The Relationship among Vocabulary Knowledge, Academic Achievement and the Lexical Richness in Writing in Swedish University Students of English

Zakaria Lemmouh
The Relationship among Vocabulary Knowledge, Academic Achievement and the Lexical Richness in Writing in Swedish University Students of English

Zakaria Lemmouh
Abstract
The main aims of the thesis are: to explore the development of Swedish university students’ vocabulary knowledge, size and depth and vocabulary use (i.e. lexical richness) in their written output, to examine the relationship between these on the one hand and their relationship to examination grades on the other, and to investigate how these relationships develop over the course of two terms of university studies.

The results showed that over one year of university studies stronger links between the two dimensions of vocabulary knowledge, size and depth are established. No relationship was found between informants’ vocabulary size and the lexical richness of their written academic essays. However, a modest relationship was found between the informants’ depth of vocabulary knowledge and lexical richness of their essays.

Furthermore, there was a modest relationship between vocabulary knowledge and academic performance. A weaker, albeit significant, relationship was found between lexical richness of student essays and academic performance as reflected in the course grade. However, the study did not show evidence of a relationship between lexical richness of student essays and essay grade, which seems to indicate that lexical richness, is not an essential criterion in teachers’ assessment of student essays.

In regard to the development of the informant’s vocabulary knowledge, there was a significant growth in their productive size and depth of vocabulary knowledge after both one and two terms. The informants’ receptive size was found to only develop over two terms of study. Moreover, they produced lexically richer essays in their second term than in their first term of study.

The results of the study are discussed in light of the Swedish university learning context, such as the effect of similar learning experience at university on learners’ vocabulary ability and the onset vocabulary ability of the informants. Moreover, the findings are discussed from the perspective of pedagogical implications and vocabulary assessment.

Keywords: vocabulary knowledge, depth, receptive size, productive size, Swedish university students, lexical richness, take-home essays, course grades, learners

©Zakaria Lemmouh, Stockholm 2010
Printed in Sweden by Universitetsservice AB, Stockholm 2010
Distributor: Department of English, Stockholm University
To my beloved sons:
Ilyas and Adam
Contents

Contents ................................................................................................................... vii
List of figures ............................................................................................................ xiii
List of tables ............................................................................................................... xv
Abbreviations .......................................................................................................... xix
Acknowledgements ................................................................................................. xx

1. Introduction ........................................................................................................... 1
   1.1 Introduction ...................................................................................................... 1
   1.2 Aims of this study and main research questions ........................................... 3
   1.3 Outline ............................................................................................................. 4

2. Conceptual framework ......................................................................................... 5
   2.1 The distinction between vocabulary knowledge and use ................................. 5
   2.2 Vocabulary knowledge .................................................................................. 7
   2.3 The receptive/productive distinction ............................................................... 7
   2.4 Size of vocabulary knowledge ....................................................................... 11
      2.4.1 Issues involved in and rationales for testing learners’ vocabulary size ............ 12
         2.4.1.1 Units of counting vocabulary .............................................................. 12
         2.4.1.2 Frequency levels ............................................................................. 15
      2.4.3 Testing vocabulary size ............................................................................ 16
         2.4.3.1 Tests of size of receptive vocabulary knowledge ............................. 17
            2.4.3.1.1 Y/N vocabulary size tests ........................................................... 17
            2.4.3.1.2 The Receptive Vocabulary Levels Test ........................................ 18
            2.4.3.1.3 The Vocabulary Size Test ............................................................ 19
         2.4.3.2 Tests of size of productive vocabulary knowledge ......................... 20
            2.4.3.2.1 The Productive Vocabulary Levels Test ................................. 20
            2.4.3.2.2 The Lex30 test of productive vocabulary .................................... 21
      2.5 Depth of vocabulary knowledge ..................................................................... 22
         2.5.1 Essentialist approaches to defining depth .............................................. 25
            2.5.1.1 The network knowledge approach ............................................... 25
         2.5.2 Comprehensive approaches to defining depth ....................................... 27
            2.5.2.1 Components approach .................................................................. 27
2.6 Models of vocabulary knowledge and ability ...........................................29
  2.6.1 Henriksen’s model of lexical competence .......................................30
  2.6.2 Daller et al.’s Lexical space ..........................................................31
  2.6.3 Chapelle’s model of vocabulary ability ...........................................32
    2.6.3.1 The context of vocabulary use ...............................................33
    2.6.3.2 Vocabulary knowledge and fundamental processes ...................33
    2.6.3.3 Metacognitive strategies for vocabulary use ............................34
  2.6.4 Summary ......................................................................................34
2.7 Measures of vocabulary use ...............................................................35
  2.7.1 Measures of learners’ communicative vocabulary use .......................36
    2.7.1.1 Word-list-free approaches to measuring lexical richness ..........36
    2.7.1.2 Word-list approaches to measuring lexical richness .................37
      2.7.1.2.1 The P_Lex ........................................................................37
      2.7.1.2.2 The Lexical frequency profile /Beyond 2000 measure .........38
3. Previous work .....................................................................................40
  3.1 Learners’ size and depth of vocabulary knowledge ...............................40
    3.1.1 Relationship between size and depth of vocabulary knowledge .......40
  3.2 The relationship between vocabulary knowledge and vocabulary use in
      writing ...............................................................................................47
    3.2.1 The relationship of vocabulary knowledge to the lexical richness of
      learners’ written output .....................................................................47
  3.3 The relationship of learners’ vocabulary knowledge and use to L2
      academic achievement? .....................................................................49
    3.3.1 The assessment of learners’ size of vocabulary knowledge in
      relation to vocabulary goals and thresholds .......................................50
      3.3.1.1 Learners’ size of receptive vocabulary knowledge ..................50
      3.3.1.2 Learners’ size of productive vocabulary knowledge ................56
      3.3.1.3 Studies on the vocabulary size needed for different tasks .......58
    3.3.2 Depth of vocabulary knowledge and academic achievement .........62
    3.3.3 The relationship of lexical richness to text quality and academic
      achievement ....................................................................................64
  3.4 The development of learners’ vocabulary knowledge and use ..............68
    3.4.1 The development of vocabulary size over time ...............................69
    3.4.2 Studies on the development of learners’ depth of vocabulary
      knowledge .......................................................................................76
    3.4.3 The development of lexical richness in learners’ written output .......78
4. Method .................................................................................................81
  4.1 Statistical concepts ...........................................................................81
    4.1.1 Test usefulness ............................................................................81
    4.1.2 Statistical analyses .......................................................................83
      4.1.2.1 Correlation analysis ...............................................................83
      4.1.2.2 T-test analysis .......................................................................84
5.2.5.1 The end of the first term of study ........................................... 125
5.2.5.2 The end of the second term of study ....................................... 126

5.3. Step 3: The relationship between the components of depth of vocabulary knowledge ......................................................... 127

5.3.1 To what extent do the components of DEPTH correlate with each other? .................................................................................. 127
  5.3.1.1 The beginning of the first term of study .................................. 127
  5.3.1.2 The end of the first term of study ....................................... 128
  5.3.1.3 The end of the second term of study ................................... 129

5.3.2 To what extent do COLL, WD and SYN, respectively, explain the variance in WD+SYN, COLL+SYN and COLL+WD, respectively? .... 129
  5.3.2.1 The beginning of the first term of study ............................... 129
  5.3.2.2 The end of the first term of study ..................................... 130
  5.3.2.3 The end of the second term of study .................................. 131

5.4 Step 4: The relationship of RVLT, PVLT, DEPTH and B2000 to academic achievement and essay grade ................................................. 132

5.4.1 To what extent do RVLT, PVLT, and DEPTH correlate with GRADE?
  ............................................................................................................. 132
  5.4.1.2 End of term one ................................................................. 132
  5.4.1.3 End of term two ............................................................... 133

5.4.2 To what extent does B2000 correlate with GRADE? ................ 135
  5.4.2.1 Term one ................................................................. 135
  5.4.2.2 Term two ................................................................. 135

5.4.3 How much emphasis do teachers at the English department at Stockholm University report placing on vocabulary features when grading student essays in the first and second term of study? ........ 136

5.5 Step 5: The development of vocabulary knowledge and lexical richness ...................................................................................... 139

5.5.1 Is there a difference between the mean score on RVLT at the beginning of term one, end of term one and end of term two? .......... 139
5.5.2 Is there a difference between the mean score on PVLT at the beginning of term one, end of term one and end of term two? .... 140
5.5.3. What is the estimated overall and frequency band vocabulary size at the beginning of term one, end of term one and end of term two? .. 141
5.5.4 Is there a difference between the mean score on DEPTH at the beginning of term one, end of term one and end of term two? .... 145
5.5.5 Is there a difference in the mean-B2000 score of all essays between term one and two? ............................................................. 147

6. Summary and discussion of results ............................................. 148

6.1 Summary .................................................................................. 148

6.1.1. To what extent are Swedish students’ receptive and productive vocabulary size and depth of vocabulary knowledge related? .................. 148
6.1.2 To what extent is Swedish students’ vocabulary knowledge related to lexical richness in academic essays? ................................................................. 150
6.1.3 To what extent is vocabulary knowledge related to academic achievement? .............................................................................................................. 151
6.1.4 To what extent is lexical richness of student essays related to essay grade and overall academic achievement? ................................................................ 151
6.1.5 How do Swedish students’ vocabulary knowledge and student-essay lexical richness develop over time? ................................................................. 152
6.2 The development of vocabulary knowledge and use among Swedish university students of English ....................................................................................... 153
  6.2.1 The state and development of receptive size ........................................ 153
    6.2.1.1 The students’ state of receptive size ..................................... 153
    6.2.1.2 The development of the students’ receptive size ................. 156
  6.2.2 The state and development of productive size .................................. 159
    6.2.2.1 The students’ state of productive size .................................. 159
    6.2.2.2 The development of the students’ productive size ............... 160
  6.2.3 The state and development of productive depth ............................... 162
    6.2.3.1 The students’ state of productive depth .............................. 162
    6.2.3.2 The development of the students’ productive depth .......... 165
  6.2.4 Vocabulary use: the lexical frequency profile of students’ take-home essays ............................................................................................................. 167
    6.2.4.1 The development of lexical richness .................................... 168
  6.2.5 Final remarks on the development of the students’ vocabulary knowledge and use ...................................................................................................... 169
6.3 The relationship between dimensions of vocabulary knowledge .......... 170
  6.3.1 The relationship between receptive and productive size .................. 170
  6.3.2 The relationship between size and depth ....................................... 170
  6.3.3 The interrelationship of components of depth ................................ 173
  6.3.4 The relationship of vocabulary knowledge to lexical richness ....... 174
  6.3.5 The relationship of vocabulary knowledge to course grade .......... 176
  6.3.6 The relationship of lexical richness of student essays to essay and course grades ...................................................................................................... 179
6.4 The development of the relationship between dimensions of vocabulary knowledge and use .......................................................................................... 180
  6.4.1 The development of the relationship between receptive and productive size ............................................................................................................. 180
  6.4.2 The development of the relationship between size and depth .......... 182
  6.4.3 The development of the interrelationship of the components of depth ................................................................................................................ 183
  6.4.4 The development of the relationship of vocabulary knowledge to the lexical richness of student essays ................................................................ 184
  6.4.5 The development of the relationship of vocabulary knowledge and lexical richness to course grade ................................................................. 185
7. Conclusion .............................................................................................................. 186
  7.1 Limitations of the study ..................................................................................... 188
  7.2 Directions for further studies ........................................................................... 190

References .................................................................................................................. 191

Appendices .................................................................................................................. 202
  Appendix 1: Productive Vocabulary Levels Test, Version 1 ......................... 202
  Appendix 2: Productive Vocabulary Levels Test, Version 2 ......................... 208
  Appendix 3: Depth of Vocabulary Knowledge Test, Version 1 .................... 214
  Appendix 4: Depth of Vocabulary Knowledge Test, Version 2 .................... 222
  Appendix 5: Teacher Assessment Procedures Questionnaire ..................... 230
List of figures

**Figure 1.** Schmitt’s (forthcoming) suggested terms for Laufer and Goldstein’s (2004) four-way categorization of different degrees of word knowledge.

**Figure 2.** An illustration of Read’s (2004) and Schmitt’s (forthcoming) classification of the different ways depth of vocabulary has been conceptualized in the research literature.

**Figure 3.** The lexical space: dimensions of word knowledge (from Daller et al. 2007: 8).

**Figure 4.** Age ranges for the informants (N = 34).

**Figure 5.** The age range when the informants started studying English in school (N = 34).

**Figure 6.** (PILOT): The average receptive vocabulary size at the four levels (3K, AWL, 5K and 10K) at the beginning of term and end of term one (N = 16). White area = no. of words not known; Blue area = no. of words known.

**Figure 7.** (PILOT): The average productive vocabulary size at the five levels (2K, 3K, AWL, 5K and 10K) at the beginning of term and end of term one (N = 16). White area = no. of words not known; Blue area = no. of words known.

**Figure 8.** Answers concerning which of the five types of lexical features are focused on in the assessment of student essays.

**Figure 9.** The average receptive vocabulary size at the five levels (2K, 3K, AWL, 5K and 10K) at the beginning of term one (N = 34), the end of term one (N = 34) and the end of term two (N = 16). White area = no. of words not known; Blue area = no. of words known.

**Figure 10.** The average productive vocabulary size at the five levels (2K, 3K, AWL, 5K and 10K) at the beginning of term one (N = 34), the end of term one (N = 34) and the end of
term two (N = 16). White area = no. of words not known; Blue area = no. of words known.

Figure 11. Intervening variables between a learners’ vocabulary knowledge and the lexical richness of the output.

Figure 12. Overlay scatterplot of the relationship between receptive and productive size at the beginning and end of term one.
List of tables

Table 1. Bauer and Nation’s (1993) word family levels.
Table 2. Nation’s (1990; 2001) list of depth of vocabulary knowledge components.
Table 3. Estimated vocabulary size of EFL learners.
Table 4. Productive vocabulary sizes of second language learners in Zimmerman (2004).
Table 5. Vocabulary size thresholds for different language tasks.
Table 6. The increase of learners’ vocabulary size.
Table 7. Range of values for interpreting alpha coefficients.
Table 8. Range of values for interpreting the strength of correlations.
Table 9. Number of items swapped between the four original versions of the PVLT.
Table 10. Extract of the first version of the depth of vocabulary knowledge test.
Table 11. Answers concerning amount of time allotted to completing the depth of vocabulary test.
Table 12. Mean score and standard deviation scores on collocations test part.
Table 13. The sequence of test administration at the beginning of the informants’ first term (Fall 06: N = 17; Spring 07: N = 17).
Table 14. The sequence of test administration at the end of the informants’ first term (Fall 06: N = 17; Spring 07: N = 17).
Table 15. The sequence of test administration at the end of the informants’ second term (Spring 07: N = 6; Fall 07: N = 10).
Table 16. Test specification of the discrete vocabulary tests: RVLT, PVLT and DEPTH.

Table 17. The number of essays handed in and graded during the informants’ first term of study.

Table 18. A section of the second pilot test format based on Nation’s (1990; 2001) components framework.

Table 19. (PILOT): Descriptive statistics of the pilot discrete vocabulary test scores: Means, standard deviations (SD) and score ranges.


Table 21. (PILOT): Changes in RVLT and PVLT after one term (N = 16).

Table 22. Overview of the variables of the study.

Table 23. Descriptive statistics of the main study discrete vocabulary test scores: Means, standard deviations (SD) and scores ranges of tests administered at the beginning of term one (both groups 06 and 07).

