Fluency in second language writing: the effects of enhanced speed of lexical retrieval
Snellings, P.J.F.

Citation for published version (APA):
CHAPTER 1

GENERAL INTRODUCTION, PREVIEW AND DIRECTIONS FOR FUTURE RESEARCH

1.1 GENERAL INTRODUCTION

Writing is a language skill that develops relatively late and can be quite effortful even in one’s mother tongue. Children have to learn the requirements of this new medium and achieve automaticity in the underlying skills such as handwriting (Berninger et al., 1997; Jones & Chistensen, 1999). Fortunately, there is the possibility of transfer of skills writing has in common with speaking. Retrieving the necessary vocabulary for writing purposes is similar to the retrieval process in speaking and by the age children start writing they can already use a vast range of vocabulary in speaking. In contrast, students writing in a second language have to learn new words in a new language not only for reading but also for writing and speaking. Fluent writing and speaking requires additional skills to those needed for reading and listening. Students have to be able to use their vocabulary productively. The nature of this productive vocabulary is a matter of debate (Henriksen, 1999; Laufer, 1998). In fact, the fluency with which a word can be retrieved for productive use is argued to be a separate dimension of vocabulary knowledge (Meara, 1996). Research by Kellogg (1994) has shown that even in the mother tongue this so-called lexical retrieval can be effortful. The finding that lexical retrieval is effortful even in student’s native language suggests that the demands of this process are even larger when having to write in a second language (L2). Therefore, enhancing the process of lexical retrieval is especially important for writers in an L2.

A related issue concerns the notion of available attention to devote to a particular task. In numerous studies it has been shown that attention capacity that is available for different cognitive processes is severely limited (for a discussion of attention limits see Kellogg, 1999; Lea & Levy, 1999). Kellogg (1994; 1996) applied Baddeley’s theory of Working Memory (WM, see Baddeley, 1986) to the domain of writing, and further elaborated on the structure of WM in writing while stressing the effects of limited WM capacity on the writing process. In view of the demands other processes such as planning and revising make on writing, and the need to coordinate all the processes at the same time, high demands of lexical retrieval may result in trade-offs among other processes and ultimately in reduced text quality (Hayes, 1996; Kellogg, 1996). Kellogg’s work has inspired a whole range of studies into the relationship between WM and the skills involved in writing (for a review see Alamargot & Chanquoy, 2001). Most of these studies have been correlational or focused on showing how increased demands may negatively affect the writing process. The latter experimental studies have demonstrated a clear causal relation between WM capacity and writing processes. Effects on writing products however, have been less well established. Clearly, the effect WM capacity has on writing quality deserves further attention. In addition, an approach focusing on beneficial effects of reduced demands on WM has evident practical advantages. Because of the effort involved in lexical retrieval in an L2, the studies reported on in this thesis aimed at reducing the demands on WM by enhancing lexical retrieval skills in a second language.

So far, most studies using reaction times have been conducted in the laboratory, with a focus on testing rather than training. Moreover, in this type of research the experimenter tested students individually. For educational purposes it is important to establish whether
enhancing lexical retrieval could be achieved in a natural setting while training students as a
group in their classroom environment.

The first aim of the research we will report on in this thesis was to investigate whether
training with an experimental computerized training program called Nelson-Writing (Nelson-
W) could enhance lexical retrieval. It has been well-established that response latencies of
paired associates decrease considerably after accuracy has been achieved and that training can
speed up naming responses of letters (LaBerge & Samuels, 1974). However, because we were
interested in writing in the L2, we needed to establish that learners increase retrieval speed on
this kind of material as well.

To investigate whether the task we used for measuring lexical retrieval was valid, we
conducted an additional study involving a group of students of comparable age and ability to
the students in our training study. The process of lexical retrieval is the reverse of lexical
access; instead of recognizing the word form and deducing its meaning, students have to
retrieve the correct word form to express the meaning they have in mind. As we were
interested in writing, the requirement that writing is part of the task seemed a critical part of
our measure. Hence, we developed a new productive measure for lexical retrieval; the Written
Productive Translation Task (WPTT). Students had to write down the correct English
translation of a Dutch word as quickly as possible while paying attention to its correct
spelling. We looked at the relationship between the WPTT and another test of lexical
retrieval, a written picture-naming task. In addition, we compared the WPTT scores to those
on a test of lexical access; a lexical decision task. To check whether the WPTT measured
more than spelling speed, we also compared the WPTT scores to an Orthographic Encoding
task. As speed of Typing is no essential part of the construct of written lexical retrieval, we
filtered out variance due to speed of typing differences.

