Applied Psycholinguistics **25** (2004), 175–200
Printed in the United States of America
DOI: 10.1017.S0142716404001092

The effect of enhanced lexical retrieval on second language writing: A classroom experiment

PATRICK SNELLINGS and AMOS VAN GELDEREN
University of Amsterdam

KEES DE GLOPPER
University of Groningen

ADDRESS FOR CORRESPONDENCE
Patrick Snellings, SCO-Kohnstamm Institute for Educational Research, University of Amsterdam, Wibautstraat 4, P.O. Box 94208, 1090GE Amsterdam. E-mail: p.snellings@uva.nl

ABSTRACT
Lexical retrieval is an essential subprocess in language production, and its efficiency is crucial for writing. To improve writing quality in a second language, we developed an experimental, computerized training for improving fluency of lexical retrieval in a classroom setting, applying techniques previously restricted to laboratory use. In a counterbalanced design with randomized assignment, each of two groups was trained on a different set of words for productive use. A previous study showed that students in both groups attained greater fluency of lexical retrieval on the trained word set in comparison to students who were not trained on that word set. The current study provides evidence that this enhanced fluency transferred to narrative writing, as students in both groups used the trained words more often in narrative texts. In addition, one of the groups showed significant improvement in their expression of essential content elements. However, no significant differences were found on global quality ratings. The results are discussed in the context of theories of limited processing capacity during text production. We also consider the implications of these results for language instruction.

Lexical retrieval is an essential process in both oral and written language production and an important aspect of fluency. It involves the selection of lexical concepts, resulting in lemmas representing a word's syntax that are subsequently morphologically, phonologically, and phonetically encoded to be either articulated (Levelt, 1989; Levelt, Roelofs, & Meyer, 1998) or written down. In contrast to the closely related process of lexical access ("word recognition"), where forms have to be recognized in order to arrive at the correct lexical concept, in lexical retrieval concepts are transformed into linguistic forms. Although the order in which the components of the process are active is reversed in lexical access and lexical retrieval, most components are interchangeable (Kroll & Stewart, 1994; Levelt, Roelofs, & Meyer, 1999).

In their influential model of the writing process, Hayes and Flower (1980) distinguished lexical retrieval ("translating") from other processes involved in

© 2004 Cambridge University Press 0142-7164/04 $12.00

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

text production, such as generating ideas, planning, organizing, and reviewing. Kellogg (1996) also stressed the importance of lexical retrieval in his discussion of the processes involved in written text production. In an experimental study, Kellogg (1994) showed that lexical retrieval is highly effortful, even for native speakers. Although planning and reviewing placed the greatest demands on cognitive resources, as measured by reaction time (RT) interference on a secondary task (an auditory probe), it turned out that translating content into sentences, which involves lexical retrieval and syntactic processing, was also highly effortful and slowed reaction times about 350 ms above baseline times. As cognitive resources are limited, it is possible that devoting too much attention to lexical retrieval could have adverse effects on the writing *process* (for a discussion of attention limits, see Kellogg, 1999; Lea & Levy, 1999). More attention being devoted to lexical retrieval may result in less attention available to planning, reviewing, and other cognitive operations necessary for writing (Hayes, 1996; Kellogg, 1996).

Within reading research, there is a well-established school of thought that cognitive resources devoted to one subprocess preclude attention being given to other important subprocesses, and that this may also be reflected in the quality of the process as a whole. Evidence has also been found for the influence of high resource demands on text comprehension (Favreau & Segalowitz, 1983; Just & Carpenter, 1992; Koda, 1996; Perfetti, 1985).

Within writing research, some studies have shown a relationship between resource demands and writing by experimentally increasing the cognitive load in writing tasks (for a review, see Alamargot & Chanquoy, 2001). Penningroth and Rosenberg (1995) showed that narrative writing with a higher information-processing load negatively affected ratings of coherence, but not of holistic quality. Lea and Levy (1999) demonstrated that a secondary task (i.e., detecting and signaling the visual alternation of letters and digits) negatively affected both writing fluency and writing quality. Scardamalia, Bereiter, and Goelman (1982), on the other hand, only found a relationship between resource demands and text length but not between resource demands and text quality. They showed that for Grades 4 and 6 children, neither slow rate of output nor interference from mechanical demands (e.g., spelling and punctuation) led to a decline in text quality. Nevertheless, in another experiment (cf. Scardamalia et al., 1982) they found that when children were forced to retain information in short-term memory, there was sometimes information loss between forecasts (i.e., the words they intended to say) and actual writing. This finding implies that a slow writing rate may be a disadvantage in that children may forget some of the words they are going to say. Finally, in a study focusing on lexical retrieval, Brown, McDonald, Brown, and Carr (1988) showed that when lexical material had to be retrieved from memory this process usurped attention from execution processes, resulting in reduced legibility and a greater number of errors.

Some correlational studies have also served to emphasize the importance of lexical retrieval for writing. McCutchen, Covill, Hoyne, and Mildes (1994) demonstrated that more proficient first language (L1) writers (across Grades 3, 4, 7, and 8) showed better performance on a reading and speaking span task than their less proficient counterparts. The more proficient writers had larger spans (i.e., they

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

remembered more last words or were able to generate grammatical sentences for longer lists) and produced longer sentences. Ransdell and Levy (1996) found a similar relationship between a writing span measure and writing. In another experiment in their 1994 study, McCutchen et al. also found that the more proficient writers had shorter latencies and higher accuracy rates on lexical decision tasks. They followed Flower and Hayes (1980) in arguing that writers have to handle a number of constraints simultaneously. The possible strain that this constraint juggling places on attentional resources could mean that the more attention that is devoted to translating and lexical retrieval, the less that can be devoted to other processes such as planning, generating ideas, organizing, and reviewing (for a review, see also McCutchen, 1996). In our view, this would seem to imply that once writers can retrieve words fluently, they can devote attention to other processes, enabling them to write better texts.

Flower and Hayes (1980) suggested the possibility of providing training to increase the efficiency of sentence production processes in the L1, for example, through sentence-combining exercises, the ultimate aim of which would be to enable writers to concentrate more on other important writing processes. To date, no studies have been carried out that specifically investigate the effect of training sentence production processes on writing quality. However, a small number of studies have shown that reducing resource demands may *facilitate* writing. Although research in this direction has been scarce, this approach is promising because it may have important practical implications for education. Van Gelderen (1997) showed that error detection and revision activities at the macrolevel of the text could be considerably improved by facilitating the revision process of L1 writers (Grades 5–6) by such means as providing standard phrases for evaluation and reformulation and providing opportunities for improving someone else's text. Recently, Chenoweth and Hayes (2001) have shown in a small cross-sectional think-aloud protocol study that second language (L2) courses resulted in direct retrieval of the words needed in their texts and there was a (small) increase in the length of the sentence parts ("bursts") that were proposed and evaluated in constructing sentences. In their view, linguistic experience led to increased lexical retrieval skills and increased facility with complex grammar, reducing resource demands and enabling writers to handle more complex and longer language structures. Finally, Glynn, Britton, Muth, and Dugan (1982) showed that in L1 persuasive writing, both preliminary drafts and final drafts improved when students (in a graduate course) were instructed not to pay attention to logical sequence or to structure demands (i.e., correct syntax and spelling) in their preliminary drafts. In particular, the students were able to use more arguments that argued for a certain position. Interestingly, a second experiment, involving undergraduates of average and low verbal ability (as measured on the Scholastic Aptitude Test), showed that when demands on preliminary draft production were absent, argument production only increased for writers with *average* verbal ability. Writers with low verbal ability did not profit from the relaxed demands, suggesting that positive effects on quality depend not only on cognitive demands but also on verbal ability.