Table 24. Cronbach’s Alpha reliability coefficients for PVLT and DEPTH administered at the beginning of term one.

Table 25. Beginning-of-term-one Pearson correlations among RVLT, PVLT, DEPTH.

Table 26. End-of-term-one Pearson correlations among RVLT, PVLT, DEPTH and B2000ALL, -LIT and -LING.

Table 27. End-of-term-one regression results with PVLT and RVLT as the predictor variables and DEPTH as the dependent variable (N = 34).

Table 28. End-of-term-two Pearson correlations among RVLT, PVLT, DEPTH and B2000ALL, -LIT and -LING.

Table 29. Beginning-of-term-one Pearson correlations among COLL, WD, SYN, RVLT and PVLT.


Table 32. *End-of-term-one regression results with DEPTH as the predictor variable and B2000ALL (N = 31) and -LIT (N = 34) as the dependent variables.*

Table 33. *End-of-term-one regression results with WD as the predictor variable and B2000ALL (N = 31) and B2000LIT (N = 34 ) as the dependent variables.*

Table 34. *End-of-term-one regression results with SYN as the predictor variable and B2000LING as the dependent variable (N = 31).*

Table 35. *End-of-term-two regression results with SYN as the predictor variable and B2000ALL (N = 11) and B2000LIT (N = 13) as the dependent variables.*

Table 36. *Beginning-of-term-one Pearson correlations among COLL, WD and SYN.*

Table 37. *End-of-term-one Pearson correlations among COLL, WD and SYN (N = 34).*

Table 38. *End-of-term-two Pearson correlations among COLL, WD and SYN (N = 16).*

Table 39. *Beginning-of-term-one regression results with COLL, WD and SYN as predictor variables and SYN + WD, COLL + SYN and WD + COLL as dependent variables (N = 32).*

Table 40. *End-of-term-one regression results with COLL, WD and SYN as predictor variables and SYN + WD, COLL + SYN and WD + COLL as dependent variables (N = 34).*

Table 41. *End-of-term-two regression results with COLL, WD and SYN as predictor variables and SYN + WD, COLL + SYN and WD + COLL as dependent variables at the end of term 2 (N = 16).*

Table 42. *End-of-term-one Spearman correlations between the independent variables RVLT, PVLT and DEPTH and the dependent variable CGRADE (N = 34).*

Table 43. *End-of-term-one significant Spearman correlations between the levels and course grade (N = 34).*
Table 44. End-of-term-two Spearman correlations between the independent variables RVLT, PVLT and DEPTH and the dependent variable CGRADE (N = 16).

Table 45. End-of-term-two significant Spearman correlations between the levels and course grade (N = 16).

Table 46. Term-one Spearman correlations between the independent variables B2000 CL1, -CL2, -CL3, -CL4. -LING and -ALL and the dependent variables EGRADE and CGRADE.

Table 47. Term-two Spearman correlations between the independent variables B2000LIT1, -LIT2, -BESS and -ALL and the dependent variables EGRADE and CGRADE.

Table 48. Change in RVLT after one term (N = 34).

Table 49. Change in RVLT after one and two terms (N = 16).

Table 50. Change in PVLT after one term (N = 34).

Table 51. Change in PVLT after one and two terms (N = 16).

Table 52. Proportion of students reaching the cut off point for mastery of the RVLT levels at the beginning of term one, the end of term one and the end of term two.

Table 53. Proportion of students reaching the cut off point for mastery of the PVLT levels at the beginning of term one, end of term one and end of term two.

Table 54. Change in DEPTH after one term among the informants who were administered the second version at the beginning of term one.

Table 55. Change in DEPTH between the beginning of term one and end of term two (N = 16).

Table 56. Change in B2000ALL after two terms (N = 13).

Table 57. Levels of the RVLT and PVLT exhibiting a significant correlation with course grade at different points in time.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2</td>
<td>Second language</td>
</tr>
<tr>
<td>SLVA</td>
<td>Second language vocabulary acquisition</td>
</tr>
<tr>
<td>RVLT</td>
<td>The receptive Vocabulary Levels Test</td>
</tr>
<tr>
<td>LFP</td>
<td>The Lexical Frequency Profile</td>
</tr>
<tr>
<td>P</td>
<td>Productive knowledge</td>
</tr>
<tr>
<td>R</td>
<td>Receptive knowledge</td>
</tr>
<tr>
<td>Y/N test</td>
<td>The Yes/No vocabulary size test format</td>
</tr>
<tr>
<td>GSL</td>
<td>The General Service List</td>
</tr>
<tr>
<td>UWL</td>
<td>The University Word List</td>
</tr>
<tr>
<td>AWL</td>
<td>The Academic Word List</td>
</tr>
<tr>
<td>TTR</td>
<td>Type/token ratio</td>
</tr>
<tr>
<td>B2000</td>
<td>The beyond 2000 most frequent words</td>
</tr>
<tr>
<td>PVLT</td>
<td>The productive Vocabulary Levels Test</td>
</tr>
<tr>
<td>VLT</td>
<td>Vocabulary Levels Test</td>
</tr>
<tr>
<td>VST</td>
<td>The Vocabulary Size Test</td>
</tr>
<tr>
<td>TESL</td>
<td>Teachers of English as a second language</td>
</tr>
<tr>
<td>EAP</td>
<td>English for academic purposes</td>
</tr>
<tr>
<td>ESP</td>
<td>English for specific purposes</td>
</tr>
</tbody>
</table>
Acknowledgements

My heart is filled with gratitude to God who has blessed me in so many ways among other things by giving me the chance to meet so many people who have supported me during the years I have been working on this thesis. Among them are my fellow doctoral students and the staff at the English Department.

I would like to express my deepest gratitude to my two supervisors Professor Philip Shaw and Associate Professor Britt Ėrman without whose advice, encouragement and feedback this thesis would not have been possible.

I would also like to express my deep gratitude to Lars Stenius Staehr whose advice and work has been a great source of inspiration. I am indebted to Norbert Schmitt who generously allowed me to see a copy of his forthcoming book. I would also like to thank Per Näsman, who advised me on statistical matters.

I would like to thank my two beloved sons, Ilyas and Adam and my wife who have been there for me during all the years I have been studying. May God bless all those who helped me in any respect during the course of my doctoral studies.
1. Introduction

1.1 Introduction

The importance of vocabulary knowledge in mastering a second language is by now well established in second language research literature (e.g. Alderson 2005; Coady and Huckin 1997; Milton 2009). This awareness has been growing in tandem with the realization that vocabulary knowledge involves more than simply knowing the meaning of words. Vocabulary knowledge is described in the research literature as a rich and complex construct (e.g. Daller, Milton and Treffers-Daller 2007; Read 2000; Schmitt forthcoming).

In order to shed light on learners’ vocabulary knowledge an important research area is to look at how different dimensions of vocabulary knowledge are related to each other. One common approach is to focus on only one or two dimensions of vocabulary ability (e.g. Abkarian 2010; Cobb and Horst 1999; Lessard-Clouston 2006; Milton and Meara 1995; Webb 2008). Although this approach is a necessary one, Schmitt (forthcoming: 166) stresses that “[w]e have seen that vocabulary knowledge is a complex construct, and that any single measure of it will give only a very minimal impression of the overall lexical knowledge constellation”.

In addition to the necessity of adopting a more holistic approach when examining vocabulary ability, we also need to examine learners’ vocabulary learning in relation to authentic learning contexts as opposed to controlled contexts where learners are subjected to a treatment or administered a language use task only as part of a study. As highlighted by Meara (1999: 565), “There is a serious shortage of good research that has looked at the behavior of real language learners acquiring vocabularies over a long timescale.”

More and more people are exposed to English and as the number of people using English in different domains increases, English will acquire the status of a basic skill which must be mastered for functioning in the global society (Graddol 2006). In the field of second language vocabulary acquisition there has been little attention to advanced learners with the extensive exposure to English via visual and digital media typical of many in the modern world. The Swedish context encompasses the type of learners

---

1 Vocabulary knowledge is used in the current study to denote a declarative knowledge dimension within the wider construct of vocabulary ability. The constructs of vocabulary ability and vocabulary knowledge are discussed in more detail in chapter two.
whose extensive exposure to English results in high proficiency. When entering university studies of English, students have already attained a comparatively high level of English language ability and it is of interest to investigate how their vocabulary knowledge and use develops in a formal language learning setting without particular focus on vocabulary. Moreover, although English is the principal foreign language taught in Sweden, we know relatively little about what milestones in L2 English vocabulary learning Swedish learners pass through at different stages of education. For instance, we do not know much about how much vocabulary a Swedish post-upper secondary learner of English entering university studies of English can be expected to know. This kind of knowledge is useful for informing teaching practices and evaluating the quality of English language teaching not only in Sweden but in the increasing number of countries where English is part of the local language ecology.

In order for researchers and teachers to be able to make informed decisions on which facets of vocabulary knowledge should be measured for obtaining an overall picture of learners’ vocabulary knowledge, they need to have an idea of how the different dimensions of vocabulary knowledge are interrelated among the specific type of learners they want to assess. Second language acquisition is characterized by a high degree of variability in the learning outcome (Ellis 1995). In contrast, among L1 learners we do not find the same inter- and intra-learner variation and if we obtain a picture of their receptive size, we can typically predict their knowledge of other dimensions of vocabulary as well (Romaine 2003).

In the Swedish context learners of English come to their university studies equipped with a varied vocabulary knowledge profile, due to different English learning experience at school. Their receptive vocabulary size might not tell us very much about other knowledge dimensions, but we would probably need to use multiple measures for obtaining a picture of the overall state of their vocabulary knowledge. By shedding light on how different dimensions of vocabulary knowledge are related among Swedish students of English, we can obtain an idea of the smallest and most practical set of predictors of these specific learners’ overall state of vocabulary knowledge. In the current study this will be achieved by looking at the relationship both statically and longitudinally between receptive size, productive size, productive depth and lexical richness of student essays among Swedish university students of English.

Although researchers within the field of second language vocabulary research have highlighted the importance of vocabulary to language proficiency, many Swedish schools and universities have yet to introduce particular vocabulary learning components.

There is a great deal of indirect data showing that vocabulary knowledge and use is strongly related to academic achievement (Nassaji 2006; Qian 1999; 2002; Staehr-Jensen 2005). However, for the Swedish context there
are no studies that have investigated directly the relationship of vocabulary knowledge and use to academic achievement. In order to be able to influence teachers to focus more on vocabulary learning, we need to show how vocabulary knowledge and use is related to academic achievement. Here, this will be achieved by looking at the relationship of the above mentioned facets of vocabulary knowledge and use to course and essay grades in the context of a university course of English Language and Literature.

1.2 Aims of this study and main research questions

The main aims of the study are to empirically examine: a) the development of vocabulary knowledge and vocabulary use in take-home essays among Swedish university students of English over two terms of university studies and then the relationship both statically and longitudinally b) between size and productive depth of vocabulary knowledge, c) between vocabulary knowledge and vocabulary use in academic take-home essays, d) of vocabulary knowledge and use (i.e. lexical richness) to academic achievement.

The following five main research questions are employed to shed light on the four areas of interest:

1. What is the relationship between Swedish university students’ receptive and productive vocabulary size on the one hand and the relationship of these two to productive depth of vocabulary knowledge on the other?
2. What is the relationship of vocabulary knowledge to lexical richness of Swedish university students’ academic essays?
3. To what extent is vocabulary knowledge related to academic achievement?
4. To what extent is the lexical richness of student essays related to essay grade and overall academic achievement?
5. How do learners’ vocabulary knowledge and student-essay lexical richness develop over time?

The first research question addresses the interrelationship between three facets of Swedish students’ vocabulary knowledge, namely receptive size, productive size and productive depth of vocabulary knowledge. The second research question concerns their respective relationships with vocabulary use operationalized as the lexical richness of learners’ written output while studying English at the university. The third and fourth research questions address the extent to which vocabulary knowledge and use are related to

---

2 Take-home essays will in the current study be used to refer to an examination task which is written outside of class and can be contrasted with the writing task of timed compositions.
4

academic achievement as apparent in essay and course grade. The fifth research question is concerned with how vocabulary knowledge and use in learners’ written output change over one and two terms of English studies at the University of Stockholm (i.e. a four- and nine-month period. The research questions will be investigated through an empirical study comprising 34 Swedish university students of English.

1.3 Outline

This thesis is divided into seven chapters with chapter one providing the background, rationale, and aims followed by the outline of the thesis. Chapter two outlines the concepts underpinning the study, and presents the different ways vocabulary knowledge and ability have been defined in the research literature. It also provides a review and evaluation of ways of assessing vocabulary knowledge and vocabulary use. In chapter three previous research related to the four areas of interest (cf. section 1.2: a) – d)) is presented and discussed. Chapter four presents the method, which includes information about the informants, test instruments, data collection and analysis procedure. Moreover, the development and validation of a productive depth of vocabulary test and a pilot study are reported.

Chapter five presents the findings of the study in five steps. The first step involves the descriptive statistics of the test scores. The other four steps address the five main research questions presented in section 1.2. Chapter six provides an interpretation and discussion of the main findings in the light of the theoretical background and earlier findings and implications for theory, assessment and pedagogy. Lastly, chapter seven concludes the study by summarizing the main contributions of the study. In addition, the chapter discusses the main limitations of the study and presents suggestions for future studies.
2. Conceptual framework

In this chapter the concepts underpinning the present study will be outlined. The chapter begins with an overview of how the facets of vocabulary knowledge (cf. sections 2.1 – 2.5) operationalized in the present study have been defined and measured in the research literature. The construct of vocabulary knowledge will be approached from the point of view of three dimensions: productive/receptive, size and depth. The review continues with an outline of three comprehensive conceptualizations (cf. section 2.6), which attempt to integrate the different components and processes involved in vocabulary knowledge and ability within one single framework. The last part of the chapter deals with measures of lexical use (cf. 2.7).

Before outlining the underlying constructs examined here a few points should be made about the main conceptual distinctions made. Most conceptualizations of vocabulary knowledge make two main distinctions, one being between size and depth of vocabulary knowledge and another being between passive/receptive or active/productive vocabulary knowledge. These two distinctions will serve as the backdrop of the review of vocabulary knowledge. The construct of depth of vocabulary knowledge is the dimension of vocabulary knowledge that is the main object of terminological and conceptual ambiguity and it will thus be explored in more detail.

2.1 The distinction between vocabulary knowledge and use

Within second language acquisition theory a basic distinction is made between capability and performance. Capability refers to learners’ L2 knowledge and performance to the actual use of that knowledge (Tarone 1988; Ellis 1989). In the present context this means that the vocabulary ability of a learner is the same irrespective of whether he or she writes an e-

---

3 In the research literature different terms such as lexical competence, vocabulary ability and vocabulary knowledge are often used interchangeably. In the current study vocabulary knowledge will be used as a cover term for receptive and productive size and depth, while vocabulary ability is defined as a wider construct encompassing more knowledge dimensions of vocabulary, such as lexical access and metacognitive strategies (cf. section 2.6.3).
mail to a friend or shortly thereafter writes an academic research paper, while the vocabulary uses differ between the two tasks.

It is important to point out that all vocabulary measures are ultimately measures of use whether the focus of assessment is the underlying vocabulary ability or vocabulary use per se. However, there is a point in making such a distinction because when we look at vocabulary features of real language use the link to the underlying ability is more indirect than when examining learners’ vocabulary use in more controlled contexts, such as the **receptive Vocabulary Levels Test** (RVLT) or even the vocabulary use in inauthentic writing tasks that are administered for the sole purpose of eliciting samples of vocabulary use for specific research purposes rather than being part of the learner’s authentic use of lexis.