Although there has been a good deal of research into the effects on writing processes
of increased demands on WM, research into the effects on writing quality has been limited.
Previous research has suggested that there is a positive relationship between efficient lexical
retrieval and writing quality (McCutchen, 2000; McCutchen, Covill, Hoyne, & Mildes, 1994).
However, this research has focused on correlations and experimental evidence is lacking. As a
consequence, our interest was in the effect of experimentally enhanced lexical retrieval on
written production. In this study we looked at three different aspects of writing quality. First,
to establish whether our training transferred to the writing task (a cartoon story), we looked at
the number of trained words used productively in this writing task. Second, we looked at the
effect on enhanced lexical retrieval on the expression of content elements. Finally, we looked
at the effects on global writing quality.

1.2 PREVIEW

The common focus of all the chapters in this thesis is the effect of enhanced lexical retrieval
skills on writing in a second language. In the following chapters we will investigate the
effects of an experimental method for enhancing lexical retrieval (Chapter 2), establish the
validity of the measure used to assess an increase in lexical retrieval speed (Chapter 3), study
the effects of enhanced lexical retrieval on L2 writing (Chapter 4), and finally, provide a
detailed description of the computerized Nelson-Writing training program (Chapter 5). Here
we will summarize each chapter followed by recommendations for research and education.
The final chapter is a Dutch summary of this introductory chapter (Chapter 6).
Chapter 2: Lexical Retrieval: an aspect of fluent second language production that can be enhanced

As there had been no previous studies focusing on the beneficial effects of enhanced lexical retrieval on second language (L2) writing, it was essential to establish whether such an increase was possible. After developing and pre-testing the Nelson-W computer program, we examined the feasibility of this experimental computerized training for fluent lexical retrieval in the L2 in an educational context. In a true experimental counterbalanced design, 100 students from four classes were randomly assigned to one of two treatments that both involved speed training. Each class was split in half, and the resulting eight groups were trained concurrently over a 4-week period by two experimenters. Each treatment used a different set of stimuli words, 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 one group were not trained in the other group’s word set. It is important to note that differences in the two sets of words are not an issue in this research design.

The words trained were assumed to be familiar to the students. Because there are no learner corpora for productive vocabulary of Dutch L2 learners we elicited words from a similar group of students on the same writing tasks that were afterwards used to establish effects of the training. Students were trained in class over a period of 4 weeks. Total time on task was 3 hours and 20 minutes (4 classes of 50 minutes). To eliminate experimenter effects, each group trained by the same experimenter contained students from both conditions. All students received the same type of computer training and the same amount of computer-assisted exercises. The crucial difference was the set of words they were trained in; either word set A or word set B. Because our objective was to increase speed, students were encouraged to speed up in the training (cf. Van den Bosch, Van Bon, & Schreuder, 1995) and they were provided with a baseline for reference to their progress (cf. LaBerge & Samuels, 1974). Each target word was trained 10 times using four different exercises, which were sequenced from receptive to productive and from relative easy to more complex.

Students started with a word order exercise (Block), proceeded with two exercises focusing on the meaning of words (Correction and Detection) and finally they completed a productive exercise (Translation). After the training students were tested using two posttests; a lexical decision task and a newly developed test for lexical retrieval, the Written Productive Translation Task (WPTT). The WPTT aims to tap the efficiency of the lexical retrieval process in L2 writing. To make clear to the students which words they had to produce we used native language words. Once the intended concept is clear to the writer, the process of retrieving the words from memory and writing them down in context is similar to the process involved in actual writing. Because lexical retrieval for writing involves retrieval of the orthography, the answers in the WPTT had to be written down, in contrast to oral translation tasks. To correct for initial differences in speed, reaction times (RTs) on untrained words from either the lexical decision task or the WPTT were used as a covariate. The first were used to correct the RTs on the lexical decision posttest, the latter to correct the RTs on the WPTT. Two one-way MANCOVA’s were carried out, one on the lexical decision scores and another one on the scores from the WPTT. The independent variable was condition (A or B) in both analyses. The dependent variables were scores on the lexical decision A and B words in the first MANCOVA and scores on the WPTT A and B words in the second.