Even though the importance of resource demands in writing has been stressed repeatedly, most of the evidence is either correlational or shows the negative effects of increased resource demands on writing. As mentioned, few studies have tried

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

to *improve* writing by reducing resource demands. Moreover, there has been no research trying to improve *L2* writing by enhancing lexical retrieval skills. When an L2 is involved, lexical retrieval is likely to be more attention demanding than in the L1 and therefore is also likely to have more negative effects on L2 writing quality. In addition, the study by Glynn et al. (1982) suggested that positive effects on argument production appear conditional on verbal ability of the writers. As lexical retrieval is an important aspect of L2 verbal ability, enhancing lexical retrieval in the L2 seems to be a promising approach. The current study addresses the lack of L2 research in this area by investigating whether experimentally enhancing lexical retrieval in the L2 has a positive effect on L2 writing quality.

In an earlier study, we demonstrated that L2 lexical retrieval could be enhanced through an experimental intervention (Snellings, Van Gelderen, & De Glopper, 2002). We showed that a 4-week, highly focused training aimed at speeding up lexical retrieval resulted in both higher levels of accuracy and faster RTs for Dutch learners of English (L2). The effect sizes in this study were medium to large. The study indicated that training learners to focus on speed and providing them with immediate feedback on both speed and correctness effectively altered and improved an essential cognitive subprocess involved in L2 production, thus potentially reducing the cognitive demands of lexical retrieval in L2 writing.

The current study, which uses the same participants, examines the effects of enhanced lexical retrieval on several aspects of written L2 narratives. In this study, we first investigate whether the lexical retrieval training results in more frequent use of the trained words in the writing tasks. Second, we investigate whether the training results in learners being better able to express the details of the story and the desired content. Third, we explore whether enhanced lexical retrieval has an effect on global communicative quality of the texts. Global text quality is the result of many different skills associated with written language production, such as audience awareness, text organizational skills, and sensitivity to style. Clearly, fluency of lexical retrieval is but one skill that may influence global text quality.

Providing answers to these research questions involved dealing with a number of important considerations that guided the development of the instruments and scoring procedures in the study. An important consideration was the need to have an appropriate measure for speed of lexical retrieval. In previous studies, picture naming tasks have often been used as a measure of lexical retrieval. Picture naming tasks are good measures for lexical retrieval because, just as in written language production, the process they tap into proceeds from meaning to word form. However, the set of words in which we were interested, which includes nouns, pronouns, articles, verbs, adverbs, adjectives and combinations of these categories, is difficult to depict unambiguously with pictures. Because a specific focus on this diverse set of words was needed, a word translation task was deemed to be more appropriate. Translation tasks and picture naming tasks tap into similar processes in that both kinds of task involve concept mediation (Potter, So, von Eckardt, & Feldman, 1984). For related word translation tasks, see De Groot, Dannenburg, and Van Hell (1994) and Potter et al. (1984).

In view of our aim to increase the retrieval fluency of familiar words, rather than to teach new words, another important consideration was the need to compile a set of familiar words to be used in the training. Word-frequency ratings may not be the

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

best indication of familiarity, especially in view of the fact that there are no learner corpuses for the productive vocabulary of Dutch English as a foreign language (EFL) learners. Thus, in order to obtain a selection of familiar (productive) words, we conducted a pilot study in which 9th grade students had to write narrative texts using the cartoons to be used in the main study (see the Method section). In view of the relevance of collocations, for reducing cognitive effort (Weinert, 1995), we also selected some word combinations used in the narrative texts from the pilot study. Following Weinert (1995), we define word combinations, collocations, or formulaic language as multiword strings (e.g., "just in time") that can be retrieved as a whole chunk and are similar to individual lexical items, rather than being derived by linguistic rules.

A third important consideration was the need to appropriately assess the three aspects of the narrative texts relevant to the research questions, namely, number of trained words used, expression of essential content elements, and global text quality. Assessing the number of trained words used in the narratives is a relatively straightforward procedure. For the rating of global text quality, there are agreed-upon procedures (Schoonen et al., 2003). Only for the assessment of content elements did a special approach need to be developed. To assess whether the details of the story and the desired content were expressed, we first defined "content elements," using the definition of "narrative schemata" described by Stein and Trabasso (1982). According to these authors, schemata "specify the components of a story that should be included in a text, the temporal sequence of the narrative events, and the type of causal relations that should connect the events." These schemata enable us to establish whether certain parts of a story have been omitted. As pointed out by Stein and Trabasso, the underlying representation of a story in memory can be inferred from a text. As a result, even though some parts of the story may be deleted from the text structure (e.g., "internal responses"), they should be inferable from the text. Stein and Trabasso divided stories into a Setting and an Episode. The latter consists of five categories that contain specific information and that serve a different function in the schema. These categories are psychological rather than linguistic entities. The five categories are Initiating event, Internal response, Attempt, Consequence, and Reaction. An Initiating event can be an action that causes the protagonist to respond emotionally and to formulate a goal. Internal response is the emotional reaction of the protagonist that causes him or her to initiate action. An Attempt is the overt action to attain a goal, a Consequence is the event indicating goal attainment, and a Reaction is the emotional response of the protagonist to the outcome of his or her actions (for the schema structures, see the section on scoring). By determining whether content belonged to one of these five categories, we were able to establish whether essential content was expressed.

METHOD

Participants

A total of 103 Dutch students participated in this study, all at the beginning of their third year of secondary education (Grade 9, age 14–15). Students were from four

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

different classes, two classes from lower general secondary education ($n = 47$) and two classes from preuniversity education ($n = 56$). Students had received 2 full years of formal English tuition (two–three lessons per week) at secondary level, and a 2-year introductory course at the elementary level, focusing mainly on basic oral communication skills (maximum one lesson per week). Students filled in a questionnaire, which was completed by 101 students, concerning their gender and language background and containing questions such as whether they spoke Dutch with their parents and whether they had acquired Dutch as their L1. Of these students 44.6% were male and 55.4% were female, and the results showed that 90 students were native speakers of Dutch. The remaining 11 students were L2 speakers of Dutch whose native language was not English (although one student who had acquired Dutch as his L1 spoke English with his father and Dutch with his mother, and another student spoke Dutch with her parents but had acquired both English and Dutch as the L1). There was an even division of L2 speakers of Dutch across the two conditions.

Design

In a counterbalanced design, students from four classes were randomly assigned to one of two speed-training treatments. Each class was randomly split in half and the two resulting treatment groups were trained concurrently by two experimenters over a 4-week period. Each treatment used a different set of stimuli, either Word Set A or Word Set B. Students in the Trained A words condition were trained on Word Set A. Students in the Trained B words condition were trained on Word Set B. Students in each group were not trained on the other group's word set. After the training, all students were tested on speed and accuracy for both word sets (Snellings et al., 2002) and also wrote two narrative texts based on cartoons. Narrative A was likely to elicit the words in Word Set A, whereas Narrative B was likely to elicit the words in Word Set B.

Materials

Selection of training words. In order to select familiar (productive) words for the training, we conducted a pilot study in which ninth grade students had to write narratives using the same cartoons as in the current study. Words that were used at least once by the pilot study students were candidates for selection. Only candidate words that were likely to be commonly included in EFL textbooks in The Netherlands were selected. In this selection process, the first author was able to use his extensive experience in teaching English to Dutch students of this age group. Another criterion for selection was that the selected words had to have straightforward translations in Dutch. It was also necessary to ensure that for each picture there were sufficient words selected to describe the events. The selection included words belonging to different grammatical classes (e.g., single nouns, nouns with articles or determiners, verbs, adjectives, and adverbs) and we also selected some word combinations. The two sets of words are contained in Appendixes A and B.