This distinction is also valid in the light of the various conceptualizations of vocabulary knowledge and ability in that these models are typically intended as frameworks for assessing ability rather than use. A case in point is Chapelle’s (1994) vocabulary ability model (cf. section 2.6.3). She posits a dimension of vocabulary ability which she terms **the context of vocabulary use**. Although the term includes the words, context and use, neither the dimension nor the model as a whole are theoretical frameworks of vocabulary use. Rather, the dimension is a specification of the ability to use words appropriately depending on the context of use, e.g. a formal or informal context.

The distinction between vocabulary knowledge and use made here is especially relevant in relation to the degree of linkage between vocabulary knowledge and authentic vocabulary use (here operationalized as the lexical richness of take-home essays) in that even if we only focus on lexical features in learners’ language use in the context of authentic tasks we should not expect to find a one to one correspondence between the underlying vocabulary ability and vocabulary use. As expressed by Witalisz (2007: 115) in relation to Meara’s (2005) claim that the Lexical Frequency Profile (LFP; cf. section 2.7.1.2.2) is not sensitive enough to discriminate between learners with similar vocabulary sizes:

> This objection emphasizes a particular perspective taken by lexical research, where the assumption is that a written text is the learner’s display of his/her productive lexicon, which directly reflects the size of his/her passive lexicon. This perspective may well be called ‘vocabulary assessment in writing’, where written output is elicited from learners in order to evaluate their vocabulary. From the perspective of writing research, the focus may be different, that is, the role of vocabulary in evaluating the text as a whole, so it is vocabulary in writing assessment rather than vocabulary assessment in writing. If the priority is the task fulfillment and lexical accuracy, low-frequency (advanced, sophisticated) vocabulary does not necessarily contribute to a higher holistic rating.
The perspective taken here is that vocabulary use in authentic communicative tasks is not a direct reflection of the underlying vocabulary ability. The relationship is mediated by factors such as higher-order skills, task demands, motivation, etc. It is precisely because the link cannot be assumed to be direct that it is of interest to examine the relationship of vocabulary knowledge to vocabulary use, since it can provide us with information about what measures of vocabulary knowledge can be employed to predict the sophistication of vocabulary use in different contexts.

2.2 Vocabulary knowledge

As noted in section 1.1 in the research literature vocabulary knowledge has been recognized as a multifaceted part of a learner’s language ability, resulting in different ways to operationalize the construct when attempting to assess learners’ vocabulary knowledge. However, a number of researchers identify certain common key dimensions of vocabulary knowledge, such as a size and a depth dimension (e.g. Daller, Milton and Treffers-Daller 2007; Nation 2001; Qian 2002). Although there is some agreement regarding what parts and sub-parts should make up a definition of vocabulary knowledge, these models are descriptive in nature and do not constitute a set of theories of vocabulary ability “that make it possible to test predictions in a more informed way” (Meara 1996; Henriksen 1999: 315).

In this section the main approaches to defining and operationalizing vocabulary knowledge prevalent in the research literature will be presented and discussed and these will provide a theoretical basis for operationalizing size and depth of vocabulary knowledge in the empirical study.

The review of the dimensions of vocabulary knowledge will be concluded by a presentation of three comprehensive models that attempt to incorporate the different dimensions of vocabulary knowledge under one single framework. These comprise two frameworks of vocabulary knowledge and one framework of the wider construct of vocabulary ability.

2.3 The receptive/productive distinction

According to Milton (2009), a common convention in the research literature is to distinguish between receptive (R) and productive (P) vocabulary knowledge. It is generally recognized that receptive knowledge normally precedes productive knowledge. Essentially, R involves the knowledge required for decoding the meaning of a word form encountered while listening or reading, whereas P involves encoding meaning in an appropriate word form while speaking or writing (Nation 2001).
Rather than pointing to R and P being different, Melka (1997) invokes the notion of a bi-polar continuum ranging from R to P and suggests that the distance between R and P should be conceptualized as increasing degrees of word familiarity.

The R/P continuum metaphor will be employed in the present study as a way of defining the underlying construct measured by the three discrete-item tests employed here. This metaphor in and of itself does not provide a detailed description of R/P mastery. In order for this metaphor to be useful as a conceptual tool for measuring and reporting learners’ R/P knowledge, we need a specification of the different stages of the R/P continuum. This should enable us to specify in more detail what degree or stage of R/P mastery a specific vocabulary test taps into and thus provides us with a conceptual basis for interpreting the results of the vocabulary tests employed as well as for comparing results across different studies (Schmitt forthcoming).

Schmitt (forthcoming) states that one way of identifying word knowledge stages in the R/P continuum is the framework provided by Laufer and Goldstein (2004) and Laufer, Elder, Hill and Congdon (2004) who propose, as part of the development of a computerized test of vocabulary size, a four-way categorization of the form-meaning link along the R/P continuum. Figure 1 displays an illustration of how the four proposed categories of degrees of word knowledge might be situated along the R/P continuum.

Figure 1. Schmitt’s (forthcoming) suggested terms for Laufer and Goldstein’s (2004) four-way categorization of different degrees of word knowledge.

The four terms for describing the four degrees of word knowledge were taken from Schmitt (forthcoming) who proposes a more transparent terminology than that of Laufer and Goldstein (2004). In the method section this four-way categorization will be employed as a basis for specifying what vocabulary knowledge facets the respective discrete vocabulary tests employed are perceived as reflecting. In other words, it will form a basis of the specification of the underlying trait of the respective vocabulary tests.

Many studies have examined the receptive and productive vocabulary sizes of L2 learners. Although it has been shown that R is larger than P and
that R precedes P in that the P/R ratio decreases as the word frequency decreases (for a discussion of word frequency cf. section 2.4.1.2), (e.g. Waring 1999, Laufer 1998, Laufer and Paribakht 1998, Fan 2000), such studies have produced inconsistent results. Studies have reported different estimates of the gap between the two as a result of employing different definitions (Schmitt forthcoming). The estimates range from 16% to 92% of the receptive vocabulary being known productively (ibid.)

In section 2.1 a conceptual distinction was made between vocabulary knowledge and vocabulary use. This distinction is relevant for how we should define receptive and productive vocabulary knowledge. According to Read 2000, the conceptual confusion regarding the definition of receptive and productive vocabulary, which as noted above has resulted in incompatible findings pertaining to the relationship between receptive and productive size, is that two different ways of defining reception and production have been used interchangeably. One way of distinguishing reception and production is by defining the former as recognition and the latter as recall. Recognition is when “test-taker are presented with the target word and are asked to show that they understand its meaning [whereas recall is when] they are provided with some stimulus designed to elicit the target word from their memory” (Read 2000: 155). Another way of distinguishing reception and production is by perceiving the difference in terms of the contrast between comprehension, which entails that “learners can understand a word encountered in context while listening or reading” and use which entails that “the word occurs in their own speech or writing” (Read 2000: 156). The second way involves viewing the difference as one between comprehending the meaning of words “in context while listening or reading” and using a word in one’s own speech and writing (Read 2000: 156).

Read (2000: 249) succinctly summarizes the source of the conceptual confusion involved in the treatment of the issue of receptive/productive knowledge prevalent in the research literature by suggesting that “one way to reduce the conceptual confusion [is] to look separately at recognition vs. recall and understanding vs. use. The former distinction applies more to discrete vocabulary tests, whereas the latter one is appropriate for embedded, comprehensive and context-dependent measures”.

In the current study, in line with Read’s (2000) suggestion a distinction is made between receptive/productive knowledge and comprehension/use. The former bi-polar continuum is defined here as linked to underlying knowledge, whereas the latter is perceived as belonging to the use domain. Accordingly, receptive and productive vocabulary knowledge are terms used for describing learners’ vocabulary knowledge without making any claims that there is a direct link between the two and real language use. Vocabulary comprehension and use will be used to refer to instances of authentic language use.
It must be stressed that this distinction is a simplification of language knowledge and use and only serves as a tool for maintaining some degree of conceptual clarity and as an interpretative framework for the current investigation of the relationship between vocabulary knowledge measured by discrete vocabulary tests and vocabulary use here operationalized as lexical richness of take-home essays. A detailed discussion of how reception and production will be operationalized in the present study will be provided in sections 2.4.3.2.1 and 2.7.1.2.2).

In regard to how receptive and productive vocabulary knowledge are related, high correlations have been found among adult EFL and ESL learners (Laufer and Paribakht 1998; Meara and Fitzpatrick 2000).

As mentioned above productive size of vocabulary knowledge has mainly been examined from the point of view of how it is related to receptive vocabulary size. In the present study it will be investigated in terms of how it is related to productive depth of vocabulary knowledge (cf. section 3.1) among other things.

Moreover, there are no studies which have looked at the relationship among Swedish learners of English. Webb (2008: 93) calls for further studies into the relationship between learners’ receptive and productive size, since “it is likely that this relationship varies from group to group”. The following two research questions address the relationship between receptive and productive size among Swedish university students of English:

1. What is the relationship between learners’ receptive and productive size?
2. How does the relationship between learners’ receptive and productive size change over time?

As indicated by previous studies (e.g. Laufer 1998; Laufer and Paribakht 1998; Meara and Fitzpatrick 2000), we should expect to find a strong overall relationship between receptive and productive size.

The practical implication of knowing how learners’ receptive and productive size are related and whether there is a change in the relationship is that it can inform researchers and teachers of the cost and benefit of employing a more practical receptive test for obtaining an overall picture of learners’ size. In many pedagogical and research assessment situations, test-administrators are limited in terms of the number of tests that can be practically employed and opt for a small set of measures for obtaining a picture of learner’s overall state of vocabulary knowledge.

The university student population in Sweden can be characterized as relatively diverse among other things in terms of language background and type of formal education, i.e. private or municipal school. In order to be able to diagnose students of English at different stages in their university studies,
it may help to characterize what the vocabulary knowledge profile looks like and whether it changes during the course of university studies of English.

This will be done by addressing among other things the relationship between students’ receptive and productive size and how the relationship develops over the course of university studies of English. The studies reviewed above have mainly looked at the relationship statically (Meara and Fitzpatrick 2000; Laufer and Paribakht 1998) or tracked the development on the basis of cross-sectional data (Laufer 1998).

2.4 Size of vocabulary knowledge

As mentioned in section 2.2 it is generally recognized within L2 vocabulary research that a definition of vocabulary knowledge should at least comprise two dimensions, size and depth of vocabulary knowledge (e.g. Read 2000, Nation 2001, Qian 2002, Qian and Schedl 2004). In his study of vocabulary knowledge and reading comprehension Qian (1999:1) underlines this basic distinction which is widely adopted in second language vocabulary research:

A recognition of depth and breadth as two primary dimensions of vocabulary knowledge is essential to understanding the relationship between vocabulary knowledge and reading comprehension. (…) it is necessary to clarify what we mean by ‘breadth’ and ‘depth’ of vocabulary knowledge. (…) breadth is defined as vocabulary size, or the number of words for which a learner has at least some minimum knowledge of meaning. Depth of vocabulary knowledge is defined as a learner’s level of knowledge of various aspects of a given word, or how well the learner knows this word.

Qian’s (1999) definition of depth is word-oriented rather than defined as referring to the overall quality of a learner’s lexicon. The former way of perceiving depth has been ascribed to the components approach, i.e. defining depth as consisting of individual elements (cf. 2.5.2), while the latter is associated with the network knowledge approach (cf. section 2.5.2.1). Even though we can only practically test a limited number of words at any one time irrespective of how we define depth it is suggested here that learners’ performance on depth tests can be interpreted as providing a rough indication of the degree of knowledge of the population of lexical items of the particular type sampled for testing rather than merely degree of knowledge of the specific test items. In other words a learner who exhibits depth of knowledge of a representative sample of words can be expected to have relatively deeper knowledge of a wider population of words of the same type, e.g. academic words or technical vocabulary of a specific field.

In the research literature, it has for a long time been acknowledged that there is more to vocabulary knowledge than mere familiarity with the meanings of words, which has prompted numerous investigations into
learners’ depth of vocabulary knowledge with a view to furthering our understanding of the complexities of second language vocabulary acquisition. To this end a wide range of conceptualizations and operationalizations of the construct have been developed of which the main ones will be reviewed in section 2.5.

The perceived vagueness of the concept of depth of vocabulary knowledge and the different ways it has been defined has prompted some researchers to argue that it should be dispensed with altogether (cf. section 2.5).

2.4.1 Issues involved in and rationales for testing learners’ vocabulary size

This section will deal with the dimension of size of vocabulary knowledge mainly from the perspective of the issues involved in measuring learners’ vocabulary size and the different rationales for assessing the construct. It will be concluded with a review of a selection of vocabulary size testing instruments.

2.4.1.1 Units of counting vocabulary

Vocabulary research often entails counting words for different reasons (e.g. to count the number of words belonging to a particular frequency range or register, to assess vocabulary growth, or to establish the minimum vocabulary size required for carrying out various language tasks). Different lexical counts can generate inconsistent results, and a recurrent issue in vocabulary studies is that size estimates are reported, but without a clear indication of how they were derived. Accordingly, in order to be able to compare and interpret different vocabulary size assessments, we must know what unit of counting the assessment is based on (Schmitt forthcoming).

Most of the item definitions are applicable either to the size of vocabulary known or to how many words a learner must know in order to understand a text. There are a number of variables that affect determinations of what constitutes a word. One is the proficiency level of the learner, because less proficient learners might not recognize and make use of the morphological links between words which means that a less inclusive item definition is more appropriate. Another is whether the estimate is of receptive or productive size. According to Nation (2006) a more restricted item definition is appropriate for estimating learners’ productive size, because having productive knowledge of one item does not necessarily entail having knowledge of another morphologically related item.

The main methods of counting have in the literature been classified as follows: One way is simply to count every item in a text and even if the same word form occurs more than once each instance is counted as one token. This
way of counting is only relevant when we want to know how many words there are in a text and not for estimating the size of learners’ vocabularies or for formulating vocabulary learning goals. A second way is to count *types*, i.e. count every different word form as one type (e.g. *laugh* and *laughed* are two types) (Read 2000; Nation 2001). A third way is to count *lemmas*, which usually means that a baseword and its inflectional forms and reduced (*n’t*) forms are counted as one occurrence (Nation 2001). The use of lemmas as the unit of counting is based on the idea of learning burden (Swenson and West 1934 cited in Nation 2001). Once a learner is familiar with the inflectional system of English, if for instance the base form *show* is already known, the effort required to learn the inflected from *showed* is negligible (ibid.). This way of counting greatly reduces the number of units in a corpus as showed by Bauer and Nation (1993) who calculated that the 61,805 tagged types in the Brown Corpus are reduced to only 31,617 lemmas. Moreover, it seems to be consistent with some psycholinguistic research into how the brain stores lexical items (cf. Indefrey and Levelt 2004).

A fourth unit of counting vocabulary which is also based on the concept of learning burden is the *word family*. It consists of a base word, its inflected forms and semantically closely related derived forms such as *–ly*, *-ness* and *un-*. There is some evidence that the concept of word family also reflects the way the mind stores and processes lexical items (cf. Bertram, Baayen and Schreuder 2000; Marslen-Wilson and Tyler 2007). The main problem in counting using the unit of word family is determining what items should and should not be included in a word family.

