Chapter 4

Findings

4.1 Introduction

This chapter presents the findings of the study, that is, the results and their interpretation. The hypotheses are presented and explained and the results of the data analysis are discussed in terms of these hypotheses. There are two general hypotheses in this study, based on the assumption that both the time spent in an immersion situation as well as the quality of the immersion experience will have an effect on vocabulary size. For the sake of convenience, all the hypotheses, as presented earlier (§1.6 and §3.2), are repeated in this chapter. The results for the length of immersion hypothesis are presented first: the general hypothesis is tested in terms of two separate hypotheses which deal with receptive and productive vocabulary respectively. The results of the quality of immersion hypothesis are then presented in the same way, with the receptive vocabulary results discussed first, followed by the productive vocabulary results. Tables of results are provided.

As stated earlier (§1.5; §3.1), the research question in this study is: what is the difference in the vocabulary size of second language (L2) speakers of English who have spent varying periods in immersion, and those who have experienced different qualities of immersion environments, and how does their vocabulary size compare to that of mother-tongue speakers of English? Children at this level of schooling should have a good command of at least the first 2000 high frequency word families in English and recognise them automatically – sight vocabulary – or recognise them in context (Coady, 1997a). They should also be progressing towards a knowledge of the 3000- word families (Laufer, 1992; Nation, 1990; Nation, 1993) (§1.2). The aim of the study is to establish how length of immersion and quality of immersion impact on the vocabulary size, both receptive and productive, of second language English speakers, and the rationale for the two general hypotheses has been provided (§3.2).

4.2 Length of immersion

The first hypothesis in the study is the length of immersion hypothesis, or Hypothesis L, which predicts that length of immersion will have a positive effect on the size of the vocabulary of Grade 7 school students. Subjects are categorised as mother-tongue speakers of English (MT), early immersion (EI) or late immersion (LI) subjects (see §3.3).

4.2.1 Length of immersion hypothesis: general

The general length of immersion hypothesis is expressed as follows:

Hypothesis L: The vocabulary of EI subjects will be significantly larger than that of LI subjects, and MT subjects will have a significantly larger vocabulary than both EI and LI subjects.

This hypothesis predicts that time spent in immersion will influence vocabulary growth (see §1.6). This is expressed as a directional hypothesis, as there is good reason to believe that subjects who have experienced a longer period in immersion will have a larger vocabulary. Research findings (e.g. Daniel and Nerlich, 1998; Kickler, 1995; Wode, 1995, 1998) suggest that we can predict an increase in vocabulary size according to years spent in immersion and that EI subjects will recognise and use more words than LI subjects. MT English speakers can also be expected to have a larger vocabulary size than both the immersion groups (Hacquebord, 1994).

This general hypothesis was tested in terms of two specific hypotheses, which dealt separately with receptive vocabulary (Hypothesis LR) and productive vocabulary (Hypothesis LP). These are discussed below.

4.2.2 Length of immersion hypothesis: receptive vocabulary

Hypothesis LR is expressed thus:

Hypothesis LR: The receptive vocabulary of EI subjects will be significantly larger than that of LI subjects, and MT subjects will have a significantly larger receptive vocabulary than both EI and LI subjects.

The null version of the hypothesis is formulated thus:

HLR_0: There will be no significant difference in receptive vocabulary size between MT, EI and LI subjects.

As explained earlier (§3.4.2), Tests A and B were drawn up to test receptive vocabulary size. Although present language testers may question the value of using only discrete vocabulary measures to test the learners' overall proficiency in a second language, there is still a place for this sort of item, especially when combined with more indirect assessment through performance on tasks, as occurred in the writing exercise component of this study. The composition and construction of the tests has been described in detail in §3.4.2.

Subjects' scores on each of the sections of the tests were calculated to provide information on the size of their receptive vocabularies. The tables below reflect the mean scores and the results of the one-way ANOVAs performed on these scores. Results were considered significant at the $p \leq 0.05$ level, highly significant at the $p \leq 0.01$ level and very highly significant at the $p \leq 0.001$ level.

Table 4.1 ANOVA: Length of immersion: Vocabulary Test A

	Groups	N	Mean	*sd*
1000-word level	MT	29	8.90	2.717
	EI	55	8.00	1.587
	LI	28	7.36	2.059
	Total	112	8.07	2.104
F-value			*4.091*	
p-value			*.019**	
2000-word level	MT	29	9.03	1.636
	EI	55	8.02	1.381
	LI	28	6.93	1.864
	Total	112	8.01	1.737
F-value			*12.670*	
p-value			*.000 ****	
3000-word level	MT	29	8.07	1.557
	EI	55	6.65	2.171
	LI	28	5.71	2.034
	Total	112	6.79	2.154
F-value			*10.152*	
p-value			*.000 ****	

** = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$*

From Table 4.1 above it is clear that there was a significant effect of the length of immersion variable on vocabulary size at all three levels of the test. As was to be expected, the differences were less significant at the 1000-word level as this level represents the most basic vocabulary, and it can be fairly safely assumed that even those subjects who have spent only a few years in immersion will have mastered these words.

Post hoc Scheffé tests (see Table 4.2 below) revealed that at the 1000-word level MT scores were higher than LI scores to a significant degree. At the 2000-word level, MT scores were higher than LI scores to a very highly significant degree and also higher than EI scores, though this difference was not as significant. At this level EI scores were also significantly higher than LI scores. A similar pattern was reflected at the 3000-word level, which again showed the most significant difference to be between the scores of MT and LI. There was no significant difference between the scores of the EI and LI, however, which suggests that at the level of non-basic vocabulary (lower frequency words) differences

in vocabulary size between the immersion groups are not pronounced because neither group has yet made the transition from basic to above-basic vocabulary.