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

Tasks and procedures

Prior to the training, a vocabulary test, to be used as a covariate, was administered. The students were subsequently trained in class for 4 weeks. To eliminate experimenter effects, each of the two trainers trained students in both conditions. All students received the same type of computer training and the same number of computer assisted exercises (but with different words). The training was given using specially designed software[1] that could be used outside the laboratory in a classroom setting, allowing a whole group of learners to be trained at the same time.

As our objective was to increase retrieval speed, students were actively encouraged to respond more quickly. The program provided them with a baseline against which they could chart their progress (i.e., previous RT on the same word). As the words were largely familiar, we assumed that working under time pressure would be a motivating factor as well, so that two instances of the translation exercise used the flashcard principle. The students encountered each word 10 times and received immediate feedback on correctness and current speed.

We consistently trained the same meaning of the words in order to ensure that mappings of stimuli to responses were consistent (Schneider & Shiffrin, 1977; Segalowitz & Gatbonton, 1995). For example, if the aim is to train the word "line" in a story about fishing, the word should always be used in the fishing sense and not in both the fishing line sense and the telephone line sense.

Students were trained in scheduled English classes. We used spaced presentations (in contrast to massed presentations), because psychological learning theories suggest that this is more effective (see discussion in Ellis, 1995). The standard duration of each training session was 50 min (except for one 45-min training session). There was a 1-week interval between each training session. After completion of the training sessions, we collected data on lexical decision skills and lexical retrieval skills in order to establish whether lexical retrieval had been enhanced. Within the next week, students completed the posttest writing tasks. All students first completed the writing task that corresponded to their training condition. Each word was trained using four different types of exercises, moving from receptive skills to productive skills. The exercises are described below.

Block exercise. The aim of this exercise was both to strengthen the receptive word knowledge and to train syntagmatic knowledge (i.e., knowledge of collocational restrictions). Students had to complete a short sentence as quickly as possible by choosing from two constituents (i.e., two "blocks"), the choice of one of which resulted in an incorrect word order.

Example

After some time

woke up she

Z M

Students were required to press either the Z or M key in order to indicate the constituent they thought should follow *After some time*.

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

Detection exercise. This exercise had a strong focus on the meaning of the words involved, in order to strengthen the connections between concepts and lexical forms. Students had to decide whether the meaning of the words and expressions was appropriate to the specific contexts provided. They had to focus on the meaning of the words in short sentences and detect semantic anomalies as fast as possible by indicating which word caused the anomaly.

Example

Three hunters (Z) had to be milked soon (M).

Students had to indicate whether *hunters* (Z) or *soon* (M) makes the sentence odd. Here, the word *hunters* makes the sentence odd because a word like *cows* is more likely.

Correction exercise. In the correction exercise, students were asked to assess words for appropriate use and, if necessary, to choose a better substitute from two alternatives. Like the detection exercise, this exercise also aimed at enhancing the connections between concepts and lexical forms. Students again had to focus on the meaning of words in short sentences. They also had to detect semantic anomalies as fast as possible, but this time they had to indicate which of two words would make the sentence logical (or accept the original word marked X).

Example

Sue felt very ill. After some *ages* (X) she felt better though.

time centuries

Z M

Students had to indicate whether *ages* (X) is semantically acceptable, or whether it should be replaced with *time* (Z) or *centuries* (M). In the example, *time* (Z) is the more acceptable alternative.

Translation exercise. In the word translation exercise, students provide translations for words presented in a meaningful context. Of the four exercise types, the translation exercise most closely matches lexical retrieval in written language production. The use of a translation task makes it possible to focus on more words at the same time (in contrast to picture tasks).

Students translated Dutch words presented in simple English carrier sentences. Both responses that provided the targeted training words and responses that offered acceptable (spelling) alternatives for the same concepts were scored as correct. In addition to the ordinary format, we developed a flashcard version in which the Dutch words and the English sentences disappeared from the screen after 15 s, thus encouraging the students to answer more quickly.

Example

voordat the bridge collapsed he had reached the other side

Correct translation: *before*

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

The order of the items in all the exercises was randomized for each presentation. In the block exercise, the order of the two constituents was also randomized. The exercises were tested in a small-scale pilot study in order to test the functioning of both software and hardware and also to gauge the exercises' effectiveness. We used information gained in this pilot study to make the instructions clearer and to place more emphasis on speed. Problems with the program and the design of the exercises were also solved. In the final experiment, students began with the block exercises, followed by the correction, detection, and translation exercises.

Dependent measures. After we trained lexical retrieval and established an increase in retrieval speed, students completed two cartoon-based writing tasks (Writing Tasks A and B). The restricted range of vocabulary elicited by the cartoon-based tasks provided a direct way of testing the effects of enhanced lexical retrieval on writing. Writing Task A was about two boys who go fishing, and Writing Task B was about a man who goes walking in the mountains and encounters a tiger (see Figures 1 and 2). In both tasks, the pictures were presented in chronological order. The dependent variables, which are "Number of trained words used," "Expression of content elements," and "Global text quality," are measured for both Writing Task A and Writing Task B.

Number of trained words used. We counted words appearing in the students' texts that were trained either in isolation or as part of a word combination. In addition, we looked at the occurrence of trained word combinations. Because our aim was to look into the effects of more efficient lexical retrieval and reduced resource demands, we only considered the use of words and word combinations that were trained literally.

Expression of content elements. On the basis of the cartoons, the first author devised the schema structures that represented the schema of each narrative. These narrative schemata assisted in identifying content elements, which were subsequently placed on a structured checklist (see Appendixes C and D). Before scoring, students' texts were divided into propositions. Scorers indicated whether each proposition contained an essential content element, added other information, or added irrelevant information. To facilitate scorer agreement, we constructed a set of guidelines and trained the scorers according to these. Decisions made while scoring the texts used in the training session were made explicit and written down, to be used as an extra guideline during scoring. An example of a guideline is that whenever a content element was not linguistically realized but was implied by other words in the text, it was also scored (e.g., "the owner warns the man" implies "there is an owner"). The structured checklist ensured that only content elements that occurred in the appropriate part of the story were scored.

Global text quality. The writing tasks instructed students to use the appropriate register for addressing their peers, to do their best to interest the reader, and to write in such a way as to produce a story that would be coherent in the absence of the cartoons.

Two independent raters scored the texts for global quality. The raters were instructed to give a general rating for the quality of the text in relation to five