Bauer and Nation (1993) provide some guidance for deciding what inflectional and derivational forms of a base word should be contained under the same word family, based on the criteria of frequency of the affix, productivity of the affix, predictability of the meaning of the affix and regularity of the written form of the head word etc. They define six levels which produce families with progressively wider memberships. From a vocabulary testing perspective these criteria provide a description of what is an item of vocabulary for the purposes of the test. Depending on whether we measure receptive or productive size one can adjust the number of levels forming the basis for determining what inflectional and derivational affixes should be included in the word family. By using a standard basis for describing the unit in vocabulary size studies, useful comparison of results can occur. Even if different studies use different levels as the basis of defining the members of word family, it is still possible to compare the results as long as the description of what is an item of vocabulary is based on a standard set of levels (Bauer and Nation 1993).

It should be pointed out that it is one thing to spell out criteria for determining what items should be included in a word family and another thing to apply these criteria. There will always be an element of subjectivity
involved when assigning a word form to a word family rather than having it as a base word on its own.

Table 1 displays Bauer and Nation’s (1993) guidelines for defining the members of a word family

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>A different form is a different word. Capitalization is ignored. -ly, -ness, -th, -y, non-, un-, all with restricted uses.</td>
</tr>
<tr>
<td>Level 2</td>
<td>Regularly inflected words are part of the same family. The inflectional categories are - plural; third person singular present tense; past tense; past participle; -ing; comparative; superlative; possessive.</td>
</tr>
<tr>
<td>Level 3</td>
<td>-able, -er, -ish, -less,</td>
</tr>
<tr>
<td>Level 4</td>
<td>-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-, all with restricted uses.</td>
</tr>
<tr>
<td>Level 5</td>
<td>-age (leakage), -al (arrival), -ally (idiologically), -an (American), -ance (clearance), -ant (consultant), -ary (revolutionary), -atory (confirmatory), -dom (kingdom; officialdom), -eer (black marketer), -en (wooden), -en (widen), -ence (emergence), -ent (absorbent), -ery (bakery; trickery), -ese (Japanese; officialese), -esque (picturesque), -ette (usherette; roomette), -hood (childhood), -i (Israeli), -ian (phonetician; Johnsonian), -ite (Paisleyite; also chemical meaning), -let (coverlet), -ling (duckling), -ly (leisurely), -most (topmost), -ory (contradictory), -ship (studentship), -ward (homeward), -ways (crossways), -wise (endwise; discussion-wise), anti- (anti-inflation), ante- (anteroom), arch- (archbishop), bi- (biplane), circum- (circumnavigate), counter- (counter-attack), en- (encage; enslave), ex- (ex-president), fore- (forename), hyper- (hyperactive), inter- (inter-African, interweave), mid- (mid-week), mis- (misfit), neo- (neo-colonialism), post- (post-date), pro- (pro-British), semi- (semi-automatic), sub- (subclassify; subterranean), un- (untie; unburden).</td>
</tr>
<tr>
<td>Level 6</td>
<td>-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re-</td>
</tr>
</tbody>
</table>

These levels form a hierarchy in which the number of inflectional and derivational forms increases as we progress along the levels.

The concept of word family is not without problems. According to Bauer and Nation (1993) ‘the meaning of the base word in the derived word must be closely related to the meaning of the base when it stands alone or occurs in other derived forms’ (253). However, they fail to provide any criteria for how to determine whether two potential members are closely related in meaning and thus should be part of the same word family. The reliability of the measures based on the word family concept is dependent on the extent to which the measures are based on the same criteria.

Gardner (2007) draws our attention to three inherent weaknesses in the concept of word families. The assumption that knowing one member of a word family automatically means knowing the rest lumps different learners at different proficiency levels and from different backgrounds into one group. It is unlikely that learners of different proficiency levels and language
backgrounds are equally capable of seeing the morphological link between the different members of a word family. Consequently, there is reason to raise doubts as to whether the concept of word family is equally psychologically valid, from a learner perspective, i.e. for learners at different proficiency levels. However, according to Gardner (2007), this mainly applies to low-proficiency learners. Advanced learners are much more likely to see the morphological connection between the different members of a word family, due to extensive morphological experience. The second point raised by Gardner (2007) is that the concept of word family does not take into account semantic aspects, such as homonymy and polysemy. In other words, in frequency lists based on the concept of word families, each word form is counted as one single word without taking into account that it may have a number of unrelated meanings. Accordingly, the Vocabulary Levels Tests (VLT) (cf. section 2.4.1.2) take into account neither other unrelated meanings (homonymy) of a word form nor other related meanings (polysemy) that can be expressed by the same word form. Thirdly, lexical counts and analyses based on the concept of word family do not include multiword items in the definition of a word.

A number of studies have stressed the important role of knowledge of collocations in language proficiency. This has been investigated within a separate field of second language acquisition. In the present study knowledge of collocations is conceptualized as a component of depth and will be approached from the point of departure of the field of second language vocabulary acquisition.

Schmitt (forthcoming: 209) contends that the main benefit of using the word family as the unit of counting is that it “reduce[s] the redundancies that can occur in word counting, that is all the semantically related members are included in a word family, they do not have to be handled again in another category, as they would have to be with lemmas and especially word forms”

The measures of learners’ vocabulary knowledge and use employed here are all based on the concept of word family or include test items sampled from word lists that are based on this unit of counting. As mentioned above this unit of counting is not without certain shortcomings which should be kept in mind, especially when interpreting productive size estimates.

2.4.1.2 Frequency levels

Frequency affects almost all aspects of lexical ability and acquisition. A high frequency word such as house is processed more quickly and acquired before a low frequency word such as stout. This insight has guided vocabulary research in general and vocabulary test development in particular (Schmitt forthcoming).

Most vocabulary size tests are based on the insight that words are generally acquired according to frequency of occurrence and that a relatively small number of words cover a disproportionate percentage of word
occurrence in language. For instance, Nation (2006) reports that the 2000 most frequent base words cover about 83% of the occurrences in the London/Oslo/Bergen corpus (LOB).

There are two principal methods of measuring vocabulary size. One is based on sampling from a dictionary and the other is based on sampling from a corpus or a word frequency list derived from a corpus. All the tests outlined in section 2.4.3 are in one way or another based on sampling from a corpus. We assume that corpus frequency corresponds to real frequency in use and therefore the main benefit of using the corpus method of word sampling is that researchers can assess learners’ vocabulary size at different frequency levels. It should be noted, however, that most vocabulary size tests are based on frequency counts of written language and that they thus reflect the word frequency of the written texts. This means that such vocabulary size tests do not capture possible gaps learners might have in their knowledge of basic everyday vocabulary that is not frequent in written language such as vocabulary denoting kitchen utensils. Another limitation of frequency counts of words is that they tend to only include orthographic words and seldom include frequencies of formulaic phrases (Schmitt in press). This is the same point Gardner (2007) made about word families but it actually applies to many current procedures underlying estimations of vocabulary size.

There are a number of commonly used sources of word frequency. The General Service List (GSL) developed by West (1953) contains a list of high frequency words which contains about 2000 basewords. Two other sources of word frequency are Thorndike and Lorge (1944) and Kučera and Francis (1967). Although these word frequency lists are outdated, they have been widely used as sources of word frequency data for a number of currently employed vocabulary size tests. A more up-to-date source of word frequency data is Nation’s (2006) fourteen 1000 BNC word lists which contain the 14 thousand most common word families in the BNC corpora.

There are also two other word lists that have been used as word sample sources for both vocabulary size and depth tests that are not frequency based. These are the University Word List developed by Xue and Nation (1984 cited in Schmitt, Schmitt and Clapham 2001) and the Academic Word List (AWL) developed by Coxhead (2000). Both are lists of vocabulary items common in academic texts. The newer AWL has replaced the UWL as the standard list of academic words.

2.4.3 Testing vocabulary size

The rationales for assessing learners’ size of vocabulary knowledge can be divided into two main types, namely applied/pedagogical and theoretical. Examples of pedagogical rationales are investigations of whether a particular group of learners reach the vocabulary size thresholds for adequate
comprehension of written and spoken texts of the sorts outlined in section 3.3.1.3 and of the effect a certain learning context or treatment has on the development of learners’ vocabulary acquisition. Examples of theoretical rationales for assessing learners’ vocabulary size are for instance examinations of how vocabulary size is related to other aspects of language ability, and of aspects of the nature of vocabulary acquisition. This necessitates the availability of simple and practical vocabulary tests that can yield reliable estimates of learners’ vocabulary size. Presently, three such measures for measuring receptive size of vocabulary knowledge are available and widely applied for pedagogical purposes and research; the Yes/No Test (Y/N Test) devised by Meara and associates (cf. Meara 2006) and the receptive Vocabulary Levels Test (RVLT) first developed by Nation (cf. Nation 2001). These two tests as well as a newly developed test called the Vocabulary Size Test (VST) described in Nation and Beglar (2007) will be reviewed below.

Due to the relative facility of measuring receptive vocabulary knowledge compared to productive vocabulary knowledge, receptive tests, and the RVLT in particular, are regarded as possessing a high degree of reliability and validity. Productive tests, however, are more contested in the research literature in this regard. The two most widely used tests of productive size of vocabulary knowledge, the productive VLT (PVLT) developed by Laufer and Nation (1999) and Lex30 developed by Meara and Fitzpatrick (2000) will also be reviewed below.

As noted in section 2.4.1.2 there are two principal methods of measuring vocabulary size. One is based on sampling from a dictionary and the other from a corpus or a corpus-based frequency list. All three test formats, the Y/N Vocabulary Size test, the two VLTs and the Lex30 are based on corpus-derived frequency lists.

The review of the vocabulary size tests below will include the following information: word-sampling method, format of the test, type of information obtained, subsequent adaptations and a short evaluation.

2.4.3.1 Tests of size of receptive vocabulary knowledge

2.4.3.1.1 Y/N vocabulary size tests
The test format was developed by Meara and associates (cf. Meara 2006). Like the two VLT tests (cf. sections 2.4.3.1.2 and 2.4.3.2.1), the word selection is based on the assumption that words are learned according to frequency of occurrence. The words are sampled from ten frequency bands (1K-10K) and the test provides a profile of words known at each frequency level as well as an overall vocabulary size estimate within the ten frequency bands. The unit of counting is the lemma.

It is a self-report test where the test taker is presented with a series of words for which the test taker is prompted to tick a Yes or No box depending
on whether the word is known. The test contains a number of non-words, which reduce the total score if ticked as known. The test has been further developed into a computerized version X_Lex which measures words up to the 5K frequency band and the Y_Lex which measures words at the 6-10K frequency bands (Meara 2005; Meara and Miralpeix 2006).

One of the main strengths of the Y/N Vocabulary Size test format is that it is simple for the test takers to do. Moreover, the simplicity of the task means that a larger sample of words can be tested, which is important for obtaining enough test items for making reliable vocabulary size estimates (Read 2000).

2.4.3.1.2 The Receptive Vocabulary Levels Test

The RVLT has been widely used in research to assess learners’ vocabularies. The RVLT is regarded as coming close to being a standard test of receptive vocabulary size (cf. Read 2000; Schmitt forthcoming).

The test is based on the word frequency data in Thorndike and Lorge’s (1944) list cross-checked against the GSL and Kučera and Francis’ (1967) frequency count and consists of items testing words at various levels of frequency: the first 2,000 words, 3,000 words, 5,000 words and 10,000. A fifth level is the AWL level which as the name suggests consists of items sampled from the AWL. The words were randomly selected from the lists and the test scores indicate the proportion of the total number of words at each frequency level the learner knows. Adding up the scores on the different levels can provide a broad measure of the learner’s total vocabulary size (Read 2000). The levels are supposed to reflect different learning objectives with the 2,000 and 3,000 levels representing high frequency vocabulary needed to function effectively in English; the 5,000 level represents the upper limit of low frequency vocabulary worth spending time on in class; the AWL level comprises words necessary to read academic texts at the university level, and, lastly the 10,000 word level contains the more common low frequency words (Read 2000).

There have been some further developments of the test. As regards the receptive test, Schmitt, Schmitt and Clapham (2001) developed and validated two enlarged parallel versions consisting of fresh word samples with 60 words to be matched with 30 definitions at each level instead of 36 words and 18 definitions in the previous versions.

Although the RVLT is widely applied in assessment and research contexts, there are some inherent weaknesses in the test format. One is that the target words are based on old frequency counts, which are up to 60 years old. This might affect the validity of the tests in that the frequency counts might be outdated, not reflecting today’s word usage. However, a relatively recent study by Hu and Nation (2000) showed that the 2,000 most frequent words sampled from the GSL accounted for between 75% and 90% of the lexical coverage in various kinds of text. This suggests that the count still has some relevance today.
Another weakness pertains to the assumption that vocabulary acquisition follows words’ frequency of occurrence in that high frequency words are acquired before low frequency words. As mentioned in section 2.4.1.2, this is not necessarily the case since L2 learners may lack knowledge of certain basic words pertaining to contexts in which the L2 learners mainly use their L1. This means that the learners might have a relatively advanced vocabulary acquired at school but lack knowledge of words that are common in spoken language (Viberg 1993). In other words, L2 learners may have gaps in their basic vocabulary that the RVLT is unable to reflect.

However, validation studies of the RVLT have produced results which show that the five levels form an implicational cline in which the lower levels receive higher scores than the higher levels (Read 1988 cited in Read 2000). This indicates that in general the test seems to reflect a natural process of vocabulary acquisition and also provides support for the assumption that the word frequency counts of the lists are at least to some extent valid today.

Since the RVLT is used in the current study, more details about the test will be provided in sections 4.2.1 and 4.2.1.1.

2.4.3.3 The Vocabulary Size Test

The Vocabulary Size Test was developed by Paul Nation and associates and is described in detail in Nation and Beglar (2007). The frequency count of the test items is based on the 10-million-word BNC spoken corpus. The target words are presented in the context of a non-defining sentence, i.e. “It is a miniature” and the test taker is supposed to match the target word with the correct definition among three distractors. The VST is made up of 14 clusters, each cluster representing a 1000-word frequency band and containing ten items which represent 100 words within that frequency band (Nation and Beglar 2007).

Schmitt (forthcoming) referring to the high reliability figures and the good technical characteristics of the test, states that the test is promising and that there are no reasons why the test should not be used. This test has a few advantages over the RVLT. One advantage is that since it tests learners’ knowledge of words up to the 14th 1000 word frequency band, there is a lower risk of a ceiling effect when assessing the vocabulary size of advanced learners. Moreover, it is better at calculating learners’ overall vocabulary size as opposed to the RVLT which was intended as a profile of vocabulary knowledge at various frequency levels.

One problem with the VST for the purpose of the present study is that it does not have a productive version based on the same word sample, which would enable the assessment of learners’ receptive and productive knowledge in tandem. The ultimate reason why it could not be considered for inclusion in the present study was simply that it was not available at the time.
Having reviewed a selection of tests of receptive vocabulary size, in the next section two tests of productive vocabulary size will be outlined.

2.4.3.2 Tests of size of productive vocabulary knowledge
As mentioned in section 2.2 studies investigating learners’ productive vocabulary knowledge have suffered from a lack of clarity of the concept, which has had negative effects on the validity of measures of productive vocabulary knowledge. In this section two commonly used tests will be outlined and discussed in light of some of the problems involved in the design of measures of productive vocabulary knowledge.

