kurtosis +2><-2) except for the immediate productive post-test and the productive post-test (Kurtosis 2.77 and 2.92). A repeated measures ANOVA on the receptive tests demonstrates significant differences. The receptive results violate the sphericity assumption, hence Wilks'Lambda = .036, $F (1,18) = 143.42$, sig. $< .001$, partial $eta^2 = .964$; Bonferroni pairwise comparisons are all significant at $<.001$, except for the difference between the immediate and delayed post-test which is significant at $<.013$.

A Friedman test for not normally distributed results on the productive tests also reaches significance (3, N 19 = 48,284 p $<.001$). Wilcoxon pairwise comparisons are significant for all pairs, with large effect sizes, except for the difference between immediate and delayed post-test (pre-test vs immediate post-test $Z = -3.824$ sig. $< .001$; effect size .877; immediate post-test vs post-test $Z = -3.141$ sig. $< .005$ effect size .721; post-test vs delayed post-test $Z = -3.764$ sig. $< .001$ effect size .864).

Table 2
Results – M (SD)

	Pre-test	Immediate Post-test	Post-test	Delayed Post-test
Receptive (/90)	41.5 (10.47)	76.47 (8.31)	84.97 (4.60)	80.08 (6.45)
Productive (/90)	15.55 (5.64)	64.76 (10.12)	77.32 (9.27)	62.05 (9.68)
T1 (/22)		16.11 (2.74)		15.58 (3.10)
T2 (/22)			17.84 (1.89)	14.89 (2.87)

Results of One Hour of Study in Class

On average, students demonstrated knowledge of 41.5/90 words on the receptive pre-test, with quite a lot of variation (N = 19, SD = 10.47). Productive scores were much lower with an average of 15.55/90. All students overestimated their receptive knowledge and often provided wrong explanations instead of opting for 'I recognize but do not know the precise meaning'; overestimates ranged between 8 and 19 items and especially low scorers tended to overestimate.

During the one-hour study period, all students went through the materials repeatedly reading and rereading and practised retrieval (both receptive and productive), with the exception of one student who looked further ahead and started making flashcards. Only nine students spent time (25 minutes on average) reading the example sentences. Other activities that were observed were reorganizing the materials (five students), paying selective attention to some words which they found more difficult (fourteen students) and copying (eight students), most of which involved writing down the words. The strategies were used as categorial variables (present or absent) for regression analysis, except for retrieval where the number of retrievals was used.

From the average scores on the immediate post-tests – receptive 76.47 (N = 19, SD = 8.31) and productive 64.76 (N = 19, SD = 10.12) – it is clear that most students studied fairly effectively. In the receptive test, there may have been a slight ceiling effect for 4 students scoring in excess of 84.5/90. Only two students scored more than one SD below the average on both receptive (resp. 56.5 and 59) and productive (resp. 39.5 and 38.5) tests. One of these was the student focusing on making flashcards, the other spent most of his time on rote copying the first half of the vocabulary.

A stepwise regression for receptive gains with previous knowledge (operationalized as their combined score on the 4000 and 5000 word levels of the Vocablab test because this differentiated them most strongly), their score on the pre-test and the strategies identified as independent variables resulted in a model (adjusted R^2 = .509, p = *.001*) with two explanatory parameters: pre-test result as a negative factor (beta = -.929, p < *.001*) and selective attention as a positive factor (beta = .466, p = *.038*). For productive gains the strongest model (adjusted R^2 = .480, p = *.005*) had three explanatory variables: pre-test result as a negative predictor, (beta = -.701, p = *.002*) and partial knowledge calculated as the difference between receptive and productive pre-test scores (beta =.336, p = *.146*) and selective attention (beta = .380, p = *.103*) as positive predictors. A two-factor model (adjusted R^2 = .436, p = *.004*) with pre-test result (beta = -.650, p = *.003*) and selective attention (beta = .569, p = *.008*) was slightly less strong.

The results on the first translation test (T1) correlate with the results on the productive test (r (17) = .831, p < *.001*).

Out of Class Studying

Most students followed the instructions quite well, spaced their study sessions and did three sessions of approximately one hour (between 140 and 180 minutes). Six students planned their sessions at shorter intervals (less than three days in between) closer to the announced post-test; one student planned only two sessions due to illness and studied for 120 minutes. Six students spent considerably less time overall (between 60 and 99 minutes).