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

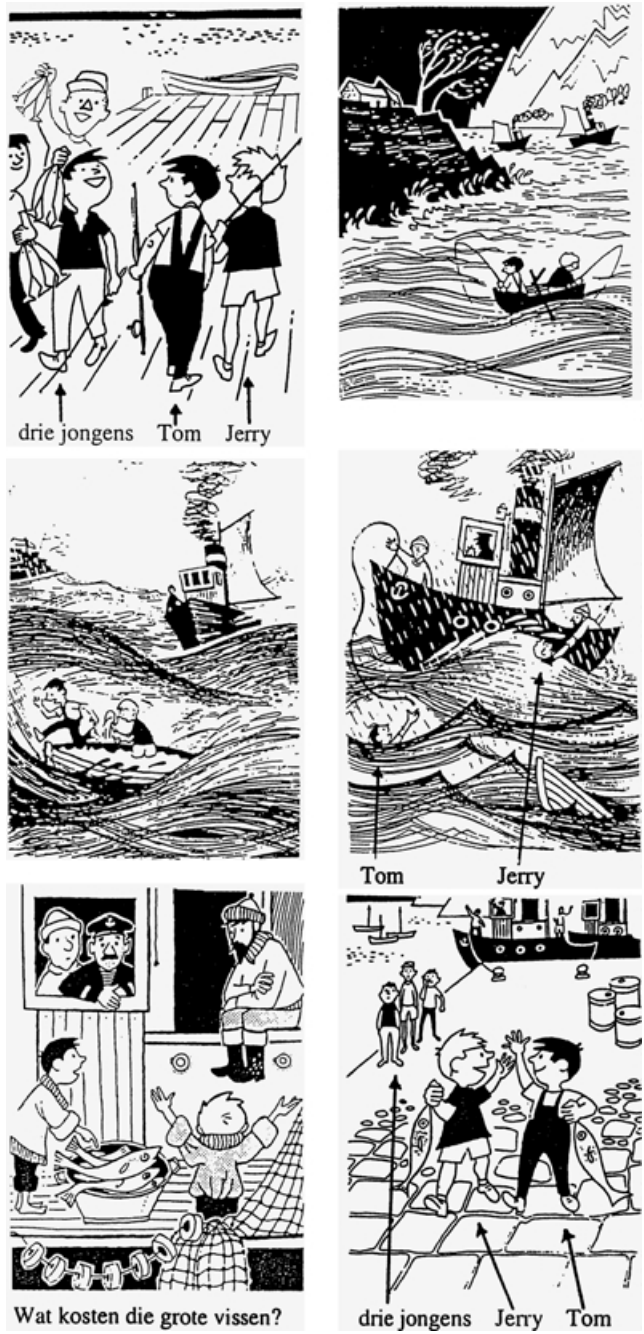


Figure 1. The cartoon used for Writing Task A. *Note.* From *Composition through pictures*, by J. B. Heaton, 1981, Harlow, UK: Longman. Copyright 1966 by Longman. Reprinted with permission. Illustrations by James Moss.

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press



Figure 2. The cartoon used for Writing Task B. *Note.* From *Composition through pictures*, by J. B. Heaton, 1981, Harlow, UK: Longman. Copyright 1966 by Longman. Reprinted with permission. Illustrations by James Moss.

benchmark texts, using any numerical value they thought appropriate. The raters were able to consult the benchmark texts while scoring. The five benchmark texts had been selected on the basis of a separate analysis according to a scaling procedure (Blok, 1986), involving six raters. The benchmarks represented a very weak text (10th percentile), a weak text (25th percentile), an average text (50th percentile = score of 100), a good text (75th percentile) and a very good text (90th percentile). Raters were instructed to concentrate on content and organization and to consider language errors only to the extent that they interfered with the understanding of the story.

Covariate. To control for initial differences in vocabulary knowledge, we used the scores on the vocabulary test as a covariate. The vocabulary test (an adapted

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

version of a test used in another NELSON project study; see Schoonen et al., 2002) consisted of multiple-choice items in English. Each item contained a neutral carrier sentence with the target word in bold print, and students had to choose between four alternative Dutch translations of the target word. The words were selected from the 5000 most frequent words in the COBUILD corpus (COBUILD, 2000). On the basis of a pretest, we removed items that were too easy ($p > .90$) to obtain a more discriminative test. In addition, we removed items with negative item–total correlations. The final test contained 64 items. Scores on the vocabulary test correlated significantly ($p < .05$, one tailed) with all three dependent variables, making the test suitable for use as a covariate. In the case of Writing Task A, $r = .24$ with Number of trained words used, $r = .19$ with Expression of content elements, and $r = .23$ with Global text quality. In the case of writing assignment B, $r = .21$ with Number of trained words used, $r = .40$ with Expression of content elements, and $r = .38$ with Global text quality.

Data collection. Both the vocabulary test and the writing assignments were paper and pencil tests. The class teacher administered the vocabulary test before the start of the training. Following the training period, a trained research assistant presented the instructions for the writing tasks.

Analyses. Means, standard deviations, and estimates of rater reliability were calculated (i.e., percentage of rater agreement for Use of content elements and Cronbach's α for the Global text quality score). The hypotheses were tested by two one-way multiple analyses of covariance (MANCOVAs) on the measures we derived from the writing tasks, with Condition (A or B) as an independent measure; Number of trained words used, Expression of content elements, and Global text quality as dependent measures; and Vocabulary knowledge as a covariate.

In the case of the MANCOVAs on the writing tasks, results of evaluation of the assumptions of independent observations, equal population covariance matrices for the dependent variables, a linear relationship between the dependent variables and the covariate, homogeneity of the regression slopes, and reliability of the covariate were satisfactory (Stevens, 1996). The only exception was the assumption of multivariate normality of dependent variables in each group for the MANCOVA on the Writing Task B measure. In the case of Expression of content elements in the Trained B words condition, the Kolmogorov–Smirnov test was significant, indicating that multivariate normality did not hold here. However, because this is the only group and dependent variable for which this assumption was not valid, the effect on power is small (Stevens, 1996, p. 247). In addition, MANCOVA is rather robust regarding a violation of multivariate normality.

Scoring. In the case of Number of trained words used, we used the summed scores of the number of isolated words, the number of words used in isolation that were trained as part of a word combination, and the number of the word combinations. We did not include derivations or inflected forms. If words were used more than once, each token was included in the total score.

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

In the case of Expression of content elements, after establishing rater reliability, a single trained scorer scored all the texts. Rater reliability was established by comparing the percentage of agreement between the scorer and the first author on a sample of 20 texts from each writing task. The percentage of agreement was 89.9% for Writing Task A and 87.4% for Writing Task B.

Finally, in the case of Global text quality, two raters scored each text independently. Panel reliability (Cronbach's α) was .84 for Writing Task A and .88 for Writing Task B.

Missing values. Only students who had been present at all the lexical retrieval training sessions were included in the writing task analyses for the trained tasks (Writing Task A, $N = 41$; Writing Task B, $N = 42$). Presence at training sessions was not required for inclusion of data from the tasks involving untrained words (i.e., Writing Task B for students trained on word set A, and vice versa, so that Writing Task A, $N = 52$, Writing Task B, $N = 50$). As a result, the analysis of Writing Task A was based on scores from 93 students and the analysis of Writing Task B on 92 students.

RESULTS

In this section we first present descriptive measures for the tests used in this study. Then, we discuss the effects of the lexical retrieval training on our dependent measures for both Writing Tasks A and B.

Descriptive and psychometric statistics

Dependent measures. As explained in the Missing Values section, the number of students in each condition differs, depending on which text is analyzed (see Table 1). When examining these data, it has to be taken into account that comparison of scores for Writing Tasks A and B is not warranted, because the writing tasks are not equivalent. Only comparisons between groups on the same writing measure are justified.

In Table 1 we give an overview of the unadjusted scores for the Writing Task A and B measures. These unadjusted scores give a first impression of the differences on the dependent variables that exist between the two experimental groups, but they cannot be used to evaluate experimental effects. These effects are analyzed in the following sections.

Covariate. The mean score for the vocabulary test was 45.19 ($N = 103$, maximum possible score = 64, $SD = 8.74$) and the reliability was .87. Thus, on average the vocabulary test was of intermediate difficulty and its internal consistency was good.