Results on the lexical decision task showed that the trained groups had higher correct scores and faster RTs than the group that was not trained on these words. In both cases the effects on speed were medium (Cohen, 1988). Results on the WPTT showed a similar picture, although this time the differences between the groups were larger. Again, the trained groups had higher correct scores and faster RTs than the group that was not trained on these words.
The effect sizes for the differences in speed were large. Looking at both the results of the lexical decision task and the WPTT, we first can conclude that the experimental training resulted in faster RTs on the lexical decision task. This is an important finding because the experimental training used a different format than the lexical decision task. Consequently, we may assume that the underlying process has been influenced and that the intervention managed to speed up lexical access. Following Daneman and Green (1986), who argued that processing skill differences in WM are highly task specific, we assumed that because the training focused on lexical retrieval, largest effects could be expected on the WPTT. Our findings indeed show that the differences are in the same direction as with the lexical decision task and this time the size of the effects is larger.

Before looking into the effects of enhanced lexical retrieval on L2 writing, we had to establish whether such an increase in retrieval speed could be achieved by an experimental training in a classroom setting. The results from this study have shown that such training can be successful.

Chapter 3: Validating a test of L2 written lexical retrieval: a new measure of fluency in written language production

To determine whether the WPTT is an appropriate measure for increased speed of written lexical retrieval, it is crucial to investigate its validity and to establish what meaning we can attribute to the scores generated by it. The study reported on in chapter 2 (Snellings, Gelderen, & De Glopper, 2002) demonstrated that L2 written lexical retrieval as measured by the WPTT can be effectively enhanced through a short, highly focused instruction. Responsiveness to experimental treatment provides significant support for the WPTT's validity. However, in view of the importance of this test for measuring fluency in L2 written language production, it has to be determined whether the construct of lexical retrieval the WPTT intends to measure is sufficiently different from related constructs such as lexical access. We would also like to determine whether the WPTT measures more than just the sub-process of speed of orthographic encoding.

The WPTT is similar to oral translation tasks used in psycholinguistic research (“forward translation”) in the sense that it also requires learners to translate from their native language into the L2 (De Groot, Dannenburg, & Van Hell, 1994). The difference lies both in the way the responses are given and the kind of material that has to be translated. Our aim was to measure lexical retrieval in an educational context. Therefore, the WPTT was designed to be applicable in a classroom setting. In this study we also used a written version of the oral picture naming (PN) method (Schoonen et al., 2002), a different test for lexical retrieval. In this task a picture is presented and the first letter of the word denoting the picture has to be typed as fast as possible. We used a Lexical Decision Task (LDT) to measure speed of lexical access. The LDT requires learners to determine as fast as possible whether a word presented on the screen is an existing word in the target language. In previous research, the LDT has been used (presumably for want of a better method) for measuring lexical retrieval (McCutchen et al., 1994). To measure speed of orthographic encoding, a test was developed (OE) in which students have to choose as fast as possible between two spelling alternatives of the same word in a small sentence or group of words. Finally, to measure typing speed, a Typing task was administered in which students are presented with individual letters on a laptop screen and have to type these letters as fast as possible.

In our validation of the test we adopt Messick’s (1989) “unified validity” framework to look at the validity of the test from different angles. In this view, different kinds of validity evidence are not alternatives but supplement each other in assessing the unifying concept of construct validity. Construct validity involves an integration of any evidence on what the interpretation or meaning of the test scores is for a particular purpose.
In this study 109 Dutch students took part, sampled from four classes and all in their third year of secondary education (grade 9, ages 14-15). All tests exhibited high internal consistencies and the results from the correlational analysis showed a strong association between the WPTT scores and the PN scores. In addition, we saw that the scores on the OE task are also related to the WPTT scores yet this correlation is somewhat lower. The LDT scores that tap the related construct “lexical access” are less strongly associated with the WPTT scores. The difference between the WPTT-PN correlation and the WPTT-LDT correlation is significant. The difference between the WPTT-OE and the WPTT-PN correlation is not significant and neither is that between the WPTT-OE and the WPTT-LDT correlation. The correlational analysis also showed that speed of Typing correlated substantially with all other tests. As speed of Typing is no essential part of the construct of written lexical retrieval, we filtered out variance due to speed of typing differences by using the residual scores from the regressions of the individual tests on speed of Typing in an additional correlational analysis. The correlations between these residual scores (PN', LDT' and OE') and the residual WPTT scores (WPTT') were lower than those between scores that were not corrected for typing but they still show the same pattern. However, these correlations were not statistically different.