Only seven students spent time studying the example sentences and studying mostly consisted of (re)reading the vocabulary lists and retrieval activities. For retrieval, five students made use of paper or online flashcards, and most students either said words aloud or wrote them down during retrieval. Eleven students consistently used oral rehearsal as a study method, and six students used written rehearsal. Five students used more elaborate strategies like using the words in a sentence (writing) or making exercises from their vocabulary practice book. The following strategies were coded as categorial variables for regression analysis: keeping sessions 3 days apart, retrieval before reading, using

flashcards, copying words, saying aloud, reading examples, using more elaborate strategies. Time spent studying was entered as a continuous variable.

Post-test results

The scores on the receptive post-test of 84.97 (N = 19, SD = 4.60) were positively skewed, and there was a clear ceiling effect. Scores on the productive post-test averaged 77.32 (N = 19, SD = 9.27) and showed a wider variation. A regression analysis on the receptive gains from immediate post-test to post-test results in a two factor model (adjusted R^2 = .752, $p < .001$) with immediate post-test result as a major negative predictor (beta = -.953, $p < .001$) and spacing as a possible positive predictor (beta = .206, $p = .133$). A regression analysis on the productive gains yields a three factor model (adjusted R^2 = .723, $p < .001$) with immediate post-test result as strongest negative predictor (beta = - .829, $p < .001$), copying words as a second negative predictor (beta = - .315, $p = .025$) and spacing as a positive predictor (beta = .357, $p = .012$).

The results on the translation post-test (T2) correlate with the results on the productive post-test (r 17) = .593, $p = .007$), but not with the results on the immediate translation test T1.

Results on the delayed post-test taken 5 weeks later show small losses for receptive use with an average score of 80.08 (N = 19, SD = 6.45) and clear losses for productive use with an average score of 62.05 (N = 19, SD = 9.68). While the receptive final scores are situated midway between the scores on the immediate post-test and the post-test, productive scores are on average slightly below the immediate post-test scores. Individual productive scores diverge strongly from immediate post-test scores (range -20; +17,5) which implies a clear effect of study strategy.

A regression analysis on losses in receptive scores compared to the post-test (adjusted R^2 = .783, $p < .001$) with the parameters from the study period and Vocablab score as predictors yields two negative correlations: with Vocablab score (beta = - .920, $p < .001$) and with the technique of retrieval before rereading (beta = -.254, $p = .044$). A further regression on changes in receptive scores compared to the immediate post-test is inconclusive. This is probably due to the ceiling effects for the better students in both earlier tests.

A regression analysis on productive losses compared to the post-test yields a three factor model (adjusted R^2 = .310, $p = .036$) with negative predictor spacing (beta = -.644, $p = .016$) and positive predictors post-test result (beta = .617, $p = .017$) and time spent studying (beta = .649, $p = .017$). A further regression analysis comparing the productive changes from immediate post-test to delayed post-test yields a three factor model (adjusted R^2 = .324, $p = .031$) with positive predictor spacing (beta = .440, $p = .079$) and negative predictors gain on the immediate post-test compared to the pre-test (beta = -.742, $p = .006$) and time spent studying (beta = -482, $p = .076$). It should be noted that two out of three predictors only approach significance.

The results on the delayed translation tests correlate strongly with the delayed productive test (T1: r (17) = .746, $p < .001$; T2: r (17) = .694, p = .001) and with each other (r (17) = .647, $p = .002$). There is also a strong correlation between the results of T2 post-test and T2 delayed post-test (r (17) = .486, $p = .035$).

Six students from the 2018 class were found prepared to take the receptive and productive tests again nine months after the conclusion of the experiment. Meanwhile, they

had restudied the words once for their final vocabulary exam four months previously. They now averaged 76.42 (N = 6, SD = 4.78) on the receptive test, which represents a loss of 5.83 compared to the delayed post-test, and 43.5 (N = 6; SD = 7,54) on the productive test, which represents a loss of 17.5 compared to the delayed post-test. Those who had scored best on the pre-test were also the top scorers on this test.