Table 4.2: Test A *post hoc* Scheffé results

Dependent variable	Years in immersion		
1000-word level	MT	EI	LI
MT		.167	.021*
EI	.167		.404
LI	.021*	.404	
2000-word level	MT	EI	LI
MT		.022*	.000***
EI	.022*		.014*
LI	.000***	.014*	
3000-word level	MT	EI	LI
MT		.010**	.000***
EI	.010**		.132
LI	.000***	.132	

* = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$

In addition, the gap between MT and EI seems to be closing at the 1000-word level. However, it does seem that EI subjects who have experienced a relatively lengthy period in immersion may be beginning to develop a basic receptive vocabulary that is commensurate with that of mother-tongue speakers of English. But it is also interesting to note the difference in vocabulary size at both the 2000- and the 3000-word levels between MT and EI. Even subjects with several years of immersion still show a considerable gap in vocabulary size at these frequency levels when compared to MT subjects. Most significant differences occurred between MT and LI groups, however. It appears from these initial results that length of immersion may be a good predictor of vocabulary size. Even at these levels of vocabulary – basic high frequency vocabulary and the beginnings of non-basic, lower frequency vocabulary – there is a significant gap in vocabulary size between MT and immersion subjects which will have to be bridged in some way if learners are to reach a level of reading vocabulary size necessary to cope with the demands of high school textbooks, which tend to be expository type texts, conceptually dense and often dealing with topics which are unfamiliar to learners (Pretorius,

2002a:189).

Although the trend that was observed in Test A was expected in Test B (see Table 4.3 below), the scores showed an anomaly: at the 2000-word level, MT, EI and LI groups all had a higher mean score than they did at the 1000- and 3000-word levels. Also, the mean scores of both EI and LI groups were higher than those of the MT group at this level. Readability of the two passages may have had some effect here: although the two passages had very similar readability scores (see §3.4.2), the passage dealing with medieval life used for Test A did contain fewer words from the 1000-word level and this may have made it slightly more difficult than the passage on the Cape Colony which was used for Test B. The content of Test A may also have been less familiar to students and thus more difficult.

Table 4.3 below reflects the results of the ANOVAs done on the scores on Test B.

Table 4.3 ANOVA: Length of immersion: Vocabulary Test B

	Groups	N	Mean	*sd*
1000-word level	MT	29	7.34	1.675
	EI	55	7.15	1.638
	LI	28	7.21	1.641
	Total	112	7.21	1.635
F-value			*.139*	
p-value			*.870*	
2000-word level	MT	29	7.93	1.981
	EI	55	8.56	2.115
	LI	28	8.18	2.480
	Total	112	8.30	2.176
F-value			*.862*	
p-value			*.425*	
3000-word level	MT	29	7.03	2.079
	EI	55	5.98	2.147
	LI	28	5.36	2.004
	Total	112	6.10	2.164
F-value			*4.733*	
p-value			*.011**	

** = $p \leq 0.05$* *** = $p \leq 0.01$* **** = $p \leq 0.001$*

As can be seen from Table 4.3 above, although Test B revealed high scores for the LI and EI subjects at the 2000-word level, these scores were not significantly higher than those of the MT subjects. The differences between scores were only significant at the 3000-word level. *Post hoc* Scheffé comparisons (see Table 4.4 below) revealed that this significant difference occurred between the scores of the MT and LI subjects. Standard deviations at the 2000- and 3000-word levels are fairly high, but similar across all three groups, which indicates that there was a wide spread among scores within groups at these levels, which was not the case at the 1000-word level.

Table 4.4: Test B *post hoc* Scheffé results

Dependent variable	Years in immersion		
1000-word level	MT	EI	LI
MT		.870	.956
EI	.870		.404
LI	.956	.404	
2000-word level	MT	EI	LI
MT		.452	.912
EI	.452		.749
LI	.912	.749	
3000-word level	MT	EI	LI
MT		.096	.012*
EI	.096		.441
LI	.012*	.441	

** = $p \leq 0.05$* *** = $p \leq 0.01$* **** = $p \leq 0.001$*

In summary, the scores on Tests A and B, despite the anomalous results in the latter at the 2000-word level, tended to reflect what had been hypothesised: there were significant differences between the MT and LI groups; the differences between EI and LI groups were not as pronounced as had been predicted, however. MT and EI groups were significantly different only in Test A at the 2000- and 3000-word levels. The gap between these groups had begun to close at the 1000-word level, the most basic level of vocabulary. A similar finding was reflected in the analysis of results of Test B. What did emerge is that some intervention will be necessary if the late immersion subjects in particular are to grow their vocabulary as they should. Research has stressed the need for automaticity in a basic

vocabulary of 2000 to 3000 word families (about 5000 lexical items) for general reading comprehension as this will allow the reader to cover 90 to 95 percent of the running words of a text (Nation, 1993). Although it must be kept in mind that language proficiency and reading, though clearly related, are 'conceptually and cognitively specific skills that develop in distinct ways and that rely on specific cognitive operations' (Pretorius, 2002a:175), below this threshold, reading strategies are ineffective and readers may find themselves reading at the frustration level (Lesiak and Bradley-Johnson, 1983). Nation (1993) even suggests that it may be necessary to actually teach this basic high frequency vocabulary, while at the same time preparing students with strategies to cope with low frequency words in context when they are encountered. Academic ability does not make up for a lack of high frequency vocabulary and even good readers will not perform well in the L2 if their vocabulary level is below the 3000 word family threshold (Laufer, 1992; Nation, 1993). Laufer (1997) argues that the number of words in a reader's lexicon is the most important lexical factor in good reading. The fact that immersion subjects show a lag in development of words in the 3000-word level is also cause for concern: a learner should be familiar with not only the 2000 high frequency words in the English language, but also with the general academic vocabulary common to many academic disciplines (Xue and Nation, 1984)(§2.3.2.1). At this level of cognitive and maturational development learners should be making the transition from basic to above-basic vocabulary (Laufer, 1992).

The third receptive vocabulary test was the Vocabulary Levels Test (VLT). The results of the one-way ANOVAs done on these scores are reflected below in Table 4.5.

Table 4.5: Length of immersion hypothesis:VLT

	Group	N	Mean	*sd*
2000-word level	MT	29	16.31	1.713
	EI	55	15.78	1.902
	LI	28	13.82	2.842
	Total	112	15.43	2.320
F-value			*11.185*	
p-value			*.000***	

3000-word level	MT	29	16.93	1.193
	EI	55	15.67	2.245
	LI	28	12.39	3.489
	Total	112	15.18	2.942
F-value			*27.204*	
p-value			*.000***	
5000-word level	MT	29	14.90	2.769
	EI	55	12.93	3.321
	LI	28	10.25	3.903
	Total	112	12.77	3.715
F-value			*13.844*	
p-value			*.000***	
UWL	MT	29	7.14	2.985
	EI	55	6.07	3.173
	LI	28	4.96	2.134
	Total	112	6.07	2.977
F-value			*4.004*	
p-value			*.021**	
10 000-word level	MT	29	7.41	4.196
	EI	55	5.36	3.346
	LI	28	4.21	2.331
	Total	112	5.61	3.547
F-value			*6.667*	
p-value			*.002***	

** = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$*

Results on the VLT revealed a steady decline in scores across all three groups as the frequency of words encountered became lower: as words became less and less familiar to subjects, so they got fewer and fewer correct answers. The scores on the last two levels, i.e. the UWL and 10 000-word level, were extremely low in comparison to scores at the first three levels (MT scores were more than 50 percent lower at the UWL level than at the 5000-word level, for instance), with many subjects simply leaving these questions unanswered and subsequently achieving very low scores. The effect of length of immersion on vocabulary size was significant at all five levels, with the first three levels showing the greatest differences. This suggests that the mother-tongue speakers have a larger receptive vocabulary than both the early and the late immersion students, but that this difference becomes less pronounced at the last two levels where all three groups perform

significantly less well. This is to be expected from Grade 7 students since at this maturational and educational level none of the groups could really be expected to perform very well on vocabulary from the low frequency levels.