Experimental effects. We expected superior scores for students in the trained words condition on the writing task measures. Our multivariate test (Wilks' lambda) for the Writing Task A measures was significant with $F(3, 88) = 18.105$, $p < .001$, $\eta^2 = .382$. Our multivariate test (Wilks' lambda) for the

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

Table 1. *Descriptives for Writing Measures A and B: Unadjusted mean (M) scores and standard deviations (SD)*

	Trained A Words Condition[a]		Trained B Words Condition[a]	
Words	*M*	*SD*	*M*	*SD*
Used A words (71)	41.05	10.24	28.87	10.10
Used B words (69)	22.06	8.45	31.19	9.13
Expressed content elements				
A (47[b])	25.39	5.64	22.54	5.38
B (31[b])	16.12	4.40	17.12	3.38
Global text quality				
Score A	100.62	10.75	96.62	12.26
Score B	95.38	11.72	95.40	10.82

Note: The maximum possible scores are in parentheses.
[a]The measures for Writing Task A are based on 41 Trained A words condition students and 52 Trained B words condition students. The measures for Writing Task B are based on 50 Trained A words condition students and 42 Trained B words condition students.
[b]The maximum possible score does not correspond with the number of lexical items trained, because words in lexical items that contain more than one word are also counted in isolation. The same procedure was followed for words that were part of a collocation. It is also possible that students used words more than once, so that the maximum possible score given here can be higher in practice.

Writing Task B measures was also significant with $F\,(3, 87) = 14.387$, $p < .001$, $\eta^2 = .332$. All reported means are derived from a MANCOVA and therefore adjusted for vocabulary knowledge.

Number of trained words used. In the case of number of trained words used in narrative A, our follow-up univariate analyses show a significant main effect for condition with $F\,(1, 90) = 32.032$, $p < .001$, partial $\eta^2 = .262$. As expected, students in the Trained A words condition ($M = 40.83$, $SE = 1.56$) used more A words than students who were not trained on these words ($M = 29.04$, $SE = 1.38$). We expected similar results for narrative B, with number of B words used higher for students in the Trained B words condition. Indeed, once again, our follow-up univariate analyses show a significant main effect for condition with $F\,(1, 89) = 27.574$, $p < .001$, partial $\eta^2 = .237$. Students in the Trained B words condition ($M = 31.32$, $SE = 1.31$) used more B words than students who were not trained on these words ($M = 21.96$, $SE = 1.20$). Effect sizes can be transformed into Cohen's f (Cohen, 1988). A small effect is an f value of .10 and this can be transformed into $\eta^2 = 0.0099$ (the transformation uses η^2 instead of partial η^2 which SPSS reports); a medium effect is an f value of .25, which equals $\eta^2 = 0.0588$; and a large effect is an f value of .40, equaling $\eta^2 = 0.1379$. In the case of both the A and B words, the effect size was large, with respective values of 25% ($\eta^2 = 0.2469$) and 23% ($\eta^2 = 0.2262$) of the variance in scores due to the training.

Expression of content elements. In the case of Expressed content elements we also expected higher scores for the trained students. For Expressed content

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

elements A, our follow-up univariate analyses show a significant main effect for condition with $F(1, 90) = 5.567$, $p < .05$, partial $\eta^2 = .058$. Students in the Trained A words condition obtained a higher score on Expressed content elements A ($M = 25.30$, $SE = .85$) than students who were not trained ($M = 22.61$, $SE = .76$). A total of almost 6% ($\eta^2 = 0.0562$) of the variance in scores, close to a medium effect in Cohen's terms, can be accounted for by the training. Although students in the Trained B words condition obtained a slightly higher score on Expressed content elements B ($M = 17.21$, $SE = .56$) than students who were not trained ($M = 16.05$, $SE = .52$), this difference is not significant with $F(1, 89) = 2.307$, $p = .13$, partial $\eta^2 = .025$.

Global text quality. Results of our univariate follow-up analyses show no significant effect of the training on Global text quality for narrative A with $F(1, 90) = 2.253$, $p = .14$, partial $\eta^2 = .024$. Nevertheless, adjusted scores are in the expected direction, with students in the Trained A words condition scoring higher on the Global text quality score A ($M = 100.39$, $SE = 1.79$) than students who were not trained ($M = 96.80$, $SE = 1.58$). For narrative B, results from the univariate follow-up analyses also show no significant effect on Global text quality with $F(1, 89) = 0.042$, $p = .84$, partial $\eta^2 = .000$. The mean Global text quality score for narrative B of students in the Trained B words condition ($M = 95.64$, $SE = 1.63$) was nearly identical to the mean score of students who were not trained ($M = 95.19$, $SE = 1.49$). Because the interaction between the covariate vocabulary and condition (homogeneous regression slopes assumption) was borderline significant ($p = 0.05$) for the dependent variable of Global text quality, we also carried out an analysis without a covariate. However, the results of this analysis were also nonsignificant.

DISCUSSION

Lexical retrieval is an important subprocess in written language production that has been neglected in research on L2 writing instruction. Thus far, research into cognitive demands on writing has been correlational or has demonstrated the importance of resource demands by showing the negative effects of increasing cognitive demands on the writing process. In contrast, research on reducing cognitive demands has been limited. In the current study, we have investigated a new approach to decreasing the cognitive effort involved in L2 writing and its effects on writing quality.

We assumed that if students had more attention to spare because lexical retrieval proceeded with less effort, this would benefit other writing processes such as planning and organization on a local and a global level. This could result in a better expression of content. The results of our study do not, however, support these assumptions unequivocally.

We will reflect on two possible reasons why a significant effect was found in one condition but not in the other. One possibility is that the link between the trained words and the content to be expressed in narrative B was less strong than in narrative A. It is possible that the content elements in narrative B could easily be expressed using words other than the trained words, thus reducing the influence of the trained words. To check the plausibility of this assumption, we calculated

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

the correlations between Number of trained words used and Expression of content elements. These correlations were substantial and positive in both writing tasks. For both tasks, using more words implied better expression of content elements. However, inspection of the confidence intervals at $p = .05$ showed that the correlations were higher for narrative A ($r = .81$, $p < .01$, one tailed) than for narrative B ($r = .56$, $p < .01$, one tailed), lending support to the idea that there was a closer relationship between trained words and content elements in narrative A.

Another possibility could be that the content of narrative B was more difficult to express. Even though students in the Trained B words condition could retrieve a selection of the necessary words with less effort, this did not help them in expressing the content in a more detailed way. To check the plausibility of this explanation, we calculated correlations between vocabulary *knowledge* and the dependent variables Expression of content elements and Global text quality. The correlations with our content and global quality measures were lower in the case of narrative A ($r = .19$ for Expression of content elements, and $r = .23$ for Global text quality) than in narrative B ($r = .40$ for Expression of content elements, and $r = .38$ for Global text quality). This would suggest that the vocabulary necessary for writing narrative A was easier than that for narrative B, which in turn points in the direction of more difficult vocabulary for expressing the content of narrative B. Confidence intervals, however, showed that these differences were not significant at $p = .05$.

Our explorative results showed that even though the data were in the expected direction in the case of narrative A, the scores on Global text quality of the trained students in both conditions were not significantly higher than the scores of the untrained students. It could be argued that we trained the wrong set of words. Therefore, to check whether our selection of trained words was functional for writing texts with high global quality, we calculated correlations between the Number of trained words used and Global text quality. The correlation was .74 ($p < .01$, one tailed) in the case of narrative A and .58 ($p < .01$, one tailed) for narrative B. These correlations suggest that the selections were appropriate for the objectives set and that word set selection does not explain the lack of effect. In addition, as was the case with the expression of content data, the correlations appeared to be somewhat higher for narrative A than for narrative B. This seemingly closer relationship between trained words (unadjusted) and Global text quality (unadjusted) in narrative A could have explained why the data were in the expected direction in the case of narrative A but not for narrative B. However, confidence intervals indicated that this difference was not significant at $p = .05$.