Discussion

     In response to research question (1), it is clear that there are significant short and long-term effects of continued study on short and long-term learning, but that these are different for receptive and productive knowledge. Productive knowledge seems to be acquired more slowly and be more susceptible to forgetting. This confirms the findings from previous research (reported in Mondria & Wiersma, 2004). The translation tests further demonstrate that full productive knowledge needed for correct use may even take longer: there are items (e.g. before tax) that most students can produce in response to a cue from the immediate post-test onwards, but that they almost never use correctly (in the example: following the head noun).

     In response to research question (2), the first thing that stands out is the variety of strategies and strategy combinations used by the learners. Still, students appeared to consider the task primarily as one of memorization: rereading, saying aloud and copying are popular strategies, while less than half the students made use of the example sentences, and only five students used more elaborate strategies like using words in sentences or making exercises. This was not only the case during the one-hour study period but also during three hours of at-home study. Secondly, all students practice retrieval in some way. While across-session spacing was built into the experimental setup, within session spacing was not, and there we see that there are differences as to the difficulty of the retrieval students subject themselves to; some students start a new study session with an attempt at retrieving all the words, while others always start by rereading and only try retrieving afterwards. There are also differences within sessions, as some students study the vocabulary per subsection, while others study the whole list in one go. This means that some students will have been misled in judging their knowledge by the fluent retrieval of materials they had read just previously (Bjork et al., 2013), while others built in more desirable difficulty (Pyc & Rawson, 2009). Several students paid selective attention to words they found more difficult by either marking them or writing a 'reduced list' and then restudying. As we will see below dropping easy items and selective attention for difficult items did not result in a disadvantage, contrary to Kornell and Bjork's (2009) findings, but this may have been an effect of the fact that students mainly dropped words they already knew before (as demonstrated in the pre-test).

For the investigation of the effect of the different study strategies (RQ 3), it was decided to include the result on the previous test in the regression analyses, since it was felt this result might affect studying behaviour. In fact, the previous result shows up as a negative predictor in both the regression analyses on the gains in the immediate post-test and in the regression analyses on the gains in the second post-test. Where the receptive post-tests are concerned, we should consider the possibility that this is due to ceiling effects for the better learners: they had much less room for improvement. There is a similar effect for the productive post-tests, where there was still room for improvement. This suggests that learning from lists may have limitations. The presence of near synonyms in the sets of

words may have played a role here (cf Tinkham, 1993); conceptual difficulty may be another factor – there were signs that the students lacked topical knowledge, and sometimes did not even know the precise meaning of the Dutch items. In some cases, the lack of progress after the immediate post-test could also be caused by the fact that students were misled by the apparent ease with which they learnt the words during the initial one-hour study session (cf Bjork et al., 2013).

Two other factors played a role in the initial one-hour study session. Selective attention for specific items had a positive impact on both receptive and productive gains. Partial (i.e. receptive) knowledge demonstrated in the pre-test had a positive impact on initial productive learning by reducing the learning burden of the items involved. The regression analyses on the post-test gains confirm the importance of between session spacing in line with the findings of memory research (Cepeda et al., 2006). The negative effect of copying words on productive knowledge poses a problem: Barcroft's (2015) TOPRA-model is not relevant here, since the act of copying words did not preclude students paying sufficient attention to the word-meaning link. Moreover, it contradicts the findings of Thomas and Dieter (1987) and Webb and Piasecki (2018). One possible explanation is that by choosing to write words down, students lost time which they could have used to increase the number of repetitions or retrievals which would have been more effective.

The regression analyses on delayed post-test losses give indications confirming the importance of between sessions spacing (productive test) and within-session spacing (receptive test). Time spent studying shows up as a negative factor for productive learning. It is not immediately clear why this would be so, but the lack of effect of time spent studying has also been found in Pauwels (2012, 2018). It may be that the better learners simply need less time to do the job.

Overall vocabulary level prior to study (as measured by the score on the 4000 and 5000 levels of the Vocablab test) is a negative predictor for losses in the delayed receptive posttest. It could be hypothesized that students with a larger vocabulary are better able to integrate new items in their vocabulary network, which helps them remember. Finally, we should note that we were not able to find any significant effects for some of the strategies identified: oral repetition/saying aloud, using flashcards, reading examples, and using more elaborate strategies. The lack of an effect for oral repetition is in line with results from previous research (Abbs et al., 2008, Kaushanskaya & Yoo, 2011), and it should also be clear that using flashcards in itself is not necessarily beneficial because students may drop words too early (Kornell & Bjork, 2009). The lack of any effect of 'making use of the examples provided' or of the use of 'more elaborate strategies' is probably due to the vagueness of these categories and as such a methodological issue.