Post hoc Scheffé tests (Table 4.6 below) revealed that the MT group scored higher than LI subjects at all five levels to a significant or a very highly significant degree. But the MT groups scored significantly higher than the EI group only at the 5000- and 10 000-word levels. Again, as was the case in Tests A and B, it seems that the gap between MT and EI subjects has begun to narrow at the levels which cover higher frequency words. It is worth noting, too, that as the words become less frequent, so the differences between the EI and LI subjects become less and less significant (.004, .262 and .345 respectively). This follows the trend that was evident in Tests A and B: it is only at the level of basic and just beyond-basic vocabulary (the first three levels of this test) that EI score significantly higher than LI subjects. This suggests that those learners who have experienced fairly long periods of immersion have begun to catch up to their mother-tongue counterparts as far as high frequency vocabulary is concerned, but are at more or less the same developmental stage as those who have experienced a shorter period of immersion when it comes to recognising low frequency words.

Table 4.6: VLT *post hoc* Scheffé results

Dependent variable	Years in immersion		
2000-word level	MT	EI	LI
MT		.560	.000***
EI	.560		.001**
LI	.000***	.001**	
3000-word level	MT	EI	LI
MT		.082	.000***
EI	.082		.000***
LI	.000***	.000***	
5000-word level	MT	EI	LI
MT		.041*	.000***
EI	.041*		.004**
LI	.000***	.004**	

UWL	MT	EI	LI
MT		.282	.021*
EI	.282		.262
LI	.021*	.262	
10000-word level	MT	EI	LI
MT		.034*	.002**
EI	.034*		.345
LI	.002**	.784	

** = $p \leq 0.05$* *** = $p \leq 0.01$* **** = $p \leq 0.001$*

As far as the results of the VLT are concerned, one can see that there was a significant difference in the size of the receptive vocabularies of these three groups, although the most pronounced difference was between MT English speakers and those subjects who had experienced the shortest period of immersion, in the direction that was predicted. From the results on the first three levels of the test it is clear that early immersion subjects have a highly significantly to very highly significantly larger receptive vocabulary than late immersion subjects. MT subjects scored significantly higher than late immersion subjects, as was predicted; there is a difference between mother-tongue speakers and early immersion subjects, albeit less significant, which suggests that EI subjects benefit from the long term effects of immersion and close the receptive vocabulary gap between them and MT subjects at the higher frequency levels. Scores on the lower frequency levels reveal that EI and LI subjects have very similar receptive vocabulary sizes, as measured by this test. Despite the maturational or age factor, the fact that immersion children are scoring significantly lower than their mother-tongue counterparts, particularly at the 3000- and 5000-word level, should raise the alarm, however: the words at this level and the levels above include the general academic vocabulary common to many academic disciplines (Xue and Nation, 1984) which children will encounter more frequently now in the expository type texts they will be faced with at high school.

There is clearly a need for some scaffolding and teaching of vocabulary to children at this stage of primary schooling and in cases where children have not yet mastered the basic levels of vocabulary. More emphasis on activities in the immersion classroom that encourage output are also necessary to increase receptive, as well as productive, vocabulary size. More student talk and exercises which

push output should be encouraged (Swain, 1996, 1998). Some intervention is called for to ensure that EI subjects continue to develop their vocabulary at the same rate as the MT subjects and that LI subjects catch up with the EI group.

4.2.3 Summary of results: receptive vocabulary

Results of all three receptive vocabulary measures show a similar tendency. Scores on all three tests revealed significant differences between the MT and those subjects who had spent the least time in immersion at almost all word levels, but the differences between EI and LI groups were generally not as pronounced as had been predicted. Test scores and ANOVA results support the theory that the difference in receptive vocabulary size between MT and EI groups is less pronounced at the levels of basic vocabulary. Long periods of immersion do seem to have a pronounced and valuable effect on basic receptive vocabulary size at least, and the differences between EI and LI subjects are also most pronounced at these levels. It is at the low frequency word levels that differences between MT and EI groups really become evident, however. This is a concern as one would expect that the EI group, after a longer period in immersion, would have developed a receptive vocabulary more similar in size to their MT English speaking classmates, and that there would be a more pronounced difference between EI and LI subjects. It could be that vocabulary is not consciously taught and children are not equipped with strategies to increase their vocabulary on their own. This may result in the so-called Matthew effect (§2.3.2.2). Also, EI subjects have learnt enough vocabulary to communicate quite effectively, and are not pushed to extend this any further. Perhaps more attention needs to be given to reading at this stage as this is how reading skill is improved and how vocabulary grows. In the process, language proficiency also improves (Elley, 1991; Pretorius, 2002b).

Laufer (1997:20) makes the point that reading comprehension is strongly related to vocabulary knowledge. Vocabulary has also proved a good predictor of reading success in second language studies (Laufer, 1992) and ‘a lack of reading ability functions as a barrier to effective academic performance’ (Pretorius, 2002a:187). The children in the present study were on the brink of

secondary school, where they would be faced more and more with expository rather than narrative texts and new, subject-specific vocabulary. Studies cited by Laufer (1997) have shown that an increase in vocabulary knowledge can bring about improvement in reading comprehension (§2.3.3.1). According to the Flesch Readability Scale (RES), Grade 7 students should be reading texts with a RES of 69 with relative ease, which is what the two chapters on which Tests A and B were based, scored. However, it seems that this is not the case for the immersion students in this study. As it stands, as far as their receptive vocabulary size is concerned, at low frequency word levels there is not much to differentiate EI from LI subjects. This indicates the need for some intervention if fossilisation of vocabulary size is to be prevented and vocabulary growth is to continue. Some structured intervention in the form of direct instruction of basic, high frequency vocabulary is necessary to ensure that such children develop an automatic knowledge of the most basic vocabulary and start moving towards a knowledge of the above-basic, 3000-plus word families in English (Nation, 1993).