The correlational analyses also yielded high intercorrelations between tests. Therefore, to filter out shared variance due to other factors than the construct involved and to look at the unique contribution of each test once the contribution of other test scores had been taken into account, we carried out a multiple regression analysis with residual scores. Results show that two predictors explained 41% of the variance in WPTT' scores and that there is a large and significant unique contribution of PN'. The unique contribution of OE' was also significant but rather small. This is proof that the WPTT' scores are a good representation of the construct of written lexical retrieval and involve more than getting fast access to orthographic information. Moreover, the unique contribution of LDT' was not significant. This corroborates the view that lexical access and lexical retrieval are distinct skills. The fact that the LDT' scores have no significant unique contribution to the scores on the WPTT' suggests that we should reconsider the validity of Lexical Decision Tasks as a tool for measuring lexical retrieval. Finally, there is still 59% variance that cannot be explained by the combination of scores on the PN', LDT', OE' tasks. This could indicate that the WPTT' is better able to tap the intended construct.

Recent research has indicated the importance of lexical retrieval for writing skills (Kellogg, 1999; McCutchen, 2000). The WPTT can be used to assess the fluency of this essential aspect of language production. We have demonstrated that the WPTT can effectively tap the construct of written lexical retrieval as evidenced by its strong relationship to the scores on the Picture Naming task scores. Crucially, the WPTT has the additional advantage of being able to tap aspects of the construct that Picture Naming tasks cannot elicit. Fast retrieval of collocations is an important aspect of fluent second language production (Wray & Perkins, 2000) that should be taken into account to obtain content relevance. The WPTT is not restricted to lexical items (concrete nouns and verbs) that can be captured by pictures, and it can also generate information about the retrieval speed of combinations of lexical items. In addition, the WPTT can be applied outside a research context for classroom purposes.

Chapter 4: The effect of enhanced lexical retrieval on L2 writing

Kellogg (1994) showed in an experimental study that lexical retrieval in L1 already puts considerable demands on working memory (WM). Spending too much attention on lexical retrieval may have adverse effects on the writing process because WM capacity is severely limited (for a discussion see Kellogg, 1999; Lea & Levy, 1999). Consequently, more attention for lexical retrieval may result in less attention that can be devoted to planning, monitoring or
other cognitive operations used in writing (Hayes, 1996; Kellogg, 1996). In a correlational study McCutchen (1994) showed that there is a relationship between lexical retrieval and writing, and that students who had higher scores on lexical decision tasks are also better writers. These findings suggest that training in one of the sub-processes of writing that manages to bring down the demands on WM will result in more attention for other processes and possibly better quality texts. In view of the difficulties lexical retrieval poses for learners of the L2, this study addressed the effects on L2 writing.

In this study we analyzed additional data from the same experiment that our first study was based on (see chapter 2). Consequently, 103 Dutch students from four classes were randomly assigned to one of two treatments that both involved speed training. After the training all students wrote two narratives on the basis of cartoons. Narrative A could be written using the words from word set A whereas the words in word set B could be used in writing narrative B. Because of the restricted range of the vocabulary elicited in cartoon tasks, these tasks provide a direct way of testing the effects of enhanced lexical retrieval on writing. We first analyzed the written texts to see whether the training resulted in increased use of the trained words. Next, we investigated whether trained students were better able to express the details of the story and the desired content. To examine whether enhanced lexical retrieval had an effect on global communicative quality of the text we conducted an additional exploratory analysis. The dependent variables were measured in both narratives. They are “Number of trained words used”, “Expression of content elements” and “Global text quality”.

We counted words appearing in the students’ texts that were trained in isolation or as part of a collocation. In addition, we looked at the occurrence of collocations themselves.