Conclusions

Methodological issues and limitations

In this study we attempted to maximize ecological validity – the only aspects that were controlled were the overall study schedule (four spaced study sessions over a 3-week period) and the maximum time spent (four hours). Studying strategies were not controlled but logged by the researcher or by the students themselves. Still, tracking student activity during study remains a difficult issue. During the in-class study sessions, it was possible to

detect retrieval behaviour when students covered up a column, or when they started putting checkmarks next to the words they retrieved successfully, but some students may have engaged in retrieval without such outward signals. The logpages gave a fairly good insight into the activities the students performed and the time they spent on them, but it was not always clear if activities (reading and writing, for example) took place concurrently. The recordings did provide more detailed information in that respect. They also showed that students used strategies flexibly. Think aloud protocols as used by Lawson and Hogben (1996) might have yielded more information but could only have been used for the individual study sessions and could have influenced performance – they would certainly have impacted on the time actually spent studying.

The coding of the different study parameters as categorial may hide relevant differences in study behaviour. Some students may have made much more, or much more systematic, use of saying words aloud than others. Some students tried to retrieve words at the start of each study session, others did so only once. Writing sometimes involved just copying, sometimes it was done in combination with retrieval. However, the limited number of participants means that a more fine-grained coding would have resulted in unique profiles that would not have been amenable to statistical analysis.
The small sample size is another issue with only 19 participants out of a possible 200 over the three years. Still, in spite of the time investment required, there was little drop-out (only two participants), and it can be argued that the quality of the data will have benefited from using motivated volunteers.

Pedagogical Implications

On the whole, students seemed to prefer so-called shallow strategies and focused on list learning. They looked at the examples but seemed to use these mainly to get a better grip on the meaning of the items. Usage information did not seem to be a major concern, in spite of their experience in struggling with the translation passage in the immediate post-test. Still, it is also clear that such shallow strategies are effective to a certain extent, especially for receptive learning of words.

Most students did seem to be aware of the positive effects of retrieval since they devoted a lot of their study time to it. However, many students tended to practise retrieval immediately after a reading session, minimizing the difficulty of retrieval, and creating for themselves a misleading impression of fluency or mastery. Bjork et al.'s (2013:438) conclusion that "assessing whether learning has been achieved is difficult because conditions that enhance performance during learning can fail to support long-term retention" is something students, and teachers, should be made aware of.

Future Prospects

In a follow-up study, we aim to investigate the data at item level comparing receptive knowledge, productive knowledge and use, which will allow us to further explore parameters of item difficulty.

Correspondence should be addressed to Paul Pauwels.
Email: paul.pauwels@kuleuven.be

References

Abbs, B., Gupta, P., & Khetarpal, N. (2008). Is overt repetition critical to expressive word learning? The role of overt repetition in word learning with and without semantics. *Applied Psycholinguistics, 29*, 627-667.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review, 105*(1), 158-173. doi:10.1037/0033-295X.105.1.158

Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research, 22* (4), 487–497.

Barcroft, J. (2009). Strategies and performance in intentional L2 vocabulary learning. *Language Awareness, 18* (1), 74–89.

Barcroft, J. (2015). *Lexical Input Processing and Vocabulary Learning.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bjork, R.A., Dunlosky, J. & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques and Illusions. *Annual Review of Psychology, 64,* 417–444.

Bransford, J.D., Franks, J.J., Morris, C.D. & Stein, B.S. (1978). Some general constraints on learning and memory research. In L.S.Cermak and P.I.M.Craik (eds.), *Levels of processing and human memory*. Hillsdale, N.J.: Lawrence Erlbaum Publishers.