4.2.4 Length of immersion hypothesis: productive vocabulary

Discrete point tests such as the multiple-choice test items and items in the VLT focus on the receptive or passive component of vocabulary knowledge and language ability, and not on the productive component. In order to form a better idea of the relative size of subjects' vocabulary (§3.4.4), a task-based component to test productive vocabulary was included in the tests administered.

Hypothesis LP is expressed thus:

Hypothesis LP: The productive vocabulary size of EI subjects will be significantly larger than that of LI subjects, and MT subjects will have a significantly larger productive vocabulary than both EI and LI subjects.

The null version of the hypothesis is formulated thus:

HLP_0: There is no significant difference in the productive vocabulary size of MT, EI and LI subjects.

Productive vocabulary data were analysed as follows: as each subject wrote a different number of words, the number of types used from each of the word-levels was calculated as a percentage of the total number of word types in each subject's piece of writing. The mean of these percentages was then calculated to indicate the proportion of word types that each group had used from each of the four levels:1000-, 2000-, 3000- and above-3000-word levels.

There was very little difference across the three groups between the proportion of word types from each level used by subjects in their writing, with the overwhelming majority of words used coming from the 1000-word level (see Table 4.7 below). A large proportion of words used by subjects across all groups in this study came from the 1000-word level, and this included the MT English speakers. As these children are approaching high school, where they will encounter academic texts and will be expected to write more formally, this should be a concern. These findings reflect those of Laufer (1995, in Laufer, 1998): better passive and controlled active vocabulary do not seem to be reflected in free production. Learners seem to have 'fossilised' their free vocabulary [beyond-2000-words per composition], and do not progress even when their passive and controlled active vocabulary improves. Even though the receptive results showed significant differences between groups, these were not reflected in the children's productive vocabulary. Again, it is clear that students will need to be pushed to produce more low frequency words to develop their productive vocabulary size.

Table 4.7: Length of immersion: productive vocabulary

	Groups	N	Mean	*sd*
1000-word level	MT	29	86.91	4.142
	EI	55	85.24	3.587
	LI	28	85.98	3.070
	Total	112	85.86	3.655
F-value			*2.044*	
p-value			*.134*	
2000-word level	MT	29	9.21	2.775
	EI	55	9.88	2.591
	LI	28	9.74	2.721
	Total	112	9.67	2.663
F-value			*.610*	

p-value	*.545*

3000-word level	MT	29	.85	1.148
	EI	55	.95	.932
	LI	28	.70	.906
	Total	112	.86	.982
F-value		*.565*		
p-value		*.570*		
Above 3000-word level	MT	29	3.03	1.947
	EI	55	3.93	1.939
	LI	28	3.58	1.684
	Total	112	3.61	1.902
F-value		*2.222*		
p-value		*.113*		

** = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$*

The results reflected in Table 4.7 suggest that not only immersion subjects, but MT English speakers as well, are all using more basic vocabulary than would have been expected at this stage of their schooling. There was not much differentiation between the average number of word types used by any of these groups. There was an overall similarity in percentages across all three groups, at all four word-levels. Referring again to Laufer (1994), it would appear that these subjects have not yet developed their productive skills in the target language (or even in their mother tongue, in the case of the MT group) to the point where the use of words from the basic levels (1000- and 2000-word levels) decreases, and the number of non-basic words used (from the 3000-word level and above) begins to overtake the basic vocabulary (§2.3.3.1). This may suggest that the gap between the EI and MT groups and between the EI and LI groups has begun to close by Grade 7 as far as using basic and above-basic vocabulary in a productive way is concerned. All groups used a majority of words from the 1000-word level; no subjects had reached the stage mentioned by Laufer (1994) above. According to Laufer (1995, in Laufer, 1998), the above-basic vocabulary of first year L2 university students makes up about 13 percent of their vocabulary while that of native speaking high school graduates can be as much as 23 percent. By comparison, these children, including the MT English speakers, are far behind this level and have a lot of catching up to do in the next four or five years.

Although there was a wide range in number of tokens used across the whole group (500-plus to as few as 150), there was a similarity in the proportion of word types used from each level, particularly by the EI and MT groups. This may, however, be the result of a test effect. The writing exercise was a very short test of free productive ability, included in the study mainly to determine whether the same kinds of differences that were evident between groups on receptive vocabulary would be reflected in productive vocabulary scores in the short essay-type task typical at this level of schooling. In this respect one cannot say that the trends that were evident in the receptive vocabulary scores were reflected in the productive scores. The one-way ANOVAs revealed no significant differences in the proportion of types used at any of the four levels by any of the groups, MT English speakers or immersion subjects. The small percentage of low frequency words used may also be due to the fact that this was a free writing exercise, and not a controlled one.

As far as lexical variation is concerned, I looked closely at the above-3000-word level and words used which had not come from the wording of the text used in the writing activity. In total, LI children used 121 of these types – considerably more than the MT (53) and fractionally more than the EI (117) subjects. Children were free to use whatever vocabulary they chose and were not prompted by the task to use any specific low frequency vocabulary. This meant that subjects did not have to take risks in using words they were not sure of and they were not forced to extend themselves lexically. Ultimately, the productive writing samples were too short overall and thus the results should be viewed with some caution.

The table below indicates the averages of tokens and types used by the groups in this study, and those of the groups in Kickler's (1995) analysis of written data from the Kiel study.

Table 4.8: Length of immersion: average number of types and tokens

Present study			
	N	Tokens	Types
MT	29	211.4	117.27
EI	55	298.92	145.05
LI	28	329.66	148
p-value		*0.780*	*0.039**
Kiel study (Kickler, 1995)			
	N	Tokens	Types
Group A	12	250.42	103.33
Group B (Immersion)	12	329.50	126.67
Group C	12	300.17	114.67
p-value		*0.088*	*0.059*

* = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$

Despite the fact that children in the present study had more than half an hour in which to write, they generally produced short texts, considerably shorter than the number of words produced in writing by most children in the German study, both immersion and non-immersion groups, except in the case of the LI subjects, who compare favourably with the immersion and one non-immersion group in the German study. In their analysis of spoken data from 11 to 12-year-old students in the German programme, Daniel and Nerlich (1998:649) found that the immersion group members spoke for an average of just over 15 minutes, using an average of 244 words (tokens) and 54 types each. What is more pertinent to the present study is the analysis of written data, however. The highest number of tokens written by the 12 immersion subjects Kickler (1995) reported on was 424, with the overall highest number of tokens (459) being used by a student in the control group from the school not following an immersion programme (Group C).