Implications for a theory of lexical competence, suggestions for future research, and educational relevance

In this study, words that were trained in a computerized setting had to be transferred and applied in a writing task. A discussion by Anderson (2000, p. 307) makes clear that transfer between skills can take place to the extent that they share the same elements. In writing a text there are elements (such as planning and monitoring) involved that are in addition to the trained elements. Our evidence shows that despite the required additional elements, the trained students used more words from the trained word set, which makes this training relevant for improving writing

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

skills. This finding is also relevant for theories of lexical competence. DeKeyser (2001) argued that rapid *understanding* of sentences in a language does not mean that one can *produce* them quickly. Henriksen (1999) provided a theoretical account of the different dimensions that play a role in the development of vocabulary. She argued that there are three possible causes for a difference between receptive and productive competence. The cause can either be a difference between input and output specifications (encoding orthograpy and sound; see Ellis, 1995), the level of "automaticity" (Meara, 1996), or the quality of the meaning representation. In our view, the learning of output specifications through practice and extensive repetition may eventually lead to automaticity, so that learning output specifications and achieving automaticity in production could be subsumed under one factor labeled "retrieval speed." The current findings have implications for such a theory of productive competence. The increase in word use as a result of the speed training shows that even in the case of relatively easy words, speed in accessing lexical items in the mental lexicon (retrieval speed) is important for productive competence.

Now that effects have been demonstrated using cartoon tasks, we also need to get a better understanding of the effects on more common text types. From a theoretical point of view, it is also important to know more about the influence of the characteristics of the words trained, such as word type and degree of familiarity, on the effectiveness of the speed training. Although we have demonstrated that increasing speed increases productive competence, it remains an issue for further research of whether actual "automaticity" (for a review of this concept, see DeKeyser, 2001) needs to be trained.

Apart from the increased use of trained words, the relationship between enhanced lexical retrieval and the expression of detailed content in L2 writing is of interest. A think-aloud study of L2 writing by Zimmermann (2000) may clarify the process underlying local effects of lexical skills on text content. Zimmerman showed that tentative formulations and simplified tentative formulations uttered before the actual act of writing are relatively frequent. Tentative formulations are formulations that are rejected and replaced by other formulations in the actual written text. Once lexical retrieval proceeds effortlessly, it could facilitate writing in two different ways. In the first place, it may help students to retrieve the necessary vocabulary directly, without having to simplify the wording of the concepts they want to express. In this case, the quality of the formulation improves directly, and attention does not need to be devoted to tentative formulations. Alternatively, enhanced lexical retrieval may lead to the rapid generation of a number of tentative formulations, freeing the writer to select the most appropriate formulation for use (see also Van Gelderen & Oostdam, 2004). In both scenarios, formulating proceeds smoothly and students do not become bogged down in retrieving tentative formulations.

We did not find significant effects on global writing quality, which is interesting in view of the theories of both Hayes (1996) and Kellogg (1996), who suggested that trade-offs among processes could have an effect on text quality. Following their line of reasoning, more attention needing to be devoted to lexical retrieval should result in less attention that can be devoted to planning, monitoring, and other cognitive operations involved in writing. However, at a global text level, other skills such as organizing and revising ability may have a bigger impact on

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

writing than lexical retrieval. Evidence that is consistent with this interpretation has been reported by a number of authors. For example, it has been argued that the effects of reduced cognitive demands only operate at the local text level and not at a global text level. Penningroth and Rosenberg (1995) only found an effect of reduced demands on local coherence but not on holistic quality, suggesting that for holistic quality other factors come into play. Penningroth and Rosenberg's study also showed that even though cognitive demands were experimentally increased, the overall level of cognitive effort, as measured by reaction time on a secondary task (i.e., pressing a key as fast as possible when a tone occurred during writing), did not increase. Writers seemed to cope with high resource demands by changing their writing strategies. Interestingly, think-aloud reports made clear that the undergraduate students in this study changed the distribution of planning, translating, and reviewing across the writing process. Under a high processing load, students did not translate in the first phase of writing ("the first third of the writing session") but seemed to be forced to catch up in the third and final phase of the writing process. In the first phase they were already reviewing, a process that is normally reserved for the final phase.

Scardamalia et al. (1982) argued that, when attention was focused at the graphical level of writing, writers had to reconstruct their plans and the gist of their texts and this impeded fluent production. Typically, this resulted in reduced quantity produced but had no effect on the rated quality of the texts. It should also be noted that even though the subprocess of lexical retrieval has become faster and more efficient, other processes essential to writing have not. Experienced writers' self-reports demonstrate the amount of effort even such writers have to invest in the writing process (for a discussion, see Torrance & Jeffery, 1999). Crucially, and in contrast to beginning writers in L2, expert writers show a far greater degree of metacognitive control. Although it is safe to assume that some of our writers (Grade 9, age 14–15) possess metacognitive skills in their mother tongue that may be transferred and used in the L2 when resource demands are not too high, other writers may lack these skills altogether, so there is nothing to transfer. Glynn et al. (1982) raised a similar issue when their study showed that only undergraduates with average verbal ability on the Scholastic Aptitude Test profited from reducing resource demands by producing more arguments in their texts. They found no effects in the case of students with low verbal ability and concluded that "additional capacity can provide writers only with the opportunity to increase their pool of persuasive arguments; it cannot remediate deficiencies in ideational fluency." In the present study, enhanced lexical retrieval may provide the *opportunity* to use metacognitive skills, but it cannot make up for lack of such skills. In future research it is therefore advisable to also measure metacognitive skills in the L1 and L2 and, if found to be lacking, to train these skills. Future research needs to determine whether training in lexical retrieval could be effective in improving the global text quality of students with different levels of metacognitive skills. Research is also needed to establish whether, for students with less-developed metacognitive skills, it might be more effective to begin by specifically training metacognitive skills, before training lexical retrieval skills.

Another relevant point to consider is that we did not measure cognitive load directly. This can be taken up more directly, either by measuring latencies on a

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

secondary task during writing or by establishing individual differences in attentional capacity on a complex span measure (see, e.g., Ransdell & Levy, 1996). It is reasonable to assume that effects of enhanced lexical retrieval may be stronger for students with a smaller span, as students with a larger span may be better able to cope with higher resource demands of lexical retrieval.

The results of this study also have significant implications for writing instruction. The increase in the productive use of words and the more detailed expression of essential content elements represent an improvement in important aspects of writing. The implication of this finding is that in teaching, apart from teaching the meaning of new words, attention should be focused on lexical retrieval as well. Because word combinations as automatic sequences can be a means of reducing cognitive effort by freeing processing time (Weinert, 1995), it might be useful for educational and research purposes to place a stronger focus on these structures. Finally, it should be stressed that a focus on enhanced lexical retrieval is not a plea for the repetitive training of words through drills. In this respect our approach is related to the communicative approach to automaticity suggested by Gatbonton and Segalowitz (1988). The training outlined in this study focuses not merely on form, but also on the meaning of utterances in meaningful contexts.