Carpenter, S.K. & Delosh, E.L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268-276. doi:10.3758/BF03193405

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T. & Rohrer, D. (2006). Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis. *Psychological Bulletin, 132*(3) 354-380. doi:10.1037/0033-2909.132.3.354

Elgort, I. (2011). Deliberate Learning and Vocabulary Acquisition in a Second Language. *Language Learning 61 (2),* 367-413. DOI: 10.1111/j.1467-9922.2010.00613x

Ellis, N. & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A, 46* (3), 533-558. doi:10.1080/14640749308401062

Fan, M.Y. (2003). Frequency of Use, Perceived Usefulness, and Actual Usefulness of Second Language Vocabulary Strategies: A Study of Hong Kong Learners. *The Modern Language Journal, 87* (2), 222–241.

Griffin, G. & Harley, T.A. (1996). Lists learning of second language vocabulary. *Applied Psycholinguistics, 17*, 443-460.

Gu, P.Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46*(4), 643–679.

Kang, S.H.K, Gollan, T.H. & Pashler, H. (2013). Don't just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychon Bull Rev, 20,* 1259-1265. doi: 10.3758/s13423-013-0450-z

Karpicke, J.D. & Roediger, H.L. III (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968. doi:10.1126/science.1152408

Karpicke J.D., Butler, A.C. & Roediger, H.L. III (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17*(4), 471-479. DOI:10.1080/09658210802647009

Kaushanskaya, M. & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes, 26*(1), 121-148. DOI:10.1080/01690965.2010.486579

Kornell, N. & Bjork, R.A. (2009). Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Memory, 16*(2), 125-136. DOI: 10.1080/09658210701763899

Laufer, B. & Shmueli, K. (1997). Memorizing New Words: Does Teaching Have Anything to Do with It? *RELC Journal, 28*, 89-108. DOI: 10.147/003308829702800106

Lawson, M. J. & Hogben, D. (1996). The vocabulary learning strategies of foreign-language students. *Language Learning, 46*, 101–135.

Mondria, J.A. (2003). The effects of inferring, verifying and memorizing on the retention of L2 word meanings. *SSLA, 25*, 473-499.

Mondria, J. & Wiersma, B. (2004). Receptive, Productive, and Receptive + Productive L2 Vocabulary Learning: What Difference Does it Make? In P. Bogaards and B. Laufer (eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing.* Amsterdam/Philadelphia: John Benjamins Publishing Company, 79-100.

Morris, C.D., Bransford, J.D. & Franks, J.J. (1977). Levels of Processing Versus Transfer Appropriate Processing. *Journal of Verbal Learning and Verbal Behaviour, 16*, 519-533.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition, 37*, 677-711.

Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning. *IRAL, 54*(3): 257–289. DOI 10.1515/iral-2015-0022

Nation, I.S.P. (2013). *Learning Vocabulary in Another Language*. Cambridge: C.U.P.

Papagno, C., Valentine, T. & Baddeley, A. (1991). Phonological Short Term Memory and Foreign-Language Vocabulary Learning. *Journal of Memory and Language 30,* 331-347.

Paribakht, T.S. & Wesche, M. (1997). Vocabulary Enhancement Activities and Reading for Meaning in Second Language Acquisition. In J.Coady and T.Huckin. *Second Language Acquisition: A Rationale for Pedagogy*. Cambridge: C.U.P., 174-200.

Pauwels, P. (2012). Vocabulary materials and study strategies at advanced level. *The Language Learning Journal 40(1)*: 47-63.

Pauwels, P. (2018). How advanced learners approach intentional vocabulary study. *The Language Learning Journal 46(3)*: 293-310. DOI 10.1080/09571736.2015.1078398

Peters, E., Velghe, T. & Van Rompaey, T. (2019). The development of an English and French vocabulary test. *ITL - International Journal of Applied Linguistics, 170*(1)*, 53 - 78 DOI: 10.1075/itl.17029.pet

Pyc, M.A. & Rawson, K.A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35*(8), 1917-1927. doi:10.3758/BF03192925

Pyc, M.A. & Rawson, K.A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437–447. doi:10.1016/j.jml.2009.01.004

Pyc, M.A. & Rawson, K.A. (2012). Why Is Test–Restudy Practice Beneficial for Memory? An Evaluation of the Mediator Shift Hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 737–746 DOI: 10.1037/a0026166

Roediger, H.L. III & Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20-27.

Rowland, C.A. (2014). The Effect of Testing Versus Restudy on Retention: A Meta-Analytic Review of the Testing Effect. *Psychological Bulletin, 140 (6)*, 1432-1463.