In the present study, the highest number of tokens used was 588, by a student from the LI group. Both immersion groups in the present study used more types than any of the groups in the German study, however, and both these groups outperformed the MT group. This underperformance on the part of the MT group may have been due to a lack of motivation and may be the effect of a general lack of interest in the activity and a lack of motivation to write. I found that I had to persist in asking some children to write at any length and many expressed reluctance to answer all the questions.

ANOVAs showed no significant differences between groups in respect of tokens in either of these studies. Researchers in the Kiel study (Daniel and Nerlich, 1998; Kickler, 1995; Wode, 1998) did, however, find nearly significant differences in the number of types used by subjects from the immersion groups ($p = 0.059$, see Kickler, 1995), despite a very limited period of immersion, and the number used by those who had experienced English only in the English-as-subject class. With regard to types, the present study showed a significant difference in numbers used ($p = 0.039$), and it is clear that MT subjects used considerably fewer types than both immersion groups.

As far as the length of immersion hypothesis and productive vocabulary is concerned therefore, the null hypothesis cannot be rejected.

4.2.5 Summary of length of immersion findings

Differences between receptive vocabulary size across the three groups were marked (as discussed in §4.2.2 and §4.2.3), with those subjects who had spent the longest period in immersion revealing a vocabulary size which was approaching comparison to that of MT speakers of English. Differences between EI and LI were most pronounced at the levels of basic vocabulary. Differences in productive vocabulary size between groups were far less pronounced, however. This reflects what research has revealed: learners generally recognise far more words than they use, and many words that are known passively or receptively never reach the productive lexicon (Laufer, 1998). According to Laufer's study with first year L2 university students, beyond-basic vocabulary made up only 12% of their vocabulary. Her findings for mother-tongue students were not very different: English speaking high school graduates revealed a beyond-basic productive vocabulary of only 23%. Passive vocabulary size did not seem to be reflected in free active production, and this is certainly borne out in the results of the present study.

4.3 Quality of immersion

The second general hypothesis in the study is the quality of immersion hypothesis, or Hypothesis Q, which predicts that the quality of the immersion environment will have an effect on the size of

vocabulary.

4.3.1 Quality of immersion hypothesis: general

The general form of the quality of immersion hypothesis, expressed non-directionally (see §3.2), is as follows:

Hypothesis Q: There will be a significant difference in the vocabulary size of MT, DI and SI subjects.

This hypothesis predicts that the quality of the language environment will have a significant effect on vocabulary size.

This hypothesis is formulated as two specific hypotheses which test assumptions about receptive vocabulary and productive vocabulary separately. Hypothesis QR deals with receptive vocabulary and Hypothesis QP with productive vocabulary. They are discussed separately below.

4.3.2 Quality of immersion hypothesis: receptive vocabulary

Hypothesis QR is expressed thus:

Hypothesis QR: There will be a significant difference in the receptive vocabulary size of MT, DI and SI subjects.

The null version of the hypothesis is expressed thus:

HQR_0: There will be no significant difference in receptive vocabulary size between MT, DI and SI subjects.

Hypothesis QR tests the prediction that there is a significant difference in receptive vocabulary size between those subjects who have experienced a different quality of immersion experience, and mother-tongue speakers of English.

As in the case of Hypothesis LR (§4.2.2), data analysis comprised computing the means at each level of the tests and doing one-way ANOVAs. As had been expected, there was a steady decline in mean scores across the three levels of Test A, with the scores for SI particularly low at the 3000-word level (see Table 4.9 below).

Table 4.9 Quality of immersion: Test A

	Groups	N	Mean	*sd*
1000-word level	MT	29	8.90	2.717
	DI	29	8.24	1.527
	SI	54	7.54	1.860
	Total	112	8.07	2.104
F-value			*4.308*	
p-value			*.016**	
2000-word level	MT	29	9.03	1.636
	DI	29	8.34	1.261
	SI	54	7.28	1.698
	Total	112	8.01	1.737
F-value			*12.535*	
p-value			*.000 ****	
3000-word level	MT	29	8.07	1.557
	DI	29	7.31	1.845
	SI	54	5.81	2.146
	Total	112	6.79	2.154
F-value			*14.237*	
p-value			*.000 ****	

** = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$*

The results of the one-way ANOVAs on the scores from Test A show that there was a significant effect of the independent variable quality of immersion on receptive vocabulary size at the 1000-word level and a very highly significant effect at the 2000- and 3000-word levels (see Table 4.9 above). *Post hoc* Scheffé comparisons (Table 4.10 below) showed that at the 1000-word level MT scores were significantly higher than SI scores; at the 2000-word level both MT and DI scores were significantly (in the case of the MT group, very highly significantly) higher than SI scores. Scores at the 3000-word level revealed the same tendency. There was no significant difference in the vocabulary size of the MT and DI groups at any level, however, and this is worth noting.

Table 4.10: Test A *post hoc* Scheffé results

Dependent variable	Years in immersion		
1000-word level	MT	DI	SI
MT		.477	.018*
DI	.477		.330
SI	.018*	.330	
2000-word level	MT	DI	SI
MT		.256	.000***
DI	.256		.016*
SI	.000***	.016*	
3000-word level	MT	DI	SI
MT		.332	.000***
DI	.332		.005**
SI	.000***	.005**	

** = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$*

As far as the results of Test A are concerned, the null hypothesis can be confidently rejected as the variable of quality of immersion has had a significant effect on vocabulary size. These results reflect the trend seen in the length of immersion results: the difference in vocabulary size between the MT group and those subjects with more intense exposure to immersion has become less distinct, particularly at the basic vocabulary levels. The most striking differences occur between native speakers of English and those subjects who had experienced shallow immersion. The results also suggest that those subjects who had not experienced as rich a target language environment, and who had had limited contact with native speakers in their school environment (School S, for instance, had

no MT subjects in the sample group) were not growing their vocabulary at the same rate as DI subjects. The significant differences between the SI and DI at the 2000- and 3000-word levels show that beyond the most basic vocabulary level the differences in vocabulary size between these two conditions does become more pronounced, suggesting that quality of immersion does have a positive effect on receptive vocabulary size.