APPENDIX A

Lexical items trained in the A condition

Dutch	English	Dutch	English
Een paar uur later	A few hours later	Kleine vissen	Small fish
Gelukkig	Luckily	Vis vangen	Catch fish
Gelukkig voor hen	Lucky for them	Vis kopen	Buy fish
Op maandag	On Monday	Hun vissen	Their fish
Of	If	Is groter dan	Is bigger than
In feite	In fact	Van de meisjes	Girls'
Urenlang	For hours	Roeide	Rowed
In de tussentijd	Meanwhile	Naar de zee	To the sea
Uiteindelijk	Finally	Merkten niet	Didn't notice
Er zijn ook	There are also	Is niet sterk	Isn't strong
Dat is de reden dat	That's the reason	Erg groot	Very big
Zij	They	Grote golven	Big waves
Besloten om	Decided to	Veel wind	A lot of wind
Huurden	Hired	Een zware storm	A heavy storm
Een kleine boot	A small boat	Begon	Began
Op weg naar	On the way to	Regenen	To rain
Wij ontmoetten	We met	Zonk	Sank
Drie jongens	Three boys	Vielen van	Fell off
De andere jongens	The other boys	Waren aan het verdrinken	Were drowning
Die	Who	Riep om hulp	Called for help
Vingen	Caught	Hadden geluk	Were lucky
Lachtten om	Laughed at	In de buurt	Nearby
Veel vissen	Many fish	Een andere boot	Another boat

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

APPENDIX A (*cont.*)

Dutch	English	Dutch	English
De vissers	The fishermen	Had een plan	Had a plan
Hen	Them	Hoeveel geld	How much money
Gooiden	Threw	Gingen terug naar	Returned to
Een touw	A rope	Terug	Back
Trekken	Pull	In de haven	In the harbor
Uit	Out of	Bedankten	Thanked
Werden gered	Were saved	Waren verbaasd	Were surprised
Een emmer met	A bucket of	Waren jaloers	Were jealous
Grote vissen	Big fish		

APPENDIX B

Lexical items trained in the B condition

Dutch	English	Dutch	English
De volgende dag	The next day	Een stok	A stick
Hoewel	Although	Voor	In front of
Ondanks	Despite	Was aan het zitten	Was sitting
Op het moment dat	At the moment that	Hoort	Hears
Onverwacht	Unexpectedly	Iets	Something
Net op tijd	Just in time	Kijkt omhoog	Looks up
Meteen	Immediately	Ziet	Sees
Snel erna	Soon after	Op	On
Zelfs nu nog	Even now	Een rots	A rock
Sinds die dag	Since that day	Was bang van	Was afraid of
Een man	A man	Viel aan	Attacked
Ging in	Went into	Probeerde	Tried
Klimmen	Climbing	Raken	To hit
De bergen	The mountains	Met	With
Kwam langs	Passed	Miste	Missed
Een restaurant	A restaurant	Brak	Broke
De eigenaar van de auto	The owner of the car	Sprong weg	Jumped away
Vertelde	Told	Stootte zijn hoofd	Hit his head
Een tijger	A tiger	Tegen een boom	Against a tree
Erg gevaarlijk	Very dangerous	Duizelig	Dizzy
Luisterde niet	Didn't listen	Pakte	Took
Was niet bang	Was not afraid	Een steen	A stone
Verliet	Left	Op het hoofd	On the head
Zag	Saw	Had gedood	Had killed
Een mooie waterval	A beautiful waterfall	Drie jagers	Three hunters
Stopte	Stopped	Juichten hem toe	Cheered him
Uitrusten	To rest	Erg dankbaar	Very grateful
Onder een boom	Under a tree	Veilig	Safe
Zette	Put		

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

APPENDIX C
Schema structure for Expression of content elements for Writing Assignment A

Code 20: all additional items
Code 100: all irrelevant items

All repetitions (also 2nd or 3rd time) get a 1 after the code: 1.1.1. Implicit elements are also counted if repeated explicitly!

1. Setting (number of propositions are counted here)
   Introduction: Tom and Jerry and boys
   1.1. Who are there? (e.g., there are two boys)
   1.2. Where are they from? (e.g., village/town)
   1.3. Why are they there? (e.g., go fishing)
   Everything that does not fit in these categories is scored additional (code 20).
2. Initiating event (picture 1). The number of content elements are counted from here on; they do not have to occur literally.
   2.1. Boys are at the harbor/near the sea/on the pier/on their way to the sea/going to the sea
   2.2. Boys meet/bump into/see/pass/run into three other boys
   2.3. Other boys are smiling/laughing/bragging
   2.4. Other boys show fish
   2.5. Other boys have fish/other boys have caught fish
   2.6. Other boys have tiny/small/little fish (the other boys show the small fish they have caught = 3)
   2.7. The sea is calm/the weather is fine/it is a nice day
3. Internal response
   3.1. Tom and Jerry are impressed
   3.2. "Feeling of Tom and Jerry: want to catch fish even more badly," "they hoped to catch fish as well"
4. Attempt (picture 2)
   4.1. Boys go fishing/boys are at sea/are in the boat/are on the open sea/are near the coast/are further from the coast/boys are fishing
   4.2. The weather changes/the weather got bad/sky turned black/wind got harder/it got windy (change; *4.2 implies 4.3*)
   4.3. Waves are high/waves are big/sea is wild/there is a big storm
   4.4. Their boat is small
5. Consequence (picture 3)
   5.1. Boat is leaking/boat is full of water/boat is damaged
   5.2. Boys are scared/afraid ("they called for help" = 20 + 5.2)
   5.3. Boys jump out/fall out (*5.3 implies 5.10*)
   5.4. Boy removes water from boat/boy tries to get water from boat
   5.5. Boys are lucky
   5.6. There is a boat/a boat sees them/fishermen see them
   5.7. The boat is big
   5.8. The boat (captain) comes closer/went to them
   (picture 4)
   5.9. The boy's boat sinks
   5.10. Boys are swimming/boys are in the water/boys are nearly drowning (*5.10 is different from 5.9*)

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

5.11. One man throws a line (rope)
5.12. Man pulls boy out
5.13. Boy grabs a line/takes hold of a line/catches a line
5.14. Boy grabs hand
5.15. Boys are saved/rescued/picked up/are helped/fishermen save the boys/boy climbs on board

(picture 5)

5.16. Boys are on the boat/on deck
5.17. Boys haven't caught fish/boys have no fish
5.18. Fishermen have fish/boys see fish/there are a lot of fish
5.19. The fish is big

6. Reaction
   6.1. Boys are disappointed: no fish
7. Internal Response
   7.1. Boys want to return with fish
8. Attempt
   8.1. Boy has an idea
   8.2. Boys want to buy fish/boys ask fish/boy asks price of the fish
   8.3. The fish must be big fish
9. Consequence
   9.1. Boys get fish/buy fish/grab fish/fishermen sell fish/give fish (score big fish here if not scored yet)

   (picture 6)

   9.2. Boys return/fishermen take them to the shore/boys are back at the harbor/boys are back home
10. Reaction
    10.1. They meet the boys again (essential: the same boys)
    10.2. Boys show the fish/let see/other boys see the fish
    10.3. The fish is bigger than that of the other boys/size of the fish
    10.4. Other boys have feeling of having lost (surprise/sad/amazed/jealous/disappointed/open mouthed)
    10.5. Boys are laughing/boys are happy/boys have feeling of success (proud)

APPENDIX D
Schema structure for Expression of content elements for Writing Assignment B

Note: Either lion or tiger can be used.
Code 20: all additional items
Code 100: all irrelevant items

All repetitions (also 2nd or 3rd time) get a 1 after the code (1.1.1). Implicit elements are also counted if repeated explicitly!

1. Setting (number of propositions are counted here)
   Introduction: Tiger
   1.1. Who is he? (hiker/hunter? a man + James = 1.1)
   1.2. Where is he from?
   1.3. Why is he there?

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

Everything that does not fit into these categories is scored additional (code 20).