To define content elements that were essential for the story we used the definition of “narrative schemata” as described by Stein and Trabasso (1982). In their view, 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 enabled us to establish whether certain parts of a story had been omitted. Only content elements that occurred in the appropriate part of the story were scored. Content elements are represented by “propositions”, or predicates and their arguments (Kintsch, 1998). Content elements can be expressed by various linguistic forms. On an abstract level the content element should be the same, e.g. “they see boys” is equivalent to “they meet boys” because “meeting” implies “seeing” each other.

The writing assignments instructed students to use the correct register for their peers, to induce interest value, and finally, to be coherent in the absence of the cartoons themselves. The texts were rated on their global quality by two independent raters. They had to give a general rating of the quality of the text considering its primary trait, “how well written is this story”. Raters were instructed to concentrate on content and organization, and to consider linguistic errors only to the extent that they interfered with the understanding of the story. Besides these general instructions the raters were provided with benchmark texts as a reference (Blok, 1986).

The hypotheses were tested by two one-way MANCOVA’s on the measures we derived from the writing assignments, with Condition (A or B) as independent and Number of trained words used, Expression of content elements and Global text quality as dependent measures, and vocabulary knowledge as a covariate.

The results showed that on both sets of words, students in the trained conditions used more trained words than students that had not been trained. In both cases the effect was medium. Regarding the expression of content elements, we saw that students in one condition used more content elements and the effect of the training was again of medium size. In the other condition, the pattern of differences was similar, the trained group had a higher mean score, but this difference was not significant. Finally, results on the global quality scores
suggested higher scores in one of the two conditions but again these differences were not significant.

The finding that the trained students used more words from the trained word set implies that the effect of the training transfers to a real writing task. This finding is also relevant for theories of lexical competence. DeKeyser has argued that understanding sentences in a language quickly does not mean that one can produce them fast (DeKeyser, 2001). Henriksen (1999) provided a theoretical account of the different dimensions that play a role in the development of vocabulary. In her view, both the type and quality of the meaning representation and the level of “automaticity” may determine productive competence. The current findings have implications for a theory of productive competence. The difference in word use after the speed training shows that even in the case of familiar words and regardless of the quality of the meaning representation, retrieval speed is an important factor in attaining productive competence. Although we have demonstrated that increasing speed increases productive competence, it remains an issue for further research whether actual “automaticity” (see for a review of this concept, DeKeyser, 2001) can be trained and established. It may well be that effects are more profound in that case.

Apart from the increased use of trained words, we have demonstrated that enhanced lexical retrieval in one trained condition had a significant effect on the expression of detailed content in L2 writing. It has been hypothesized by Zimmerman (2000) that struggling for the first preliminary formulation may become unnecessary and that students do not get stuck at finding tentative formulations. As a result, more attention can be dedicated to finding a more precise and better formulation. Nevertheless, in the other trained condition we found no significant difference. This is an issue for further research.

We did not find significant effects on global writing quality. These results are 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 for lexical retrieval should result in less attention that can be devoted to planning, monitoring or other cognitive operations used in writing. On the other hand, these results offer opportunities for further research. Penningroth and Rosenberg, as well as Scardamalia et al. (Penningroth & Rosenberg, 1995; Scardamalia, Bereiter, & Goelman, 1982) have argued that the effects of reduced cognitive demands only operate at the local text level. In the section “Directions for future research and implications for education” we will elaborate on this issue.

Chapter 5: An experimental computer program to train lexical retrieval
We showed that the computerized Nelson-Writing training program (Nelson-W) was effective in enhancing both lexical retrieval and lexical access (Snellings et al., 2002, Chapter 2). In addition, the resulting increased fluency in lexical retrieval had positive effects on the use of lexical material productively, and in one trained condition also on the expression of essential content elements in writing (Snellings, Gelderen, & Glopper, submitted, Chapter 4). The program provides feedback on the basis of measures of reaction time and accuracy. So far, this type of information has been predominantly used for testing objectives and has been restricted to laboratory conditions. Because the Nelson-W program is the first training program that specifically focuses on speed of lexical retrieval, it may serve as a useful tool for future research into this area. In addition, as it has proven its potential in an educational setting it can also be adapted for educational purposes. Therefore, it seemed worthwhile to give a detailed description of the design principles underlying the Nelson-W program, the way it functioned in a classroom experimental setting and examples of exercises used in an experimental application. Here we will discuss the general design principles.
1) A focus on speed. Because our aim is to stimulate fast lexical retrieval for production purposes, it is important to encourage students to respond as quickly as possible (Van den Bosch et al., 1995). Consequently, we explicitly encourage students to answer as fast as possible while keeping correctness in mind. To achieve this, we stress attention to speed in the on screen instructions at the beginning of every exercise. In addition, feedback displays the exact speed in seconds as well as speed of previous reaction on the same item as a baseline for reference to their progress. Emphasis on speed is also obtained by complementing the numerical information with two different icons, the choice of icon depending on whether the response is faster or slower than the previous response on that same item. In addition to the speed feedback, it is also possible to restrict presentation time, as a further incentive to answer as fast as possible. This “flashcard principle” entails that the item disappears from the screen after a specified time but students can still give their answer. At the end of each exercise we show students a results screen in which both current results (Percentage correct, Average reaction time) and results from the previous sit of the same exercise (also Percentage correct and Average reaction time) are presented.