2.4.3.2.1 The Productive Vocabulary Levels Test
The productive version (PVLT) currently exists in two parallel versions, one of which consists of item sets taken from three older versions, whereas the other is made up of levels from one older version (Laufer and Nation 1999). The PVLT is based on the same word sampling procedure as the RVLT. In the PVLT, the test takers are presented with 18 sentences, at each level, with a blank in which a variable number of the initial letters are provided in order to ensure that only the target word fits (Read 2000).

In contrast to the RVLT the validity of the PVLT has been questioned by a number of researchers (e.g. Meara and Fitzpatrick 2000; Read 2000; Fitzpatrick 2007a; Webb 2008). The points of criticism can be divided into three main areas pertaining to validity, extrapolation, and test construction. Furthermore, a number of researchers (e.g. Melka 1997; Read 2000; Webb 2008 and Schmitt forthcoming) have called for a clearer definition of productive knowledge so that we know what it is we set out to measure (cf. 2.3).

The point of departure of the criticism leveled at the PVLT seems to be Laufer and Nation’s (1999: 41) claim that the test scores can be interpreted as the approximate number of words ‘readily available for productive use’. Although they do stress that the PVLT measures controlled productive ability defined as “the ability to use a word when compelled to do so by a teacher or researcher, whether in an unconstrained context such as a sentence-writing task, or in a constrained context such a fill-in task where a sentence context is provided and the missing word has to be supplied”, the claim is perhaps too broad since it also encompasses the ability to use a word in an unconstrained context such as in a sentence-writing task (Laufer and Nation 1999: 37). Some researchers have even questioned whether the PVLT should be regarded as a productive test (e.g. Read 2000; Webb 2008). As mentioned in section 2.3 in accordance with Schmitt’s (forthcoming) suggestion, we should regard the test as a form-recall test and not a productive test in the sense that it provides an estimate of the number of words a learner can use in his/her speech or writing (cf. section 2.7).
Concurrent validity of the test has been assessed by Laufer and Nation (1995) and Fitzpatrick (2007a), who showed that the PVLT scores correlate with scores on the LFP (described in section 2.6.1.2.2) and a productive translation test. However, it is important to point out as argued by Fitzpatrick (2007a) that the PVLT seems to address aspects of receptive knowledge as well as of productive. According to Fitzpatrick (2007a) this is not necessarily a drawback because by having tests that measure different points on the receptive/productive continuum (cf. section 2.2) we can obtain important information about learners’ vocabulary knowledge. In the present study the PVLT is employed as a measure of controlled cued productive vocabulary knowledge which should be perceived as “the first steps in productive mastery” i.e. being able to use a word freely in one’s speech and writing (Schmitt forthcoming: 95).

Since the PVLT is employed in the present study, further details about the test are provided in sections 4.2.1 and 4.2.1.2.

2.4.3.2.2 The Lex30 test of productive vocabulary

The Lex30, which is in the form of a free word association task, presents the test takers with 30 stimulus for which they are prompted to supply responses. Since “[t]here is no predetermined set of response target words for the subject to produce, the “Lex30 resembles a free productive task” (Meara and Fitzpatrick 2000: 22).

The stimulus words are sampled from Nation’s first 1000-word list (Nation 1984 cited in Meara and Fitzpatrick 2000) and have been selected to meet three main criteria. In order to ensure that the stimulus words are known by the subject and thus minimize the employment of receptive knowledge on the part of the test taker the words are highly frequent. Words that typically elicit a very narrow range of responses such as black or dog were avoided in favor of words that elicit a wide variety of responses. In order to encourage responses which are not common words, a minimum of 50% of the most common responses to the stimulus words by native speakers had to fall outside the first 1000 frequency band. A point is awarded for every response that falls outside the first 1000 frequency band (ibid).

The reliability of the test was assessed by means of a test-retest method which showed that there were no significant differences between the two sets of scores. Concurrent validity was tested in two ways, one by comparing native and non-native speakers’ scores on the test and the other by correlating the scores of the same test takers on two different tests. The comparison of native and non-native speakers’ scores showed a significant difference, which suggests that the measure can discriminate between learners at different proficiency levels. Meara and Fitzpatrick (2000) showed a positive correlation of .84 between the Lex30 and the Y/N-test. Further validity assessment was carried out in Fitzpatrick (2007a) by correlating the Lex30 with the PVLT and a productive translation test. While the scores
from the PVLT and the translation test showed a significant high correlation, the scores from the Lex30 only correlated modestly with the other two productive tests. This led Fitzpatrick (2007a) to argue that the lack of a strong correlation between the Lex30 and the other two tests is due to the tests measuring different aspects of vocabulary knowledge. However, by the same token one could argue that the strong correlation between the Y/N-test and the Lex30 is because they measure the same aspect of vocabulary knowledge.

As mentioned in section 2.2, receptive and productive vocabulary have been described as forming a continuum and it is argued by Fitzpatrick (2007a) that the stronger correlation between the PVLT and the translation test should be viewed as evidence for them occupying a position on the continuum closer to the receptive end than the Lex30. She supports her argument by pointing out that both the PVLT and the translation test rely on more activation properties which are defined as the degree to which the production of target responses is stimulated by a specific test. The PVLT is said to have the following three activation properties: L2 semantic stimulus, L2 orthographic stimulus and L2 collocational stimulus. The translation test has two activation properties the L1 semantic stimulus and the L2 orthographic stimulus. The Lex30 has one activation property: the L2 semantic stimulus (ibid.).

There are two main reasons why the Lex30 was not selected for the present study. First, the format is too dependent on the test-takers’ motivation to supply low-frequency words. Since the informants of the current study have no particular stakes in the testing, a more controlled format was deemed more appropriate. Second, since as described by Meara and Fitzpatrick (2000) the test resembles a free productive task it was deemed redundant to include such a test in addition to a measure of lexical richness.

Having discussed a few commonly employed vocabulary size tests, we will now examine the concept of depth of vocabulary knowledge.

2.5 Depth of vocabulary knowledge

As mentioned in section 2.3 whereas size measures focus on how many words are known, depth measures focus on how well words are known. However, Schmitt (forthcoming: 236) points out that the two notions should not be perceived as discrete, since “it can be said that all size measures are also depth measures in the sense that some quality of knowledge, no matter how minimal, must be operationalized as the criterion of sufficient knowledge [of words]”. This seems to be a valid description with the qualification that this “quality of knowledge” is limited to the form-meaning
link and that conceptions of depth of vocabulary knowledge often include other knowledge components as well.

There are a number of different conceptualizations of depth of vocabulary knowledge. Read (2004) suggests that there are three common ways of conceptualizing depth in the research literature: **precision of meaning** defines depth of vocabulary knowledge in relation to the degree to which a learner knows the meaning potential of a specific word. **Network knowledge** defines depth of vocabulary knowledge in terms of how well a learner's lexicon is organized which is assumed to be a reflection of the learner’s ability to link the word to related words and distinguish it from other words. Finally, the **components** approach is based on the idea that depth of vocabulary knowledge subsumes not only semantic features, but also other features of a word such as its orthographic, phonological, morphological, syntactic, collocational and pragmatic characteristics.

Schmitt (forthcoming) supplies us with a more detailed taxonomy of the different ways depth has been conceptualized which is illustrated in Figure 2.

![Figure 2](image-url)

*Figure 2. An illustration of Read’s (2004) and Schmitt’s (forthcoming) classification of the different ways depth of vocabulary has been conceptualized in the research literature*

He makes a distinction between conceptualizations of depth that focus on declarative knowledge and those that focus on procedural knowledge (e.g. automaticity). All three above-mentioned approaches to defining depth are concerned with learners’ declarative knowledge and these will be the focus of the following outline. Schmitt (forthcoming) makes a further distinction between approaches that are more oriented towards conceptualizing depth as the quality of knowledge of individual words and approaches that focus on the quality of the lexicon as a whole i.e. the organization of the lexicon.

Read (2004: 211) argues that the different ways of conceptualizing depth have contributed to conceptual ambiguity and he questions whether “depth
can meaningfully encompass the different uses to which it has been put”. Read’s (2004) critique of the term depth is based on the numerous ways it has been defined in the research literature. In the present study a two-way categorization of the different ways depth has been operationalized will be suggested. This way of approaching how depth has been defined in the research literature provides a perspective which removes the apparent contradictions by viewing the different conceptualizations as forming a continuum in terms of the number of elements that are included in the definition of depth, rather than perceiving them as necessarily incompatible definitions.

In the next sections each approach to defining depth of vocabulary knowledge in terms of declarative knowledge will be outlined in turn. As mentioned I will categorize the different conceptualizations of depth by distinguishing between two overall ways of defining depth of which I term one essentialist approaches and the other comprehensive approaches. The first one encompasses what Read (2004) terms the precision of meaning approach and the network knowledge approach. The second one encompasses what Schmitt (forthcoming) calls the components approach. The principal difference between these two main ways of defining depth is that the first one defines depth according to a perceived notion of what features best captures its essence whereas the second defines depth as encapsulating a wider set of sub-skills or components. Another difference is that the conceptualizations within the essentialist framework work with a more research-driven definition, i.e. depth has been defined with research in mind in order to facilitate the examination of learners’ lexical competence by identifying “a small number of critical dimensions” and as way of facilitating theory building by reducing the complexity of the construct (Meara 1996: 51). In regard to the merits of postulating two dimensions (i.e. size and organization) to theory building, Meara states that:

> These two dimensions have the advantage that they are relatively independent of the items that contribute to them, and do not require a detailed understanding of the way individual lexical items function.

Of course this view is only relevant if researchers attempt to operationalize all aspects of what is involved in knowing a word. Moreover, irrespective of how depth is defined when operationalizing the construct test-developers and test-administrators have to be familiar with the features of words that are tested in order to be able to judge the performance of the test-takers.

Within the comprehensive approaches, depth has been defined in relation to pedagogical goals. An attempt has been made to identify a wide range of sub-skills that might be conceived as making up depth with the aim of providing teachers and learners with a detailed specification of what sub-skills are involved in knowing a word both productively and receptively (cf.
Studies that have examined depth conceptualized as components have mainly been concerned with investigating the relationship between different components and the variable relationship of the components to size (e.g. Schmitt and Meara 1997; Nurweni and Read 1999).

As noted above the construct of depth has been called into question (cf. Read 2004; Milton 2009). For instance, Read (2004: 224) argues that depth is “inherently ill-defined” and the construct which the term is used to describe is too complex for being neatly classified under one cover term. It is important to acknowledge that we are dealing with a somewhat contested concept and that there is not as of yet a consensus on how to define depth or whether it is a useful term for examining learners’ vocabulary knowledge.

2.5.1 Essentialist approaches to defining depth
As noted in the previous section there are two ways of defining depth that are here categorized as essentialist, viz. the precision of meaning and the network knowledge approach. Only the latter will be reviewed in detail.

2.5.1.1 The network knowledge approach
An essentialist way to define depth of vocabulary knowledge is the network knowledge approach in which depth of vocabulary knowledge is seen as the ability to link a given word to other related words, e.g. paradigmatically, which includes synonymy relationships and quality relationships, i.e. \(x\) is a quality of \(y\) and syntagmatically which includes collocational relationships (Schmitt forthcoming). The lexicon is regarded as a network and the number of links between words reflects the degree to which words are integrated into the learner’s mental lexicon (Read 2000).

Learners’ depth of vocabulary knowledge conceptualized as network knowledge has been assessed both receptively and productively. A widely used test format operationalized as network knowledge is the Word Associates Test (WAT) which is a receptive test originally developed by Read (1993; 1998 cited in Read 2000) and adapted by a number of other researchers (e.g. Greidanus and Nienhuis 2001; Greidanus, Bekx and Wakely 2005; Qian 2002; Staehr-Jensen, 2005; Albrechtsen, Haastrup and Henriksen 2008). The test format exists in different versions, but typically presents the test-takers with a number of options for each target word which are linked to it in the three ways mentioned above. Although widely used as a measure of depth, it has been critiqued by Milton (2009) who questions whether the test should be perceived as a valid measure of depth. He argues that the word connections presented in the test tap into learners’ size of vocabulary knowledge, since once the learner knows the basic meaning of the test items, he/she can easily identify the correct word associates.

Productive formats typically contain a number of prompt words which are intended to elicit associations. According to Schmitt (forthcoming) associa-
Association responses have typically been categorized into three types: paradigmatic, syntagmatic and clang (i.e. association responses which sound like the target word, e.g. *kill*/*fill*).

An inherent weakness in the approach is that only a limited number of associational links have been employed as measures of learners’ organization of the lexicon. This weakness has been addressed by Fitzpatrick (2006), who has used Nation’s (2001) word knowledge taxonomy (outlined in section 2.5.2.1) to identify 17 subcategories of associational links.

There are as yet no standard tests that measure depth of vocabulary knowledge operationalized as network knowledge productively. Although it is relatively easy to develop a productive test of associations, there are problems concerning the analysis of the generated association data. One of the most fundamental problems is that learners’ association data tend to be compared to a native speaker norm. The main problem with comparing learner associations with native data is the lack of homogeneity among native speakers. Fitzpatrick (2007b) found a wide variety of responses to AWL prompt words among 30 native speakers. As pointed out by Schmitt (forthcoming) unless a study uses very frequent prompt words, native speaker responses are likely to be extremely heterogeneous.

In several studies by Meara and his colleagues (e.g. Meara 1996; Wilks and Meara 2002; Meara and Wolter 2004) in response to what they perceive as an unproductive focus on several different aspects of word knowledge, such as the way depth is defined according to Nation’s (1990; 2000) components approach (cf. section 2.5.2.1, it is suggested that depth of vocabulary knowledge should be viewed as a property of the lexicon as a whole and measured on one single scale rather than on several different subscales reflecting each of the different aspects of word knowledge. As noted in section 2.5.1 they identify two dimensions of vocabulary knowledge, size and structure or organization. The mental lexicon is conceptualized as a network made up of nodes which represent the individual words stored. According to this conceptualization the number of nodes corresponds to size and the number of connections between these nodes represents depth or organization (Meara and Wolter 2004). They argue that focusing on the depth of knowledge of individual words “implies that we need to test very many words in ever-increasing detail, and this very quickly leads us into serious logistical problems which constrain the types of hypotheses that we can test” (Meara and Wolter 2004:86). Instead of testing a small sample of words in ever increasing detail they propose a test with a representative sample of words assessing word associations.

In my view although many researchers have called for a holistic view of different human abilities and sub-abilities and although we can all intuitively acknowledge that human abilities are interrelated, for practical reasons it is often difficult to operationalize such conceptualizations. Moreover, the compartmentalization of human abilities does not exclude theory building in
which the different elements are seen as parts of a whole. Another problematic issue in Meara and Wolter’s (2004) account of the mental organization is exactly what features of word knowledge are contained within the nodes and what the individual connections signify. A relevant question is whether other knowledge features other than, for instance synonymy, quality (i.e. x is a quality of y) and collocational associations are involved in the organization of the mental lexicon. As noted above, in response to this, Fitzpatrick (2006) has developed a richer system of categorizing the ways in which words are associated which draws on Nation’s (2001) word knowledge taxonomy (cf. section 2.5.21) and which acknowledges form based associations such as derivational links between words.

One can look at the value of depth from two perspectives. One is in terms of vocabulary access. The more we know about a word the more entrenched it is and therefore more easily accessed. From this point of view it is difficult to see why only associational links (i.e. paradigmatic and syntagmatic links) between words might facilitate access. The second perspective has to do with vocabulary use. The more we know about a word the more we can use a word in different contexts. This applies to both comprehension and use. We can probably identify a selection of components that play a prominent role in both use and access. It might thus be disadvantageous for our understanding of vocabulary knowledge to limit the number of components beforehand. Such an approach to defining depth that identifies a wider set of word knowledge aspects is Nation’s (1990; 2001) components framework outlined as a subset of comprehensive approaches in section 2.5.2.1.