The same trend was evident in the scores on Test B, although DI scores were higher than MT scores at the 1000-word level (see Table 4.11 below). Another surprising result was that the SI group's mean score at the 2000-word level exceeded both the other groups' scores. The ANOVA showed a very highly significant difference only at the 2000-word level, with a significant difference at the 3000-word level.

Table 4.11 Quality of immersion: Test B

	Groups	N	Mean	*sd*
1000-word level	MT	29	7.34	1.675
	DI	29	7.45	1.723
	SI	54	7.02	1.572
	Total	112	7.21	1.635
F-value			*.773*	
p-value			*.464*	
2000-word level	MT	29	7.93	1.981
	DI	29	7.28	2.218
	SI	54	9.06	2.004
	Total	112	8.30	2.176
F-value			*7.715*	
p-value			*.000* ***	
3000-word level	MT	29	7.03	2.079
	DI	29	5.83	2.494
	SI	54	5.74	1.895
	Total	112	6.10	2.164
F-value			*3.867*	
p-value			*.024**	

* = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$

Post hoc Scheffé tests (Table 4.12 below) revealed that the significant differences lay between DI and SI at the 2000-word level, but not in the direction that had been expected as SI scores were

significantly higher than DI scores. MT scores were significantly higher than SI scores at the 3000-word level. The results are not conclusive here, and the null hypothesis cannot be confidently rejected on the basis of the results of Test B.

Table 4.12: Test B *post hoc* Scheffé results

Dependent variable	Years in immersion		
1000-word level	MT	DI	SI
MT		.972	.689
DI	.972		.525
SI	.689	.525	
2000-word level	MT	DI	SI
MT		.481	.064
DI	.481		.001**
SI	.064	.001**	
3000-word level	MT	DI	SI
MT		.098	.032*
DI	.098		.984
SI	.032*	.984	

* = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$

Scores on the third test, the VLT, were computed in a similar way and ANOVAs were conducted (Table 4.13 below). Mean scores reflected the same tendency as the scores on Tests A and B, with a steady decline in scores across all three groups as the words became more infrequent, but with the expected difference in vocabulary size reflected: that is, MT scored consistently highest, with DI consistently higher than SI. The results of the one-way ANOVAs revealed significant differences at all levels of the VLT, with these being highly significant at the 3000- and 5000-word levels in particular. Again, as was reflected in the scores on Tests A and B, deep immersion seems to have a positive effect on receptive vocabulary size, to the extent that there is a less pronounced difference in scores of DI and MT subjects than is the case between MT and SI scores.

Table 4.13 Quality of immersion: VLT

	Groups	N	Mean	*sd*
2000-word level	MT	29	16.31	1.713
	DI	29	15.62	2.484
	SI	54	14.85	2.382
	Total	112	15.43	2.320
F-value			*4.077*	
p-value			*.020 **	
3000-word level	MT	29	16.93	1.193
	DI	29	15.83	2.929
	SI	54	13.89	3.039
	Total	112	15.18	2.942
F-value			*13.537*	
p-value			*000 ****	
5000-word level	MT	29	14.90	2.769
	DI	29	12.97	3.868
	SI	54	11.52	3.586
	Total	112	12.77	3.715
F-value			*8.985*	
p-value			*.000 ****	
UWL	MT	29	7.14	2.985
	DI	29	6.62	3.310
	SI	54	5.20	2.550
	Total	112	6.07	2.977
F-value			*4.984*	
p-value			*.008***	
10 000-word level	MT	29	7.41	4.196
	DI	29	5.69	3.743
	SI	54	4.59	2.610
	Total	112	5.61	3.547
F-value			*6.579*	
p-value			*.002***	

** = $p \leq 0.05$ ** = $p \leq 0.01$ *** = $p \leq 0.001$*

Post hoc Scheffé tests (Table 4.14 below) revealed that at all levels the most significant differences were between MT and SI. At the 2000-word level MT scored significantly higher than SI. At the 3000-word level both MT and DI scored highly significantly to very highly significantly higher than SI. MT scores were significantly higher than SI scores at the 5000-word level, UWL and 10 000-word level. The absence of any significant differences between the MT and DI groups at any of the five levels suggests that deep immersion may have the effect of narrowing the gap between mother-

tongue speakers of English and immersion subjects. It is also an effect of the fact that at the levels of low frequency words (5000-word level and above) the vocabulary size of MT subjects has not developed due to maturational factors. At only one point, the 3000-word level, was there a significant difference between DI and SI subjects. This is the level that differentiates the two groups: this is the point at which subjects begin to move from a basic to an above-basic vocabulary. It appears that SI subjects are not doing this at the same rate as DI subjects. Levels above the 3000-word level do not differentiate as well, as the subjects in this study were schoolchildren whose maturational level was not such that they could be expected to have a vocabulary much beyond the first 3000 words.

Table 4.14: VLT *post hoc* Scheffé results

Dependent variable	Years in immersion		
2000-word level	MT	DI	SI
MT		.511	.022*
DI	.511		.339
SI	.022*	.339	
3000-word level	MT	DI	SI
MT		.290	.000***
DI	.290		.008**
SI	.000***	.008**	
5000-word level	MT	DI	SI
MT		.111	.000***
DI	.111		.199
SI	.000***	.199	
UWL	MT	DI	SI
MT		.791	.016*
DI	.791		.106
SI	.016*	.106	
10000-word level	MT	DI	SI
MT		.157	.002**
DI	.157		.374
SI	.002**	.374	

** = $p \leq 0.05$* *** = $p \leq 0.01$* **** = $p \leq 0.001$*

The results discussed here were to be expected – immersion in a rich target language environment such as School D has apparently facilitated vocabulary growth as far as receptive vocabulary is concerned and on the basis of these results the null hypothesis can be rejected.

4.3.3 Summary of results: receptive vocabulary

The results of all three measures of receptive vocabulary suggest that shallow immersion has a less positive effect on vocabulary size than deep immersion, when compared to the scores of the benchmark group of MT English speakers. If one disregards the discrepancies between Tests A and B, the MT group scored consistently significantly or very significantly higher than the SI subjects. The DI group, on the other hand, showed no significant differences when compared to MT scores. A richer immersion environment such as that represented by School D seems to have a positive effect on receptive vocabulary size. Differences between the DI and SI groups become significant at those levels of vocabulary which differentiate basic from above-basic vocabulary, suggesting that deep immersion is more conducive to vocabulary growth than shallow immersion.