Sanaoui, R. (1995). Adult Learners' Approaches to Learning Vocabulary in Second Languages. *The Modern Language Journal, 79*(1), 15-28.

Schmitt, N. (1997). Vocabulary Learning Strategies. In: Schmitt, N. and M.McCarthy. *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: C.U.P., 199-227.

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research, 19*(1), 28–42 DOI: 10.1177/1362168814541726

Schuetze, U., & Weimer-Stuckmann, G. (2011). Retention in SLA Processing. *CALICO Journal*, *28*, 460–472.

Soderstrom, N.C., Kerr, T.K. & Bjork, R.A. (2016). The Critical Importance of Retrieval—and Spacing—for Learning. *Psychological Science, 27*(2), 223–230. DOI: 10.1177/0956797615617778

Steinel M.P., Hulstijn, J.H. & Steinel, W. (2007). Effects of Direction of Learning, Direction of Testing, Idiom Imageability, and Idiom Transparency. *SSLA*, 29, 449–484. DOI: 10+10170S0272263107070271

Thomas, M.H. & Dieter, J.N. (1987). The Positive Effects of Writing Practice on the Integration of Foreign Words in Memory. *Journal of Educational Psychology, 79*(3), 249-253.

Thorndike, E.L. (1908). Memory for Paired Associates. *Psychological Review 15,* 122-138.

Tinkham, T. 1993. The Effect of Semantic Clustering on the Learning of Second Language Vocabulary. *System, 21*(3), 371-380.

Van Hell, J.G. & Candia-Mahn, A.C. (1997). Keyword Mnemonics Versus Rote Rehearsal: Learning Concrete and Abstract Foreign Words by Experienced and Inexperienced Learners. *Language Learning 47*(3), 507–546.

Webb, W.B. (1962). The Effects of Prolonged Learning on Learning. *Journal of Verbal Learning and Verbal Behaviour 1*, 173-182.

Webb, S. (2005). Receptive and Productive Vocabulary Learning: The Effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition 27*, 33-52.

Webb, S. (2009). The Effects of Receptive and Productive Learning of Word Pairs on Vocabulary Knowledge. *RELC*, *40*(3), 360-376. DOI: 10.1177/0033688209343854

Webb, S. & Piasecki, A. (2018). Re-examining the effects of writing on vocabulary learning. *ITL - International Journal of Applied Linguistics, 169*(1), 72-94. DOI: 10.1075/itl.00007.web

Appendix A
The Vocabulary Assignment

Work Relations

Coding system

C/U: countable vs uncountable
pl/sing/GP: plural vs singular vs collective noun
attrib: adjective only used before a noun
T/I: transitive (passivizable) vs intransitive
Adv: adverb
Infml: informal
Euph: euphemism

Workers and management

the workforce[1] , the labour force[1] [C, usu. sing]	het personeelsbestand
the workforce[2], the labour force[2] [C, usu. sing]	de actieve bevolking
manpower [U]	werkkrachten, arbeidskrachten
the staff (C, usu. sing, GP), the personnel	het personeel, de werknemers
a member of staff, an employee	personeelslid, werknemer
blue-collar [attrib]; blue-collar worker	arbeiders-; arbeider
white-collar [attrib]; white collar worker	bedienden-; bediende
Labourer, (farm)hand	Handarbeider, landarbeider
workman	werkman, arbeider
skilled [attrib]	geschoold
menial [attrib]	ongeschoold
shopfloor [attrib]; (on) the shopfloor	arbeiders-; (op) de werkvloer
a (sales) rep(resentative)	vertegenwoordiger
a PA (personal assistant), private secretary	Privésecretaris/esse
manager (sales-, marketing-, …)	kaderlid, manager
an executive	directielid, directeur
a managing director, CEO	algemeen directeur, afgevaardigd bestuurder
board (of directors)	raad van bestuur
chairman/-woman/-person of the board	voorzitter
promote sb (sth)[T]	bevorderen
demote sb [T]	degraderen