2) The importance of correctness. In the on screen instructions at the beginning of every exercise we explicitly encourage students to keep correctness in mind. In addition, the feedback uses two different icons to indicate whether the response is correct or incorrect. Whenever an answer is incorrect, information on speed is not shown on the screen, only a red cross appears to stress that correctness has to be taken into account. It is very important that whenever students make an error they carefully examine the correct answer. Therefore, before using the program, it is possible to specify the minimal number of times a student must try to give the correct answer.

3) Flexibility of stimuli and instructions. We programmed the program in such a way as to provide optimal flexibility for each individual user. In the Nelson-W program it is possible to use written stimuli, as well as pictures and sounds or a combination of these. The program further allows for easy construction and adaptation of stimuli. In the case of the translation exercise, the input is not restricted to single words and if the aim is to train collocations, combinations of words can also be entered. Because the only requirement is that stimuli are entered in a text file, most languages are possible, as long as the character set is installed on the computer used.

4) The avoidance of sequential learning. To make sure students learn the items individually and are able to use them on their own, the order in which items are trained in all exercises is randomized each time before the students commence an exercise.

5) Progression from receptive to productive learning. Under the assumption that productive knowledge is conditional upon receptive knowledge, the Nelson-W program is designed to incorporate exercises aimed at both kinds of knowledge. In our experimental application, each specific word can be trained with four different types of exercise, moving from receptive skills to the target production skills. In our approach we therefore started with a receptive exercise focusing on the correct position of the target words in the sentence (Block exercise). The focus then changed to the meaning of the items but once more using a receptive format. Students have to decide whether the target item makes sense in the context of a sentence and are either required to choose an alternative if it does not (Correction exercise), or to indicate which of two words in a sentence makes the sentence odd (Detection exercise). Finally, students practice the specific skills necessary for productive language skills. In the translation exercise students have to translate Dutch words in an English carrier sentence as fast as possible in suitable English equivalents.

6) Administrative flexibility. If an experiment involves different schools, it is possible to choose the specific school the program is going to be used at for that particular session. The
program incorporates a database so that students only have to type in their identification number.

7) The ability to monitor the individual's training processes. To be able to keep track of each student's personal progress the Nelson-W program saves a personal logfile for each student.

The Nelson-W program provides a useful tool for teachers who want to train students to use words productively in writing. Apart from being an effective training instrument, the Nelson-W program generates logfiles that provide important information about the learning process. Finally, we suggest a number of new exercises that could make the program a valuable asset to both language education and research into language production.

1.3 DIRECTIONS FOR FUTURE RESEARCH AND IMPLICATIONS FOR EDUCATION

This is the first set of studies that has explicitly focused on the effects of enhanced lexical retrieval on L2 writing quality. The results have provided new insights into the process of writing and the factors that influence writing quality. Furthermore, the Nelson-W program appears to be a promising approach for enhancing students' productive language skills by focusing them on both speed and correctness.

In the studies reported on in this thesis, we have shown that speed of lexical retrieval, an important sub-process of language production, can be experimentally enhanced. We have demonstrated increased efficiency by showing large effects on a measure of lexical retrieval, the written productive translation task (WPTT). In addition, we saw medium effects on the related skill of lexical access, as measured with a lexical decision task.