2.5.2 Comprehensive approaches to defining depth

2.5.2.1 Components approach
An example of a comprehensive way of defining depth is Nation’s (1990; 2001) components approach in which depth is defined as encompassing a wide range of sub-skills or knowledge elements. Nation (1990; 2001) has proposed a list of sub-skills involved in knowing a word in depth which are outlined in Table 2.

In addition to listing the relevant components of knowing a word the receptive/productive distinction is incorporated in the list, since it is believed that being able to understand a word in verbal or written communication involves different components from being able to produce the word. The framework is intended to specify what is demanded of the learner to understand or actively/productively use a word in different contexts. It has also been used in vocabulary research as a theoretical framework of depth of vocabulary knowledge.
Table 2. Nation’s (1990; 2001) list of depth of vocabulary knowledge components.

<table>
<thead>
<tr>
<th>Form</th>
<th>spoken R</th>
<th>What does the word sound like?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>How is the word pronounced?</td>
</tr>
<tr>
<td>written R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>word parts R</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>What word parts are needed to express the meaning?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meaning</th>
<th>form and meaning R</th>
<th>What meaning does this word form signal?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>What word form can be used to express this meaning?</td>
</tr>
<tr>
<td>concepts and referents R</td>
<td></td>
<td>What is included in the concept?</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>What items can the concept refer to?</td>
</tr>
<tr>
<td>associations R</td>
<td></td>
<td>What other words does this make us think of?</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>What other words could we use instead of this one?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use</th>
<th>grammatical functions R</th>
<th>In what patterns does the word occur?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>In what patterns must we use this word?</td>
</tr>
<tr>
<td>collocations R</td>
<td></td>
<td>What words or types of words occur with this one?</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>What word or types of words must we use with this one?</td>
</tr>
<tr>
<td>constraints on use (register, frequency…) R</td>
<td></td>
<td>Where, when and how often would we expect to meet this word?</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Where, when and how often can we use this word?</td>
</tr>
</tbody>
</table>

Note: In column 3, R = receptive knowledge, P = productive knowledge.

The list makes use of the distinction between receptive and productive knowledge, and provides a specification of what components of depth are involved in receptive vis-à-vis productive knowledge. Accordingly, the three main groups of components, form, meaning and use, in turn are subdivided respectively into three components, involve different knowledge types depending on whether they are known receptively or productively.

The first main component form involves knowledge of pronunciation, spelling and word parts, i.e. inflectional and derivational affixes. Meaning consists of knowledge of form-meaning links and the various semantic relationships of words, such as polysemy, hyponymy and synonymy. The third component, use, includes knowledge of the grammatical patterns a word can fit in, its collocational profile and its register and frequency constraints. Accordingly, the model provides an elaborate and comprehensive description of the different knowledge types constituting depth of vocabulary knowledge. The model must be seen as an account of an ideal state of affairs and not an account of the amount of knowledge a learner typically has for any given word in her/his vocabulary.

The comprehensive nature of the list in that it both identifies a great number of components and also defines them in accordance with whether the language use is receptive or productive in nature, makes it applicable to
different research contexts. For instance, in a study in which depth of vocabulary knowledge is examined in relation to a productive language skill it might be more relevant to operationalize the productive components. Another aspect is that the all-inclusiveness of the list constitutes a good starting point for the study of vocabulary acquisition since it allows us to examine different components to reach an understanding of their internal relationship and whether they are involved in what it means to know a word qualitatively. Nagy and Scott (2000: 272) argue that it would be possible to reduce some aspects of word knowledge to a single continuum if research was able to show implicational relations between these aspects. In a similar vein Schmitt and Meara (1997) and Schmitt (1998) stress the need for studies that investigate the interrelationships among the various aspects of word knowledge. Schmitt and Meara (1997:19) note that “a better understanding of these interrelationships could go a long way toward developing an explanatory model of vocabulary acquisition”.

In view of Fitzpatrick’s (2006) use of this model for categorizing associational links, one might argue that there is no necessary contradiction between this model of depth in which several different word knowledge aspects are identified as parts of depth and viewing the mental lexicon as a network. Words are probably linked to each other in numerous ways so that every encounter with a word in different forms and contexts adds new links between the nodes. There is no evident reason to believe that only paradigmatic and syntagmatic word relations reflect the organization of the lexicon. One might very well acknowledge that many components are involved in depth and that these components, such as knowledge of word inflection and derivation to different degrees have implications for the rest of the network. The idea that many components are involved in depth of vocabulary knowledge does not mean that all components must be tested on one occasion. By limiting the number of components on any one test occasion, a larger sample of words might be tested. This also enables us to identify which components are relevant signs of the organization of the mental lexicon and the interconnection of the components. Once again, it seems a bit premature to equate the organization of the lexicon with paradigmatic and syntagmatic relationships between words.

2.6 Models of vocabulary knowledge and ability

In the past decade or so there have been attempts to define vocabulary knowledge in terms of more comprehensive frameworks that integrate different dimensions of vocabulary knowledge under one overarching construct, in an attempt to formulate a construct specification for vocabulary assessment.
The main rationale for these models is to provide researchers with a construct definition guiding the operationalization and measurement of vocabulary ability as well as explicating what dimensions of the construct results can be applied to.

2.6.1 Henriksen’s model of lexical competence

Henriksen (1999) defines lexical competence as consisting of three interrelated continua. The first continuum, the partial-precise dimension refers to the development of learners’ word meaning knowledge, which progresses form a broad knowledge of meaning to a more precise understanding of the meaning of a word.

It involves *item learning*, namely adding new items to the lexicon by creating extensional links. An item is pushed along the partial-precise continuum by a process of building a network. The degree of knowledge of this network is equated with depth of vocabulary knowledge and constitutes the second continuum. As a learner develops a word’s paradigmatic and syntagmatic relations to other words in the lexicon, more and more meaning characteristics become known, which leads to the given word moving along the partial-precise continuum.

The third continuum, receptive-productive, involves procedural knowledge as opposed to the two other continua which relate to declarative knowledge. It refers to the degree to which words can be used productively and it is posited that words move along this continuum as a word becomes more and more entrenched in the mental lexicon. Hence, the dimension denotes the extent to which a word is available for use or in other words the degree to which it has become automatized. Also, this dimension is closely related to the other two declarative dimensions in that development in them is seen as a crucial prerequisite for a word becoming productively known.

The two underlying ideas in Henriksen’s (1999) three continua of lexical competence are that of interdependency of the described dimensions and conceptualization of vocabulary knowledge as a continuum in which knowledge moves along different interrelated continua. However, it seems a bit superfluous to posit a dimension that is responsible for the labeling and packaging and another responsible for the network building process and in extension a word’s development along the partial-precise continuum. These two interrelated declarative processes could be carried out within one single dimension that involves both item learning and network building. Especially when we are dealing with older learners who come equipped with a well developed L1 lexicon and also possibly a L2 one, the distinction between intake (dimension 1) and system changing (dimension 2) is difficult to maintain. Another problem with this approach is that she does not specify whether the model describes the development of individual lexical items or the mental lexicon as a whole. Moreover, the model is somewhat simplistic.
in that it is too word-centered and does not touch on the larger picture of the context of learning and inter- and intra-learner variation in how vocabulary is acquired.

The main merit of this model is the emphasis on lexical competence as a developing system and on the interaction between different subcomponents of this system in learners’ vocabulary acquisition. However, Henriksen’s conceptualization of lexical competence seems to be limited to the developmental aspect of vocabulary ability and should not be perceived as model of vocabulary ability in the same sense as Chapelle’s (1994; cf. section 2.6.3) model which also includes the context of vocabulary use.

Henriksen’s (1999) model could be useful for guiding research which investigates the more minute details of lexical-item intake and which tracks the development of individual items within the three proposed dimensions. When dealing with the development of the whole lexicon or specific dimensions of it, this model is quite difficult to apply. The main prediction one can glean from it is that dimensions of vocabulary knowledge are highly interrelated. This issue has been investigated in a number of studies, some of which will be outlined in the next chapter. In regard to the present study the aspect of interconnectivity of dimensions of vocabulary knowledge will be used to discuss the results.

2.6.2 Daller et al.’s Lexical space

A second way of describing vocabulary knowledge is that of Daller et al. (2007), who suggest that we can conceive of it as a three-dimensional space, where each dimension reflects a facet of vocabulary knowledge.

![Figure 3. The lexical space: dimensions of word knowledge (from Daller et al 2007: 8).](image)

31
This model is not intended as a comprehensive definition of the broader construct of vocabulary ability, but rather as a metaphor for what Chapelle (1994) terms vocabulary knowledge and fundamental processes and constitutes an explanatory space for locating learners’ vocabulary knowledge. They exemplify how this model can be used in this way by providing the following analogy:

Some learners may have large vocabularies but are very limited in the speed and ease with which they can recall these words and put them to use communicatively. The learners ought to be placed well along the breadth axis but less far along the fluency or depth axes. Other learners may appear to have different characteristics and possess comparatively few vocabulary resources but considerable fluency in calling these to mind and using them in communication. These learners would occupy a different location in the lexical space, less far along the breadth axis but further along the depth axes. (Daller et al 2007: 9)

Although one might find such learners, the question is whether this analogy reflects a typical learner. As will be outlined in the next chapter, it seems to be the case that vocabulary knowledge dimensions for most learners are highly interrelated and it is fairly uncommon among learners exhibiting a well developed vocabulary knowledge dimension at the same time as other dimensions are relatively undeveloped. As stressed by Daller et al. (2007) this model should not be understood as a detailed model which can be empirically tested.

2.6.3 Chapelle’s model of vocabulary ability

One of the first proposed definitions of vocabulary ability was that of Chapelle (1994), which includes both vocabulary knowledge and the ability to put that knowledge into use, hence, the term vocabulary ability. The key aspect of this model is the emphasis on vocabulary ability “as a capacity for language use in context” (Chapelle 1994: 163). Drawing on Bachman’s (1990) interactionalist construct definition of communicative language ability, Chapelle’s definition of vocabulary ability encompasses vocabulary knowledge and the mental processes required to access that knowledge as well as the context in which vocabulary is used. A third component is posited which controls the interaction between the two. These three components make up vocabulary ability:

- **The context of vocabulary use**
- **Vocabulary knowledge and fundamental processes**
- **Metacognitive strategies for vocabulary use**
2.6.3.1 The context of vocabulary use
From a testing perspective context refers to the sentence or structure surrounding the target word. In this sense context is supposed to reflect natural language use.

From a communicative perspective, which is the perspective held by Chapelle, context is more than linguistic context. Vocabulary use is affected by the social and cultural situation. The main underlying idea is that context can affect lexical meaning, and knowledge of this needs to be part of learners’ vocabulary ability. For instance learners need a different kind of vocabulary ability to take part in an informal conversation with friends than in a formal political debate on a given issue. Moreover, it is important to define vocabulary ability in relation to the context in which it is examined.

In this study, I have followed this recommendation in two ways by operationalizing vocabulary knowledge and by selecting tasks for examination in relation to the academic setting. The first point is achieved by investigating learners’ depth of productive knowledge of academic words. Since the academic learning context place higher demands upon learners’ vocabulary knowledge, it was deemed as relevant to operationalize depth as quality of productive knowledge of academic words. The second is realized by examining authentic texts which are part of students’ academic examination. These two steps enable one to explore the relationship between vocabulary knowledge and vocabulary knowledge in light of the specific learning context of higher learning.

2.6.3.2 Vocabulary knowledge and fundamental processes
The second component in Chapelle’s (1994) framework consists of four subcomponents: vocabulary size, knowledge of word characteristics, lexicon organization, and fundamental vocabulary processes.

The first sub-component Vocabulary size, simply denotes the number of words a person knows within a certain context. According to Chapelle (1994), a learner’s size of vocabulary knowledge should not be measured in an absolute sense but instead should be defined and assessed in reference to a particular context of vocabulary use. In the present study, this is done by using a vocabulary size format that includes both low frequency vocabulary and academic vocabulary, the knowledge of which is important in an academic setting.

Knowledge of word characteristics, the second sub-component, is very similar to the notion of depth of vocabulary knowledge as defined by Nation (2001) in that it includes knowledge of phonemic, graphemic, morphemic, syntactic, semantic, pragmatic and collocational features.

The third sub-component, lexicon organization, refers to how lexical items are organized in the mental lexicon. Viewing the present study from the perspective of Chapelle’s (1994) distinction between depth of vocabulary
knowledge and lexical organization the test employed can be said to only measure the former, i.e. knowledge of word characteristics (here termed depth of vocabulary knowledge). However, it must be pointed out that it is difficult to understand exactly how the two dimensions are different from each other since a well developed knowledge of word characteristics dimension is tantamount to a well organized lexicon. Also from an assessment perspective it is difficult to envision a test which only taps lexicon organization without also tapping knowledge of word characteristics, since, as discussed in section 2.5.1.2, tests designed to measure network knowledge utilize items that measure test-takers ability to link words with each other, i.e. precisely the knowledge aspects contained in the latter dimension (i.e. knowledge of word characteristics).

The fourth sub-component, fundamental vocabulary processes, involves the processes associated with lexical access and includes the following aspects: productive and receptive use of vocabulary knowledge and being able to parse and compose words morphologically.

2.6.3.3 Metacognitive strategies for vocabulary use
The third main component denotes the mostly subconscious strategies we use to handle vocabulary in communication. These become conscious only when we are involved in unfamiliar or cognitively demanding communication tasks, e.g. talking with a foreigner, or deciphering illegible handwriting or having to carefully choose one’s words in an argument. These strategies are particularly important for less proficient learners who must devise and execute plans for achieving their communicative goals despite a limited vocabulary.

2.6.4 Summary
Three comprehensive frameworks of vocabulary ability and vocabulary knowledge have been outlined. The models of both Chapelle and Henriksen stress the interconnectivity of different components of vocabulary knowledge and ability. Daller et al. (2007), however, seem to make the case on the basis of their model that among some learners there might be an imbalance between the proposed fluency axis and the size axis. Moreover, Henriksen’s (1999) model can be distinguished from the other two in that her model focuses on the developmental aspect of vocabulary ability.

According to Chapelle (1994) context of use is a crucial notion when studying vocabulary knowledge. In the present study, the conceptualization of depth of vocabulary knowledge has been selected with reference to advanced learners in an academic language learning context. Accordingly, in view of the higher cognitive demands and the more complex productive tasks such as essay writing, which EFL learners studying English at the tertiary level are subjected to, it was deemed necessary to employ a
conceptualization of depth in which depth is defined with reference to productive knowledge. Moreover, the rationale behind the selection of the VLT format as a measure of size of vocabulary is based on the context in which vocabulary is examined. The fact that it has a productive version, and also that it measures knowledge of both low frequency words and academic vocabulary are two main reasons for including it in a study of informants studying at the tertiary level. Another reason is that vocabulary is examined in relation to the ‘productive’ skill of writing, or more precisely the use of advanced vocabulary in academic writing tasks.

The main differences between Henriksen’s and Chapelle’s models lie in how depth of vocabulary knowledge is perceived. Whereas Henriksen (1999) equates depth with lexical organization, Chapelle (1994) defines depth along the lines of the components approach and thus perceive lexical organization as a separate dimension. As discussed in section 2.5.2 this view is in line with the argument that there is no necessary opposition between the components approach and seeing the organization of the lexicon as a network.