4.3.4 Quality of immersion hypothesis: productive vocabulary

Hypothesis QP is expressed thus:

Hypothesis QP: There will be a significant difference in the productive vocabulary size of MT, DI and SI subjects.

The null version of the hypothesis is expressed thus:

HQP_0: There will be no significant difference in productive vocabulary size between MT, DI and SI subjects.

The ANOVA revealed no significant differences between the groups. As already discussed in the case of length of immersion (§4.2.4), this could be partly attributed to the fact that the total number of word tokens written by each subject was very low – and contrary to what might have been expected, the MT English speakers tended to write much less than the immersion subjects**.**

Table 4.15 Quality of immersion: Productive vocabulary

	Group	N	Mean	*sd*
1000-word level	MT	29	86.91	4.142
	DI	29	85.95	3.921
	SI	54	85.24	3.131
	Total	112	85.86	3.655
F-value		*2.013*		
p-value		*0.139*		
2000-word level	MT	29	9.21	2.775
	DI	29	9.39	2.824
	SI	54	10.07	2.499
	Total	112	9.67	2.663
F-value		*1.206*		
p-value		*0.303*		
3000-word level	MT	29	0.85	1.148
	DI	29	1.04	1.137
	SI	54	0.77	0.784
	Total	112	0.86	0.982
F-value		*0.73*		
p-value		*0.484*		
Above 3000-word level	MT	29	3.03	1.947
	DI	29	3.62	2.235
	SI	54	3.92	1.628
	Total	112	3.61	1.902
F-value		*2.127*		
p-value		*0.124*		

** = $p \leq 0.05$* *** = $p \leq 0.01$* **** = $p \leq 0.001$*

The lack of any significant differences at any of the four levels means that as far as productive vocabulary is concerned, the null hypothesis cannot be rejected. As was remarked above (§4.2.4), the written vocabulary of all subjects in the study, regardless of their group, was limited, with the majority of words coming from the 1000-word level, and very little differentiation between groups. As far as an analysis of tokens and types is concerned, Table 4.16 below reflects the average numbers used by students in this study, compared to those used by subjects in Kickler's (1995) analysis of data from the Kiel study.

Table 4.16 Quality of immersion: average number of types and tokens

Present study			
	N	Tokens	Types
MT	29	211.4	117.27
DI	29	272.27	139.6
SI	54	331.14	149.51
p-value		*0.359*	*0.016***
Kiel study (Kickler, 1995)			
Group A	12	250.42	103.33
Group B (Immersion)	12	329.50	126.67
Group C	12	300.17	114.67
p-value		*0.088*	*0.059*

** = $p \leq 0.05$* *** = $p \leq 0.01$* **** = $p \leq 0.001$*

MT and DI subjects used considerably fewer tokens than the subjects in the immersion group in the German study; SI subjects used fractionally more on average. As for types, SI again used more on average than both MT and DI groups, and the averages for both immersion groups in this study were higher than those of the immersion group in the German study. ANOVAs revealed no significant differences in the German study, although there was a nearly significant difference in the number of types used ($p = 0.059$). A significant difference ($p = 0.016$) was reflected in the present study with respect to types. As for words used from the beyond-3000-word level which were not to be found in the wording of the passage, the SI subjects used considerably more (83) than the DI or MT subjects (65 and 53 respectively). The generally poor showing of the MT subjects in all respects – types, tokens and words from the above-3000 level – has been discussed above in §4.2.4.

4.3.5 Summary of quality of immersion findings

Quality of immersion does seem to have had a positive effect on vocabulary size, at least in the case of receptive vocabulary. Subjects in the SI group scored consistently lower than those in the DI group, and significantly lower than the MT subjects at all levels. As in the case of the LP hypothesis, however, effects of quality of immersion on productive vocabulary size were not significant. Laufer's (1998) theory is again borne out: many words that become part of a subject's passive or receptive repertoire never make it into productive use. Productive vocabulary seems to reach a plateau beyond

which it does not expand. The test effects could be responsible for this: subjects were not pushed to expand their output and tended to write in an informal way, with few making the effort to express themselves at any length. They may also not have taken the activity particularly seriously: I was not their teacher, and this 'wasn't for marks', which was a concern expressed by several of the children. The 3000-word level seems to be the level that differentiates DI from SI subjects. At the basic word levels there is not much difference between these groups' scores.

4.4 Two-way ANOVAs

Once the one-way ANOVAs had been conducted the next phase in the data analysis was the running of two-way ANOVAs. This was done to check for any interaction effects of the two independent variables. These results are discussed in the following section.

4.4.1 Interaction effects

In order to determine whether there was an interaction effect between the two independent variables, two-way between-subject ANOVAs were done on the scores on the tests and the writing exercise. Two-way ANOVAs are used with factorial designs, that is designs that have two or more independent variables, as is the case in the present study (§3.7). Interaction overrides main effect, which is why researchers almost always hope that main effects will turn out to be significant. They can then say that there is a clear difference that relates to the independent variable.

In the present study, the two-way ANOVAs revealed a significant interaction effect at one point only – the 5000-word level of the VLT (p = .016). The crucial corollary point is therefore that all the other tests revealed no significant interaction effect. In other words, a strong claim for the effects of the two independent variables of length and quality of immersion can be made where the one-way ANOVAs reveal significant differences, as these are not blurred by an interaction or moderating effect.

4.5 Conclusion

As far as accepting or rejecting the null hypotheses in this study is concerned, the following seems clear: in the case of receptive vocabulary, both length and quality of immersion had a significant effect on the size of vocabulary, with length exerting a marginally stronger influence. ANOVAs revealed slightly more significant effects for the length of immersion variable in general (26 instances as opposed to 24 for the quality of immersion variable), and more very highly significant differences specifically (11 as opposed to 9).

Across both variables, differences were most pronounced when comparing the receptive vocabulary size of mother-tongue speakers of English to that of subjects who had the least experience of immersion (that is, late and shallow immersion subjects). But there were also significant differences within the immersion categories, between EI and LI subjects (4) and between DI and SI subjects (4).