We have also examined the validity of the WPTT. We have shown that the WPTT is a reliable measure of L2 written lexical retrieval. Moreover, we have demonstrated that the association with other theoretically related tests is in accordance with the claims made in the research literature. The findings further show that instead of using lexical decision tasks as a substitute for measuring written lexical retrieval, it should be recommended to use a test specifically constructed for this aim, like the WPTT. To use the WPTT as a measure for lexical retrieval it is crucial to obtain a measure of speed of typing. As speed of typing is not part of the theoretical construct of written lexical retrieval but an artefact due to the method used, we have to stress the importance of filtering typing speed out as a correction. As the correlations derived from the regression analysis with residual scores show, failing to take these factors into account may lead to inflated correlations. Had we measured speed of writing on an electronic writing pad as Bonin (2001) did, handwriting skills would have been important instead and they should have been controlled for. Similarly, spoken Picture Naming tasks should in fact be corrected for articulatory speed, possibly by voice onset measures for meaningless sounds.

Apart from a direct effect on speed of lexical retrieval, our experimental training with the Nelson-Writing training program (Nelson-W) had a positive effect on the quality of L2 writing. After the training, students used more words that were trained in their condition in their texts than students who were not trained in these words. In addition, students in one experimental condition became significantly better in expressing the detailed content of the cartoons that elicited the narratives. In the current study we did not find significant effects on global writing quality. These results are interesting in view of the theories of both Hayes and Kellogg (Hayes, 1996; Kellogg, 1996), as these authors both predict that reducing demands on WM should have beneficial effects on global writing quality as well. Here, we will discuss
some reasons for the discrepancy between the predicted findings and the data found in the present study.

It could be argued that because the link between the trained words and the content to be expressed was not strong enough, students did not need the training to express the necessary content. However, students trained in one condition significantly outperformed the students in the other condition. The other way round there was no significant difference but the raw difference was in the same direction. Therefore, we believe that the trained words were good instruments for writing these texts. It seems more likely that because lexical retrieval is only one sub-process of writing, on a global text level other skills such as organising and revising ability obscure the differences in efficiency of lexical retrieval because they may compensate for less efficient retrieval processes. This interpretation is in accordance with previous research, as both Penningroth and Rosenberg, as well as Scardamalia et al. (Penningroth & Rosenberg, 1995; Scardamalia et al., 1982) have argued that the effects of reduced cognitive demands only operate at the local text level. The research by Scardamalia et al. (1982) suggested that when a screen was placed over the writing paper while writing, children failed to subsequently write all the words they had already formed in their minds (their "forecasts") due to memory loss. Typically, this resulted in reduced richness of content but not in lower global quality. Penningroth and Rosenberg (1995) found an effect of increased information-processing load on local coherence (connections between sentences) but not on holistic quality.

Another form of compensation could be that learners who were not trained found ways of avoiding a high cognitive load. The research by Penningroth and Rosenberg (1995) showed that even though cognitive demands were experimentally increased, the overall level of cognitive effort as measured by RT on a secondary task (pressing a key as fast as possible when a tone occurred during writing) did not increase. This was due to the fact that writers coped with a high WM load by changing the distribution in time of their writing strategies. In the first phase ("thirds" of the writing session), instead of translating writers under a high load reviewed more. In the third phase, writers under a high load translated more and reviewed less. Future research should therefore take up these matters more directly, by measuring cognitive load while writing. It is also possible to compare the effects of different types of facilitation on the reduction of cognitive load. Measuring cognitive load could be accomplished by using reaction times on a secondary task during writing. It is also important to establish individual differences in WM on a complex span measure (see for example Ransdell & Levy, 1996). It is reasonable to assume that effects may be stronger for students with a smaller span, as students with a larger span are better able to cope with the higher demands of lexical retrieval on WM capacity.

Finally, it has to be noted that even though the sub-process of lexical retrieval has become faster and more efficient, other essential processes in writing have not. In particular, experienced writers' self reports demonstrate the amount of effort even these writers have to invest in the writing process (see for a discussion Torrance & Jeffery, 1999). In addition, experienced writers have more metacognitive knowledge. Although we may assume that some of our writers (grade 9, age 14-15) possess metacognitive skills in their L1 that may be transferred and used in the L2 when the demands on WM are not too high, other writers may lack these skills so there is nothing to transfer. Although enhanced lexical retrieval may have provided the opportunity to use metacognitive skills, metacognitive skills themselves may be lacking. Future research should therefore measure the metacognitive skills in the LI as well.