Chapelle’s (1994) model captures to a greater extent the communicative aspect of vocabulary ability by positing a dimension of context and metacognitive processes of use, which both interact with and govern learners’ vocabulary knowledge and fundamental processes. Furthermore, Chapelle (1994) adopts an interactionalist construct definition of vocabulary ability, which includes a declarative knowledge component and the capability for implementing that knowledge in language use in context.

Again it must be stressed that none of these models are intended as predictive frameworks which can be tested empirically. The main contribution of these models is that they enable us to navigate between and put labels on different components of a very rich and complex construct. In order to be able to formulate a more detailed model of vocabulary acquisition, we have to each contribute a small piece of the puzzle and these models can be seen as descriptive frameworks enabling us to get an idea of how the pieces of the puzzle might fit.

2.7 Measures of vocabulary use

Whereas the previous sections have been concerned with outlining ways of describing vocabulary knowledge and the wider construct of vocabulary ability, in this section we are going to deal with ways of measuring vocabulary use. As noted in section 2.1 a distinction is made in the current study between a learner’s vocabulary knowledge and the actual use of that knowledge in authentic communicative settings. This distinction is an important one to make, because in authentic communicative language tasks there are many intervening factors between a learner’s vocabulary
knowledge and the kind of vocabulary he or she makes use of. For instance, a learner with a well developed vocabulary knowledge might in certain language tasks make use of less sophisticated vocabulary. This distinction is also reflected in the way the two aspects have been tested with discrete/selective/context-independent measures used to measure learners’ vocabulary knowledge or ability and with embedded/comprehensive/context-dependent measures used to measure learners’ vocabulary use (Read 2000).

In the next section different ways that learners’ vocabulary use in writing tasks have been measured will be outlined and discussed.

2.7.1 Measures of learners’ communicative vocabulary use
There are a number of measures of learners’ vocabulary use in writing. Read (2000) makes a basic distinction between qualitative and objective approaches to assessing learners’ vocabulary use in writing. Rating scales are an example of the former approach and can be further subdivided into global and analytic rating scales. The former employs a rating scale that provides a description of several aspects of language use without a specific reference to lexical quality. The latter involves the use of several scales each focusing on one aspect of writing ability, such as vocabulary, content and organization.

Objective measures involve the counting of various statistics that reflect learners’ use of vocabulary in their written production. The general term describing the features measured by these statistics is lexical richness (Read 2000). A further distinction is made between word-list and word-list-free approaches to measuring lexical richness in oral and written texts (Daller and Xue 2007). These will be outlined and discussed in the next two sections.

2.7.1.1 Word-list-free approaches to measuring lexical richness
The common denominator of word-list-free approaches is that the degree of lexical richness is determined on the basis of different word distributions and word type proportions in the texts analyzed without using extra-textual reference points such as word-frequency lists to establish the degree of lexical richness. In other words word-list-free approaches involve the assessment of how well a learner uses the vocabulary s/he knows, not what type of vocabulary is used as defined by a pre-existing word list (Laufer and Nation 1995). One of the most common word-list-free methods is the type/token ratio (TTR) or lexical variation which shows the ratio in percent between the number of different words in a text and the total running words. Another common measure is lexical density which is the proportion of content words, in relation to function words (Read 2000).

One particular limitation of the TTR is that it is sensitive to text length. Although a number of efforts have been made to overcome this limitation such as Malvern and Richards’ D formula, type/token-based measures, have
been strongly criticized as unreliable (e.g., van Hout and Vermeer 2007; McCarthy and Jarvis 2007; Vermeer 2000).

The main reason for not employing a word-list-free method of assessing lexical richness is that the present study seeks to establish whether type of vocabulary, as defined by frequency, learners use in their writing is related to other dimensions of vocabulary knowledge and academic success in view of the assumption that the writing task of take-home essays as opposed to timed compositions facilitate the use of more advanced vocabulary. In other words whereas word-list free approaches focus more on the diversity of vocabulary use, word-list approaches focus more on the sophistication of vocabulary use. Moreover, as pointed out by Meara and Bell (2001) one gains much more insight into lexical richness by assessing the quality of learners’ vocabulary use by means of extrinsic criteria, such as word frequency lists. The following example of Meara and Bell (2001) illustrates the inherent weakness of the TTR approach:

- example 1: the man saw the woman.
- example 2: the Bishop observed the actress.
- example 3: the magistrate sentenced the burglar.

Although these sentences are different in terms of the quality of vocabulary a TTR approach will produce the same results for each of these examples.

2.7.1.2 Word-list approaches to measuring lexical richness
Having outlined and discussed word-list-free approaches, this section will provide an outline of word-list approaches. In the next two sections two word-list measures of lexical richness will be presented and evaluated.

2.7.1.2.1 The P_Lex
Similarly to the Beyond 2000 measure (described in sections 2.7.1.2.2 and 4.2.2), the P_Lex measures the lexical sophistication of texts in terms of the proportion of difficult words defined as beyond the 2000 most frequent words in English. It can be used on relatively short texts. The programme divides the text into ten-word segments and constructs a profile in the form of a curve showing the number of difficult words per segment. The data fed into the programme is matched to a theoretical curve on the basis of the Poisson distribution formula. This enables the programme to assign a lambda value ($\lambda$) to a profile which typically ranges from .5 to about 4.5. Higher and lower are possible but very infrequent. The lambda value can be used as a measure of the lexical richness of a text (Meara 2009). The higher the lambda value, the lexically richer a text is.

The measure has been shown to have an acceptable level of test-retest reliability and some evidence of concurrent validity, i.e. a modest correlation with the RVLT (Meara and Bell 2001).
2.7.1.2.2 The Lexical frequency profile /Beyond 2000 measure

The Lexical Frequency Profile (LFP) was developed by Laufer and Nation (1995) and shows the percentage of words learners use at different vocabulary frequency levels in their written texts. Briefly, the LFP is able to calculate lexical richness through a computer programme termed **Range** (available at [http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx](http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx)). The programme exists in two versions. The first is based on three word family lists: the first 1,000 most frequent words (word list one), the second 1,000 most frequent words (word list two) (both based on the frequency counts of the GSL) and the AWL (word list three). For all the words that fall outside these lists only number of tokens and types and their proportion is calculated since no word family list is available for these words. These words are termed ‘not in the lists words’. The second version is based on the BNC 10 million token spoken corpora and consists of 14 lists with 1,000 word families each, which are organized in accordance to frequency of occurrence with the first list containing the 1,000 most frequent words in the corpora. There are two additional lists 15 and 16 that contain proper nouns and most interjections, exclamations, hesitation procedure, etc., which are common in spoken English. The two versions differ on another point as well. In the word lists for the BNC version the word families are set at level 6 (cf. section 2.4.1.1). In the word lists for GSL/AWL-version the word family level is set at 3. Although, level 3 reflects to a greater degree than level 6 what would be perceived as the same words for productive skills, Nation (2006) suggests that the lemma would be the most sensible unit to use. This might affect the validity of the instrument, if the scores are used as indicators of the writer’s vocabulary size available for use.

The idea behind LFP is that it provides a calculation of the proportion of words belonging to different frequency levels. The higher the proportion is of AWL words and ‘not in the lists words’ in a written text, the higher the degree of lexical richness of that text. Laufer and Nation (1995: 312) provide a simple example of how the LFP works.

Let us imagine a composition of an intermediate learner which consists of 200 word families. Among the 200, 150 belong to the first 1,000 most frequent words, 20 to the second 1,000, 20 to the AWL and 10 are not in any list. To calculate the LFP, we convert these numbers (the numbers of word families at each frequency level) into percentages out of the total of 200 word families. The LFP of the composition is therefore 75%-10%-10%-5%.

The LFP shows the proportion of four types of words. There is a condensed profile called the Beyond 2000 (B2000) which distinguishes between the first two thousand words and the academic and not-in-the-lists words. Since the B2000 can be used to generate one single measure showing the percentage of the beyond-2000 words in written samples, it facilitates...
correlational analyses and will therefore be the measure used in the present study.

If we compare some of the concurrent validity figures obtained for the LFP with those of the P_Lex, the LFP seems to have somewhat higher degree of concurrent validity. For instance, Meara and Bell (2001) report a correlation between P_Lex and PVT scores at between .34 and .57 as compared to .6-.8 for the LFP (Laufer and Nation 1995). More recent evidence has also shown that the P_Lex has a weaker concurrent validity than the LFP. Daller and Xue (2007) found that LFP scores of oral texts by 50 Chinese EFL students correlate at a modest to high level with three out of four examined measures of lexical richness, whereas the P_Lex only shows a weak correlation with one out of four. It should be pointed out that Tidball and Treffers-Daller (2008) found that the French version of the LFP was not able to discriminate between the oral production of learners at different levels of proficiency.

The main advantage P_Lex has over the LFP measure is that it is a better measure of shorter texts typically extractable from lower-level learners. Although the LFP has its inherent weaknesses it has stronger concurrent validity, it has been more widely used in a number of studies, and it has been shown to provide reliable information about the level of sophistication of learners’ vocabulary use and it has also been shown to discriminate between the writing of learners at different proficiency levels (e.g. Andrews 2009; Coniam 1999; East 2006; 2009; Goodfellow, Lamy and Jones 2002; Horst and Collins 2006; Lee and Muncie 2006; Lenko-Szymanska 2002; Li 1997; Morris and Cobb 2004; Lim and Galaczi 2010; Vidaković and Barker 2010).

Since the LFP is employed as a measure of lexical richness in the current study, further details about the measure are provided in section 4.2.2.
3. Previous work

The previous chapter discussed relevant components that comprise the overall construct of vocabulary knowledge and how these have been measured. This chapter provides an account of and discusses previous work on learners’ vocabulary knowledge from the perspective of the relationship between facets of vocabulary knowledge and their development over time. The chapter is divided into four main sections focusing on the relationship between learners’ size and depth of vocabulary knowledge, the relationship of different components of vocabulary knowledge to lexical use, the relationship of vocabulary knowledge and use to L2 academic achievement, and the development of the specific components of vocabulary knowledge.

Each section is concluded by a set of research questions which will be further investigated (here). These research questions expand on the five main research questions presented in section 1.2.

3.1 Learners’ size and depth of vocabulary knowledge

This section will deal with empirical studies examining the relationship between size and depth of vocabulary knowledge and the interrelationship between different components of depth of vocabulary knowledge.

3.1.1 Relationship between size and depth of vocabulary knowledge

As mentioned in section 2.3, vocabulary knowledge has been described in the research literature as having two primary dimensions: size and depth. A large and growing body of literature has investigated the relationship between learners’ size and depth of vocabulary knowledge. An understanding of how size is related to depth among different learner types can inform researchers and teachers about what constellation of measures is appropriate for obtaining a picture of the overall state of learners’ vocabulary knowledge. In many testing situations, there are limits to the number of tests that can be used to this end. Therefore an insight into how different dimensions of vocabulary knowledge are interrelated among a specific group of learners and across different stages of vocabulary development might
improve the effectiveness and practicality of vocabulary knowledge assessment. Solely establishing that there is a strong relationship of size to depth without specifying kind of knowledge, i.e. receptive or productive, type of learner or learning context, limits the applicability of the results to real assessment situations.

In regard to empirical findings on the nature of the relationship of size to depth, it has been shown in a number of studies that there is a strong relationship between having a large vocabulary size and qualitative word knowledge (e.g. Greidanus, Boogards, van der Linden, Nienhuis and de Wolf 2004; Qian 1999; 2002; Qian and Schedl 2004). However, in order to gain an insight into how learners’ vocabularies develop as their overall L2 language proficiency improves we need to know what the relationship between size and depth looks like at specific stages of learners’ L2 proficiency development. Further light should be shed on whether we would typically expect a stronger relationship between size and depth among advanced learners than among beginner or intermediate learners of the same age. We also need to consider whether different learning contexts might affect the way the relationship between size and depth develops. Many studies into the relationship between adult learners’ size and depth of vocabulary knowledge provide somewhat unclear information about the overall proficiency level of the informants and/or the learning context in which the L2 is acquired, which makes it difficult to apply these findings to real vocabulary assessment contexts, such as the university setting investigated here.

One case in point is Greidanus et al (2004) who found a correlation of .79 between the scores on a receptive vocabulary size test and a receptive depth measure. The informants are reported to be advanced Dutch learners of French on the basis of having received 2-3 hours of instruction in French a week over a period of 6 years. However, since the Netherlands is arguably a low exposure learning context as far as French is concerned (cf. section 3.3.1), one is inclined to question whether the informants should be described as advanced learners of French. Accordingly, these results indicating a strong relationship between adult learners’ size and depth of vocabulary knowledge are difficult to assign to any specific proficiency level.

Another example is Qian (1999) who found a similar relationship between learners’ size and depth. In contrast to the informants in Greidanus et al. (2004), the informants in Qian (1999) were all learning English at the time of the study in a second language learning context (cf. section 3.3.1). However, since the vast majority of the informants had only been in Canada for between one and six months and since the only proficiency level information reported is that the informants have a vocabulary size of at least 3000 word families it is difficult to determine the informants’ proficiency level more precisely.
In Qian 2002 and Qian and Schedl 2004 the results suggesting that there exists a strong relationship between adult learners’ size and depth are based on a relatively large sample, 217 and 207 informants, respectively. Moreover, although both do specify the proficiency level of the informants as intermediate and beyond, the results reported reflect the whole sample, which prevents information of how the relationship between size and depth might vary according to proficiency level. Accordingly, one criticism of many of the studies is that they either provide vague information about the informants’ proficiency level or that different proficiency levels are lumped together making it difficult to ascertain whether the relationship between size and depth is equally strong among learners at different proficiency levels.

It should be pointed out that the main purpose of these studies was not to investigate the relationship between size and depth. For instance, in the case of Qian (1999; 2002) the aim of the studies was to examine the relative contribution of the size and depth of vocabulary knowledge of adult learners to reading comprehension. Nevertheless, we are presented with data suggesting a strong relationship between size and depth in adult learners. The fact remains that based on the data reported by these four studies, we do not know whether the relationship between the two dimensions is stable across proficiency levels and learning contexts.

There is some evidence indicating a weaker relationship between size and depth among less proficient learners than among more proficient learners, which would suggest that we cannot predict their depth of vocabulary knowledge solely on the basis of their relative vocabulary size. Among less proficient learners, it seems that one would have to resort to employing a wider set of vocabulary measures for the assessment of the learners’ overall state of lexicon. For instance, Nurweni and Read (1999) found among Indonesian university learners of English (N = 324) that the relationship between depth and size varied according to their sub-level proficiency defined as their achievement on a national examination in English for high school students. Sub-level proficiency is used here to refer to proficiency differences within broad categories like “beginner”, “intermediate” and advanced”.

The sample consisted of 324 first year students from different faculties. Based on their average vocabulary size, overall, the informants seem to have been at a low to intermediate level of English proficiency, although with a fair amount of variation. The informants were divided into three subgroups: High, Middle and Low. In the case of the High level students there was a correlation of .81. For the Middle and Low level students a moderate (.43) and a weak (.18) correlation, respectively, were found between the two measures. However, these results are difficult to interpret because the subgroups could either be interpreted as reflecting three sub-levels of low-proficiency learners or three distinct proficiency levels, i.e. low-, intermediate and high-proficiency levels. Accordingly, the results could
either be conceived of as indicating that among low-proficiency learners the relationship between size and depth varies according to sub-level of proficiency or that the relationship simply varies according to overall proficiency. The limited overall vocabulary knowledge of the informants makes it difficult to determine how the High level group would compare to learners in other learning contexts who are more exposed to English and who on average have a higher level of proficiency, such as Swedish university students of English.