With regard to length of immersion hypothesis, results suggested that neither EI nor LI subjects had as yet made the transition from basic to non-basic vocabulary which Laufer (1994) postulates (§4.2.4). Results for the quality of immersion hypothesis, however, suggest that some DI subjects may have already begun to make this transition. Differences between MT and EI subjects' scores were not generally significant, and this tendency was reflected in scores for MT and DI subjects. This may suggest that the EI subjects, after a period of immersion of five or so years, and DI subjects, may have begun to build up a basic receptive vocabulary (that is, at the higher frequency levels at least: 1000- and 2000-word levels) which is almost commensurate with that of MT English speakers.

Generally, results for the quality hypothesis reflect the same tendencies evident in the length of immersion results. It seems that quality of immersion did make a difference in vocabulary size, albeit less than length of immersion, particularly when SI subjects were compared to MT subjects; differences between the SI and DI groups were not significant overall. But what is evident is that

immersion, be it length or quality, does have a significant effect on receptive vocabulary size. The longer the period of immersion and the richer the quality of the immersion environment, the larger the receptive vocabulary.

Results from the ANOVAs performed on the productive vocabulary scores were less clear cut. No significant differences between totals of word types used were revealed at any of the four levels for either the length or quality of immersion variable. As discussed above (see §4.2.5 and §4.3.5), a free writing exercise such as the one used in the present study may not have demanded that the subjects use all the vocabulary they knew: subjects were not forced to extend themselves and could have chosen safe options which they knew were correct, rather than opting for vocabulary that was risky and error-prone because it was less familiar. A controlled-active vocabulary exercise may have produced results that were more genuinely reflective of their productive vocabulary size, with regard to both the length and quality of immersion variables. Laufer (1994, 1998) showed that what she called free-controlled exercises elicited more low frequency words than free writing exercises. In free-controlled exercises learners are prompted to use particular words; free writing exercises, on the other hand, allow learners to take the 'low risk' way and avoid words they are less familiar with, the use of which could result in errors. In spite of subjects in Laufer's studies increasing their passive vocabulary and making progress in controlled active vocabulary size, they did not put this knowledge to use when choosing their own vocabulary. This finding is borne out in the present study. Immersion subjects' scores on Test A and B and the VLT reveal that they are recognising words from lower frequency levels, but they are not using these in their writing. Free active vocabulary seems to reach a 'plateau' beyond which it does not progress (Laufer, 1998: 266) and this is reflected in the results here. Corson, referring specifically to academic vocabulary, notes that although many new words may enter children's passive vocabularies as they read, only those that are encountered repeatedly will be learned in the sense that they are readily available for productive use (Corson, 1997:702).

In this regard, Swain (1998) argues that the type of processes involved in producing language are often very different from those involved in comprehending language. Because of this, students need more opportunities for 'sustained oral use of the target language' (1998:129) and she refers to

ongoing studies into Canadian immersion programmes which have revealed the efficacy of introducing the sort of tasks that encourage what she refers to as 'metatalk'(Swain, 1998:131). By this is meant the language that learners use to reflect on their own language use. Results from various studies that Swain and various colleagues looked at (Swain and Lapkin, 1995, Swain, 1995 and Tarone and Liu, 1995, in Swain, 1998) 'pointed to the potential usefulness of collaborative work in promoting output and second language learning' (Swain, 1998:138). These studies also revealed that immersion children were speaking [French] 'relatively infrequently' (Swain, 1998:128) in the immersion classroom and not much outside of it. Swain stresses the importance of language learners doing something with their language once they have completed the collaborative tasks, such as writing, and the vital role that teacher feedback plays in the process. Thus input from the teacher and the communicative context of the immersion situation is not enough on its own – there must be pushed output as well, in which children are encouraged to think about their language learning processes (§2.2.2). Swain (1998) strongly advocates the inclusion of collaborative tasks in the immersion curriculum.

It can be claimed, then, that neither length nor quality of immersion had any significant effect on productive vocabulary size. As has been posited above, immersion subjects in particular will need to be made to focus on pushing their output beyond the more familiar, high frequency words if they are to reach a level of vocabulary knowledge at which they can cope adequately with the type of writing demanded at high school.

Subjects wrote generally very low proportions of words from any of the three levels above the 1000-word level. Those words that were on the lists of above-3000 words are worth noting though, and may indicate something of their input, a chief source of which, next to school content subjects and teacher talk, is probably television. Words such as 'ambulance', 'dehydrated', 'clinic', 'cholera', 'sanitation', 'environment', 'inhabitants' and so on may be words they have learnt from content subjects and from watching the type of reality programmes that are very popular today. A more formal assignment may have stimulated the use of a wider range of vocabulary by some of the subjects, particularly the early immersion, deep immersion and English MT subjects. The present

study focused primarily on vocabulary size: the Kiel studies, (e.g. Daniel and Nerlich, 1998; Kickler, 1995; Wode, 1998) on the other hand, looked at size but also at individual variation within groups, lexical sources and errors, distribution of lexical items according to word classes as well as lexical fields and sense relations. They found that the immersion children's vocabulary was more diversified, in that they used more words that were not used by any of the other groups, and also used more synonyms. They also used more words on average than the other groups, producing significantly higher numbers of both types and tokens. The results from the present study show that there was little difference between groups, immersion and English mother tongue. However, as far as variation of vocabulary and words used which had not come from the text was concerned, immersion subjects actually used more of these types than did MT English speakers.

Although the null hypotheses cannot be rejected in all cases, there is strong evidence to suggest that both length and quality of immersion do lead to an increase in vocabulary size, at least as far as receptive vocabulary is concerned. In the case of both these variables, there was a pronounced difference between the MT speakers of English and subjects with least exposure to immersion. Those subjects who had been in immersion for longer periods, or who had experienced deep immersion, showed evidence of a receptive vocabulary size, at the basic levels at least, commensurate with that of their MT English classmates. This study in part confirms findings from the Kiel studies (e.g. Daniel and Nerlich, 1998; Kickler, 1995; Wode, 1998) that time in an immersion class, although regardless of the quality of that immersion situation (those studies dealt with classes that were essentially foreign language classrooms and teachers who were not mother-tongue speakers of English) can produce positive results, although it is only the receptive data in the present study which really supports this. In other words, subjects drew on additional input sources. Daniel and Nerlich (1998) concluded that immersion students in general could be assumed to have a larger vocabulary than non-immersion students studying the target language as a school subject. The present study supports Wode's claim that 'IM [immersion] creates better opportunities for students to activate their language-learning abilities than any other teaching methodology today' (1999:256), while at the same time highlighting the need for a focus in the classroom on activities designed to develop vocabulary size, both receptive and productive.