In view of the learning process, the logfile data recorded by the Nelson-W program can provide us with insights into the development of lexical retrieval skills. Logfile data can tell us something about the stability of the skills involved. We could determine whether students show consistent behavior once they have learned the right responses or whether they keep on
making mistakes. If the latter is the case, it is informative to look into the kind of mistakes made. In addition, we can see for which words this is the case; which words are consistently mastered and which words show less consistent behavior. Logfile data also provide the tools to investigate the individual development in reaction times for different types of learners, as these data show the progress from each moment in time to the next. We would like to see whether there is a stable speed pattern to be found and if this is the case, whether it depends on the type of words involved for a particular student. It is obvious that the relation between RTs and correctness has to be taken into account. It is an interesting question whether speed increases only after a particular item has been answered consistently correct or whether responses increase in speed before this point. One instance of the latter could be the occurrence of spelling errors that are simply the result of carelessness. In contrast, when students are still struggling to find an adequate answer, speed may not be the main focus for these students, an assumption that was confirmed in our pilot study. Another interesting possibility is to determine whether students who start slower are able to benefit more from the training.

On the basis of the results in the present study we feel justified in carrying out further research into the effects of training lexical retrieval on both writing and speaking. In view of the importance of efficient lexical retrieval for speaking (Levelt, 1989; Poulisse, 1997), the effects of experimental training on speaking deserves further attention. Although we have demonstrated in this study that the methodology used to train written production can be implemented in a classroom setting, an approach aimed at speaking would make additional demands on implementation. For example, if the aim is to practice fluent retrieval of combination of words for speaking, speech recognition abilities should be incorporated in the program.

Because of the restricted range of the vocabulary elicited in cartoon tasks, these tasks provided a direct way of testing the effects of enhanced lexical retrieval on writing. Now that effects have been demonstrated on cartoon tasks, we can investigate the effects on other writing tasks. From a theoretical point of view it is important to know more about the relation between a training with a focus on speed and characteristics of the words trained, such as the familiarity of the word. The Nelson-W program is especially suited for this kind of research as it can also measure written lexical retrieval of words that cannot be depicted by pictures, and it can incorporate single words and collocations at the same time.

The results in this study have important implications for writing instruction. Training students in words that they are already familiar with is a laborious task and all too easily results in uninspiring drills. Therefore, we would like to stress that a focus on enhanced lexical retrieval is not a plea for a repetitive training of words through drills. Because of this, the current training does not focus on structures but on the meaning of utterances in a meaningful context. Moreover, the game features (e.g. the fighter jet plane icon in the case of a faster response) improve the attractiveness of the practice environment whereas the focus on both speed and correctness prevents students from ignoring the meaning of the words.

For research purposes we trained the same set of stimuli in each group of students. Clearly, when using this program for educational purposes the choice of stimuli is all-important. Ultimately, we need to get a better understanding of which words are most suitable for a specific training, either because they are prominent in a text type that certain writers have to use (e.g. in writing scientific papers), or because these words are applicable in more general text types and in speaking contexts. It seems also sensible to adapt the level of item difficulty to the abilities of the students. To maximize the impact of the training for practical purposes, inclusion of words used in a particular teaching method is warranted as well.

The productive use of words and the expression of essential content elements are important aspects of writing. At the moment however, it is not clear how to establish...
productive competence (Henriksen, 1999). The findings in the current study show that enhancing lexical retrieval is one approach that effectively increases production in writing. The implication of this finding is that in teaching, attention should be focused on speed of lexical retrieval as well. Simply teaching words until their meaning is known may not be sufficient. When language production is required, teaching has succeeded only when students can retrieve words with little effort. Furthermore, if we take into account that students encountered each word only 10 times over a 4-week period and the total training time was 3 hours and 20 minutes (4 classes of 50 minutes), the time on task required seems reasonable in an educational context. In view of the importance that has recently been attributed to collocations for reducing cognitive effort (Weinert, 1995), it might be useful for purposes of teaching and research to put a stronger focus on combinations of words. Nevertheless, metacognitive skills remain important as well for writing high quality texts. It may well be that training in lexical retrieval could suffice for students with sufficient metacognitive skills. In contrast, students who lack these metacognitive skills may be better served by a training that combines metacognitive skills with a training in enhanced lexical retrieval.