7.2. Pay

fee [usu pl]	ereloon, honorarium
wages [usu pl], pay [U]	(week)loon
salary [C], pay [U]	salaris
retainer	voorschot (op een honorarium)
remuneration	verloning
pay slip	loonbriefje
a rise, a raise	loonsverhoging
wage reduction, wage cut	loonsvermindering
allowance	onkostenvergoeding
fringe benefit; perk [infml]	extralegaal voordeel; extraatje (infml)
before tax [postmodifier]	bruto
net	netto

Engaging and dismissing

job	job
post, position	(vaste) betrekking, functie
vacancy, opening	vacature
to employ sb [T]	tewerkstellen
engage sb, hire sb, take sb on [all T]	aanwerven
to assign sb to [usu.passive]	aanstellen als, tewerkstellen in, een opdracht geven
to dismiss sb [T], fire sb [T] =	ontslaan =
sack sb [T], give/get the sack [both infml BrE]	aan de deur zetten, afdanken /afgedankt worden
to resign [I, T] [fml], leave [I,T]; give up one's job [T]	ontslag nemen =
quit [I,T]] [infml]	opstappen
to make sb redundant [T, usu passive] [euph]	laten afvloeien
lay sb off	op (tijdelijke) werkloosheid plaatsen
redundancy [C, usu pl]	ontslagen, afvloeiingen
unemployment [U]	werkloosheid
the unemployed [pl]	werklozen
employment [U]	tewerkstelling
employer	werkgever
unemployment benefit [U]; the dole [BrE]	stempelgeld
be on unemployment benefit; on the dole [BrE]	uitkeringsgerechtigd werkloos zijn

Unions and Industrial Action

a (trade) union	vakbond
a strike [C]	staking
to strike (for/against)[I]	staken
industrial action [U]	vakbondsactie
stoppage	tijdelijke werkonderbreking
wildcat strike	spontane staking
collective bargaining	collectieve arbeidsonderhandelingen, CAO-overleg
shop steward	vakbondsafgevaardigde

Working conditions

to work nine to five	een kantoorjob hebben
to work shifts	in ploegen werken
shift work [U]	ploegenarbeid
the night shift	nachtploeg
part-time [attrib, Adv]	deeltijds
full-time [attrib, Adv]	voltijds
to work overtime	overuren maken
temporary employment/work	interim/uitzendwerk
permanent employment/work	vast werk
steady job [infml]	vast werk
a temp	uitzendkracht
to temp [I]	uitzendwerk doen

Appendix B
Logpage

Student logpage (2017)
(send in a logpage immediately after each of the study periods)

NAME:
DATE:

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31									

MONTH:

1	2	3	4	5	6	7	8	9	10	11	12

TIME SPENT STUDYING TODAY: IN THE RELEVANT CELLS
1. Indicate the order in which you performed the activities (rank them 1, 2, 3 etc.)
2. Write down the number of minutes after each activity
3. If you repeat an activity, add it in the same box after its new rank

e.g.: 1/10m; 4/2m in the first box means that the first activity you performed was reading the complete list with translation equivalents for ten minutes, and the fourth activity was (re)reading it for two minutes

Activity performed ↓	Materials used →	T-equivalents only	T-equivalents with examples	Examples only	Own materials (please describe)
Read					
Read aloud					
Check ability to translate into English					
Check ability to translate into Dutch					
Copy words					
Write examples					
Other (please describe below)					

Comments:

Appendix C
Talk Aloud Instruction

Talk aloud instruction
During your three one-hour sessions, I would like you to record an audio description of what you are doing.

1.   At the start of each session, state the date and the time. Leave the recording device to run during the complete session. If you take a break, or are interrupted, you can pause the recording. When you restart, state the time.

2.   During the session:
-       At the start of **each** new activity, describe what you are going to do in some detail, and refer to the materials you are going to use (e.g. I am going to read through the complete list of words; OR I am going use the complete list to cover the left-hand column and I am going to try and recall each word, checking as I go along; OR I am going to read the example sentences with the first set of words and try to recall what each word means OR I am going to reread my list of most difficult words …);
-       When you turn a page, say so;
-       During the activity, say when you do something that is different (e.g. when you are reading, you suddenly decide to write down one specific item OR you are reading the list, and suddenly decide to look up an example OR…);
-       When you end an activity, say so;
-       When you take a short break, say so;

3.   After you have completed your study session, send me the audio file as a mail attachment.
Thank you for participating.