2. Initiating event (picture 1). The number of content elements are counted from here on; they do not have to occur literally.
   2.1. He arrived at a restaurant/at a hostel/at an inn/a house/he was near a hostel.
   2.2. He meets another man/there is a waiter/owner of the restaurant.
   2.3. The owner warns man/says that it is dangerous.
   2.4. The owner says there are tigers.
3. Internal response
   3.1. Man wouldn't listen/disregards the story
   3.2. Man is not afraid/is brave
4. Attempt (picture 2)
   4.1. Man goes into the mountains/he is walking in the mountains
   4.2. Man goes unarmed
   4.3. Man is tired/man wants to have a rest

   (picture 3)

   4.4. Man stops/man takes a break/man sits down/man sits against a tree
5. Interrupting event
   a. Initiating event
   5.1. Man sees/man hears tiger/tiger appears
6. Internal response
   6.1. Man is frightened
   6.2. Man wants to defend himself
7. Attempt (picture 4)
   7.1. Tiger attacks/tiger tries to get at him/tiger jumps at him
   7.2. Man jumps away/steps aside/man gets up
   7.3. Tiger misses man
   7.4. Man tries to hit tiger/man tries to hit tiger with stick
   7.5. Man misses tiger with stick
   7.6. Tiger hits a tree/tiger bumps his head
   7.7. Tiger is dizzy/knocked out (not headache) *after hitting the tree*

   (picture 5)

   7.8. Man hits tiger with a rock
8. Consequence
   8.1. Tiger is dead/unconscious; *after being hit with rock*/knocked out (not asleep)/man kills tiger

   (picture 6)

   8.2. Man has one foot on tiger (triumphant)
   8.3. Hunters/people from the village see him/watch him/hunters come
   8.4. Hunters are armed
   8.5. Hunters think he is a hero/hunters cheer him
   8.6. He has been lucky
9. Reaction
   9.1. Man has feeling of success/man is proud/man is happy

ACKNOWLEDGMENTS

Part of the findings reported here were also originally presented at the 11th Conference of the European Second Language Association (EUROSLA), Paderborn, Germany,

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

September 27, 2001. This research was supported by a grant from the Dutch Organization for Scientific Research (NWO, 575-36-001). This article benefited from the comments of the other NELSON (an acronym in Dutch that translates as "Reading and Writing in Dutch and English in [Secondary] Education") project members, Jan Hulstijn, Rob Schoonen, Annegien Simis, and Marie Stevenson. We also thank Marie Stevenson for her valuable contribution in the revision stage of this work and Fiona Rose for her help in scoring the content elements. Finally, we thank the Editor, the anonymous reviewers, and Gert Rijlaarsdam for their useful and valuable comments on an earlier version of this manuscript.

NOTES

1. This software was a joint development of the current study and a reading intervention study by Annegien Simis and Jan Hulstijn. Although the exercises in both studies differed, the framework was identical and was programmed by Boudewijn de Roovere. For a more detailed description, click the NELSON tab on the following site: http://members.ams.chello.nl/b.roovere/.

REFERENCES

Alamargot, D., & Chanquoy, L. (2001). *Studies in writing: Vol. 9. Through the models of writing.* Dordrecht: Kluwer Academic.

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). New York: Worth.

Blok, H. (1986). Essay rating by the comparison method. *Tijdschrift voor Onderwijsresearch, 11*, 169–176.

Brown, J. S., McDonald, J. L., Brown, T. L., & Carr, T. H. (1988). Adapting to processing demands in discourse production: The case of handwriting. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 45–59.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing. Generating text in L1 and L2. *Written Communication, 18*, 80–98.

COBUILD. (2000). *Bank of English corpus*. Retrieved from http://titania.cobuild.collins.co.uk/about.html.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Groot, A. M. B., Dannenburg, L., & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language, 33*, 600–629.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge: Cambridge University Press.

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning, 8*, 103–128.

Favreau, M., & Segalowitz, N. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & Cognition, 11*, 565–574.

Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Hillsdale, NJ: Erlbaum.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22*, 473–492.

Glynn, S. M., Britton, B. K., Muth, D., & Dugan, N. (1982). Writing and revising persuasive documents: Cognitive demands. *Journal of Educational Psychology, 74*, 557–567.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 1–27). Mahwah, NJ: Erlbaum.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.

Heaton, J. B. (1981). *Composition through pictures* (19th ed.). Harlow, UK: Longman.

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303–318.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.
Kellogg, R. T. (1994). *The psychology of writing*. New York: Oxford University Press.
Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57–71). Mahwah, NJ: Erlbaum.
Kellogg, R. T. (1999). Components of working memory in text production. In G. Rijlaarsdam & E. Esperet (Series Eds.) & M. Torrance & G. C. Jeffery (Vol. Eds.), *Studies in writing: Vol. 3. The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 43–61). Amsterdam: Amsterdam University Press.
Koda, K. (1996). L2 word recognition research: A critical review. *The Modern Language Journal, 80*, 450–460.
Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*, 149–174.
Lea, J., & Levy, C. M. (1999). Working memory as a resource in the writing process. In G. Rijlaarsdam & E. Esperet (Series Eds.) & M. Torrance & G. C. Jeffery (Vol. Eds.), *Studies in writing: Vol. 3. The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 63–82). Amsterdam: Amsterdam University Press.
Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1998). *A theory of lexical access in speech production.* Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.
McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*, 299–325.
McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology, 86*, 256–266.
Meara, P. (1996). *The vocabulary knowledge framework*. Vocabulary Acquisition Research Group. Retrieved from http://www.swan.ac.uk/cals/calsres/vlibrary/pm96d.htm.
Penningroth, S. L., & Rosenberg, S. (1995). Effects of a high information processing load on the writing process and the story written. *Applied Psycholinguistics, 16*, 189–210.
Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
Potter, M. C., So, K. F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*, 23–38.
Ransdell, S., & Levy, C. M. (1996). Working memory constraints on writing quality and fluency. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 93–105). Mahwah, NJ: Erlbaum.
Scardamalia, M., Bereiter, C., & Goelman, H. (1982). The role of production factors in writing ability. In M. Nystrand (Ed.), *What writers know* (pp. 173–210). New York: Academic Press.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.
Schoonen, R., Van Gelderen, A., Glopper, K. d., Hulstijn, J., Snellings, P., Simis, A., & Stevenson, M. (2002). Linguistic knowledge, metacognitive knowledge and retrieval speed in L1, L2 and EFL writing. In G. Rijlaarsdam (Series Ed.) & S. Ransdell & M.-L.Barbier (Vol. Eds.), *Studies in writing: Vol. 11. New directions for research in L2 writing* (pp. 101–122). Dordrecht: Kluwer Academic.
Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing; the role of linguistic knowledge, speed of processing and metacognitive knowledge. *Language Learning, 53*, 165–202.
Segalowitz, N., & Gatbonton, E. (1995). Automaticity and lexical skills in second language fluency: Implications for computer assisted language learning. *Computer Assisted Language Learning, 8*, 129–149.

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press

Snellings, P., Van Gelderen, A., & De Glopper, K. (2002). Lexical retrieval: An aspect of fluent second language production that can be enhanced. *Language Learning, 52*, 723–754.

Stein, N. L., & Trabasso, T. (1982). What's in a story: An approach to comprehension and instruction. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 213–267). Hillsdale, NJ: Erlbaum.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Torrance, M., & Jeffery, G. (1999). Writing processes and cognitive demands. In G. Rijlaarsdam & E. Esperet (Series Eds.) & M. Torrance & G. C. Jeffery (Vol. Eds.), *Studies in writing: Vol. 3. The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 1–11). Amsterdam: Amsterdam University Press.

Van Gelderen, A. (1997). Elementary students' skills in revising: Integrating quantitative and qualitative analysis. *Written Communication, 14*, 360–397.

Van Gelderen, A., & Oostdam, R. (2004). Revision of form and meaning in learning to write comprehensible text. In G. Rijlaarsdam (Series Ed.) & L. Allal, L. Chanquoi, P. Largy, & Y. Rouiller (Vol. Eds.), *Studies in writing: Vol. 13. Revision of written language: Cognitive and instructional processes* (pp. 103–123). Dordrecht: Kluwer Academic.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics, 16*, 180–205.

Zimmerman, R. (2000). L2 writing: Subprocesses, a model of formulating and empirical findings. *Learning and Instruction, 10*, 73–99.

https://doi.org/10.1017/S0142716404001092 Published online by Cambridge University Press