Two other studies that have investigated whether the relationship of size to depth varies according to sub-level proficiency is Staehr-Jensen (2005) and Akbarian (2010) who found evidence among university students of English consistent with the findings of Nurweni and Read (1999). These results indicate that while the depth and size of relatively advanced learners are more integrated, the dimensions are more distinct among relatively less proficient learners (Akbarian 2010). From an assessment perspective, employing a more practical size test for obtaining a picture of learners overall state of vocabulary knowledge, i.e. size and depth, is only appropriate among more advanced learners as they could be expected to perform more consistently across different measures.

There is some contradictory evidence of a stronger relationship of size to depth among less proficient learners. Zareva (2005) found a stronger relationship of size to depth among less proficient learners among Bulgarian learners of English. However, the estimates of size and depth employed by Zareva (2005) suffer from being derived from the same test. Accordingly, the high correlations between the scores of two tests, i.e. the size and the depth tests may be an artifact of the measure employed. Another study which has shown a stronger relationship of size to depth among less proficient learners than among more proficient learners is Albrechtsen, Haastrup and Henriksen (2008). The informants in this study consisted of Danish EFL learners at three different educational levels: grade 7 of comprehensive school, grade 10 of upper secondary school and university undergraduates of English. The results showed a progressively weaker significant correlation of size to depth as the learners’ proficiency increases – beginner: .852, intermediate: .689 and advanced: .546.

Several studies have produced estimates of the nature of the relationship of size to depth, but so far little attention has been paid to establishing the nature of this relationship among more specific types of learners and in an authentic learning context. Those studies that have looked at the relationship among learners at specified proficiency levels have produced somewhat inconsistent results of the exact nature of the variation of this relationship in relation to proficiency level or sub-level proficiency.

In addition to obtaining an insight into whether it varies according to developmental factors, such as proficiency level, we need to identify other sources of variation in the relationship. There is some evidence which
indicates that the relationship of size to depth varies according to kind of vocabulary knowledge, i.e. receptive or productive knowledge. Intuitively, one would expect that there is a stronger relationship between the same kind of knowledge e.g. between productive size and productive depth. Again from an assessment perspective it is of importance to have an idea about what kind of depth of vocabulary knowledge can be predicted from different size tests. The relationship of size to depth as a function of type of knowledge has mainly been investigated from the perspective of receptive size. For instance, Schmitt and Meara (1997) found a weaker relationship of the same kind of knowledge among receptive size and receptive depth operationalized as knowledge of word associations than between different kinds of knowledge, namely receptive size and productive depth. Another study that has looked at the relationship of receptive size to depth according to kind of knowledge is Albrechtsen et al. (2008) who also found evidence of a stronger relationship between different kinds of knowledge i.e. receptive size and productive depth than between same kind of knowledge, i.e. receptive size and receptive depth. Albrechtsen et al (2008) suggest that the paucity of statistically significant correlations between the receptive size test and the receptive depth test scores is explained by differences in sensitivity of the three tests employed, i.e. a receptive size, a receptive depth and a productive depth measure. Both depth tests target higher frequency levels than the RVLT. Moreover, the informants were given more time to complete the receptive depth test than they were to finish the productive one. The somewhat counter-intuitive findings of Schmitt and Meara (1997) and Albrechtsen et al. (2008) may be artifacts of the test administration.

In addition to knowing how the relationship might vary according to factors such as proficiency level and kind of knowledge, it is also important to know how size is related to different components of depth. An understanding of the nature of the relationship of size to different elements of depth might tell us what kind of skills we can predict from learners size.

If we look at how size is related to different components of depth, receptive size has been shown to be strongly related to knowledge of synonyms and collocations (e.g. Staehr Jensen 2005; Gyllstad 2007). Schmitt and Meara (1997) found that receptive size was moderately related to word association knowledge and weakly related to verbal suffix knowledge. These studies suggest that size is at least a reasonably good predictor of receptive knowledge of word associations, such as collocations and synonyms, but not a very good predictor of word derivation knowledge. However, there is some evidence of a modest relationship of receptive size to receptive knowledge of word derivations. For instance, Mochizuki and Aizawa (2000) found a modest relationship between these two dimensions among Japanese high-school and university learners of English. However, although this study helps in shedding light on how size is related to depth,
again different proficiency levels are lumped together, which make it difficult to apply the results to any specific group of learners.

Another issue relating to the validity of the construct of depth is the extent to which different components that have been indentified as constituting elements of the construct are indeed interrelated. An understanding of the degree to which different elements of depth are interrelated might contribute to the formulation of a clearer definition of depth i.e. a specification of the quality of depth. This knowledge might also aid test administrators isolating the smallest and most practical predictors of depth of vocabulary knowledge.

Most studies that have found a strong relationship between size to depth have operationalized depth as network knowledge. Encapsulating depth in a single measure does not tell us much about how size is related to specific elements of depth. Moreover the relationship has mainly been examined at the receptive end of the receptive/productive continuum. Read (2004: 223) points out that:

A broader range of measures is needed before we can be more confident about the extent to which depth in some sense makes a contribution to the assessment of the lexical knowledge of L2 learners.

In accordance with Read’s recommendation, in the present study depth will be conceptualized according to the components approach. By doing this one is able to investigate depth of vocabulary knowledge in terms of how different components relate to other dimensions of vocabulary knowledge and also how these are interrelated. Schmitt and Meara (1997: 19)\(^4\) state that “a better understanding of these interrelationships could go a long way toward developing an explanatory model of vocabulary acquisition”. Although they only found a weak to modest relationship between the informants’ derivative suffix knowledge and word association knowledge, they argue that different components of word knowledge should be perceived as interrelated. Schmitt (1998), however, suggests that the strength of the relationship varies between different word knowledge types, even to the extent of some being unrelated.

Summing up, on the whole, the studies reviewed have found a strong relationship of size to depth among adult university learners of English. It is not clear, however, whether the nature of this relationship varies according to factors such as the proficiency level of learners. There seems to be evidence of a variation in the relationship of size to depth according to proficiency level (cf. Akbarian 2010; Nurweni and Read 1999, Staehr Jensen

\(^4\) It should be pointed out that in later works Meara (e.g. Meara and Wolter 2004) argues for perceiving vocabulary knowledge as consisting of two dimensions, size and organization, and that the components approach is not a productive way of conceptualizing vocabulary knowledge.
2005). However, these findings are somewhat contradictory (cf. Albrechtsen et al. 2008).

In addition to proficiency level, another identifiable factor of variability in the relationship of size to depth is type of knowledge, i.e. receptive or productive knowledge. Studies have reported somewhat counter-intuitive results of a stronger relationship of receptive size to productive depth than between the same kind of knowledge receptive size and receptive depth (e.g. Albrechtsen et al. 2008; Schmitt and Meara 1997).

As noted in section 2.5 it is a moot point whether we can treat depth as a single construct rather than a set of separate components (cf. Milton 2009). In order to further increase the analytic value of the cover term depth, we need to shed light on what different word knowledge types are covered by the term. One way would be to examine the interrelationship of components that have been isolated as components of depth. There is some evidence of a strong relationship between at least two such components, namely knowledge of synonyms and collocations (Staehr Jensen 2005). Another component that has been identified as a component of depth of vocabulary knowledge is knowledge of word derivations. However, there is only weak evidence of it being related to other possible components of depth (Schmitt and Meara 1997).

Apart from Schmitt and Meara (1997), none of the studies mentioned have examined how the relationship between size and depth develops for the same individual over time. Moreover, as noted above, depth has mainly been investigated at the receptive end and conceptualized as network knowledge.

The following four research questions address the relationship between size and depth as well as the interrelationship between components of depth among Swedish university students of English:

3. How are advanced adult EFL learners’ size and depth of vocabulary knowledge related?
4. How does the relationship between advanced EFL learners’ size and depth of vocabulary knowledge develop over time?
5. To what extent are different components of depth interrelated?
6. How does the interrelationship between components of depth change over time?

This review suggests that we should expect a strong relationship of size to depth. There is inconclusive evidence of how this relationship might vary according to proficiency level. However, there seems to be more data pointing to a stronger relationship between the two with increasing proficiency, which indicates that we should expect a similar development longitudinally (Akbarian 2010; Nurweni and Read 1999; Schmitt and Meara 1997; Staehr Jensen 2005).
Studies that have investigated the interrelationship of components of depth conceptualized according to the components approach have found a modest to high interrelationship. A modest to high interrelationship among the components examined here is thus anticipated. As suggested by Schmitt and Meara 1997, a stronger interrelationship between the components with increasing proficiency is expected.

In regard to the relative relationship that receptive and productive size displays with productive depth, we should expect to find a stronger relationship between the same kind of vocabulary knowledge i.e. productive size and productive depth.

3.2 The relationship between vocabulary knowledge and vocabulary use in writing

This section focuses on research that has investigated the relationship between vocabulary knowledge and use. This relationship has mainly been studied from the perspective of vocabulary size. In regard to vocabulary use, the studies reviewed are limited to those examining the construct from the point of view of lexical richness in learners’ written texts.

3.2.1 The relationship of vocabulary knowledge to the lexical richness of learners’ written output

Much research has highlighted the importance of vocabulary knowledge in L2 language use (e.g. Alderson 2005; Mecarty 2000; Staehr 2005; Zimmerman 2004). Milton (2009) suggests that a requirement for learners to be able to improve their L2 communicative skills is an increase in their vocabulary knowledge. A prerequisite for fully revealing the role of vocabulary knowledge in L2 communicative competence is to shed light on the scope of its contribution to various language skills. From an assessment perspective, it is of practical relevance to uncover the extent to which specific vocabulary knowledge dimensions are reliable indices of various types of language use in authentic tasks. Such knowledge might inform teachers and researchers about what dimensions of learners’ vocabulary knowledge should be assessed for obtaining a rough indication of how well they might be expected to execute different language tasks. Moreover, such insight might explicate what facets of learners’ vocabulary knowledge should be targeted for focused vocabulary instruction in order to improve the quality of L2 language use in specific tasks.

The ability to use a varied and sophisticated vocabulary has been acknowledged in the research literature as an important facet of language use (e.g. Engber 1995; East 2009; Nation 2001; Tidball and Treffers-Daller
Lexical richness (cf. 2.6.1) measures have frequently been used as practical and straightforward indices of quality of vocabulary use in the spoken and written output of learners. This has prompted much research into the relationship between vocabulary knowledge and lexical richness of learner output.

Scores on receptive measures of vocabulary size have been found to correlate positively with different lexical measures of written compositions (e.g. Arnaud 1984; 1992; Goodfellow, Lamy and Jones 2002; Jarvis 2002; Llach and Gallego 2009). In general these studies have found a modest correlation between receptive vocabulary size and lexical richness of timed compositions. One should expect a higher correlation between scores for productive vocabulary size and scores for lexical richness. However, there seem to be somewhat conflicting results in regard to the relationship between productive vocabulary size and lexical richness.

For adult EFL/ESL learners, Laufer and Nation (1995) found modest to high correlations between scores on the PVLT and LFP scores whereas Laufer and Paribakht (1998) found only a modest correlation between the two variables for the same age-group. Moreover, Laufer (1998) who, examined, among other things, the relationship between productive vocabulary size and lexical richness among adolescent EFL learners found no significant correlation. This led her to conclude that “[b]etter passive and controlled active vocabulary do not seem to be reflected in free production” (Laufer 1998:267). She proposes that this might be due to risk-avoiding on the part of the informants in that they prefer relying on high-frequency words that are more deeply entrenched in their mental lexicon.

Although there are a great number of studies on the relationship of specifically size to learners’ sophistication of vocabulary use, this has mainly been examined in the context of restricted writing tasks, such as in Laufer (1998). In addition, there seem to be conflicting results in regard to the relationship between productive vocabulary size and lexical richness (e.g. Laufer and Nation 1995; Laufer and Paribakht 1998; Laufer 1998). All the studies have examined the relationship between vocabulary size and lexical richness in timed compositions rather than those without time-limits (see section 3.3.3 for a discussion of differences between timed compositions and take-home essays). Moreover, all look at size and there are no documented studies that have examined the relationship between depth of vocabulary knowledge and lexical richness.

The question how different facets of vocabulary knowledge are related to vocabulary use can provide valuable insight into what kind of vocabulary learning and focused vocabulary instruction is required for learners to incorporate more advanced words into their written production. In terms of assessment, by exploring the nature of the relationship of vocabulary knowledge to learners’ inclusion of less frequent vocabulary into their written production, we can obtain more effective and practical predictors of
the sophistication of learners’ vocabulary use. However, too little attention has been paid to the relationship of learners’ vocabulary knowledge and the incorporation of advanced words into their written production of authentic texts in the university setting. Accordingly, there is a need to further investigate the relationship of vocabulary knowledge to the lexical richness in this context in order to uncover its contribution to real vocabulary use. The nature of the relationship of vocabulary knowledge to lexical richness in authentic learner writing tasks will be addressed by the following research question:

7. To what extent are vocabulary knowledge (receptive and productive size and productive depth) and lexical richness of take-home essays related?
8. How does the relationship between vocabulary knowledge (receptive and productive size and productive depth) and lexical richness change over time?

Since a number of previous studies have found a modest relationship between size and lexical richness, we should expect a weaker albeit significant positive relationship of size to lexical richness of the task type examined here. The rationale for anticipating a weaker relationship is that the task type of take-home essays offers the learners the possibility of compensating for a small vocabulary size or a heavy reliance on high-frequency vocabulary in their use, such as the time to tend to the lexical richness of the output and making use of dictionaries.

In regard to the relationship of depth to lexical richness, Bulté et al. (2008) suggest that lexical sophistication, which is the type of lexical richness examined here (cf. 2.7.1.1), should primarily be perceived as an index of qualitative word knowledge. On the basis of their construct definition, we should also expect to find a positive relationship of depth to lexical richness among the sample investigated here.

3.3 The relationship of learners’ vocabulary knowledge and use to L2 academic achievement?

Although there is a consensus in the research literature that there is a strong relationship of vocabulary knowledge and use to academic performance, few studies have actually investigated this in terms of looking at the relationship directly (cf. Milton 2009). Instead this consensus is to a large extent based on studies that have looked at the relationship from the perspective of indirect measures of academic achievement such as reading comprehension etc. In the light of the importance of vocabulary size to reading
comprehension a number of studies have set out to investigate whether learners in different learning contexts reach relevant vocabulary learning goals.

3.3.1 The assessment of learners’ size of vocabulary knowledge in relation to vocabulary goals and thresholds

3.3.1.1 Learners’ size of receptive vocabulary knowledge
Size of vocabulary knowledge is widely recognized as having a strong relationship to reading comprehension. Accordingly, learners’ size of vocabulary knowledge has mainly been examined from the point of view of whether it reaches vocabulary size goals judged as necessary for appropriate reading comprehension. Since a threshold of 3000-5000 word families has been identified as a crucial minimum for comprehending fiction and academic books a number of studies have focused on this interval. The 3000 word threshold should be regarded as the absolute minimum for reading less advanced texts (Laufer 1992), whereas the 5000 word threshold should be seen as a minimum requirement for reading academic texts (Sutaraysh, Nation and Kennedy 1994). Although there exist more recent and higher estimates for how much vocabulary is needed for reading comprehension, these thresholds have influenced a number of earlier studies on whether learners have enough vocabulary to manage university studies in an L2 (cf. Nation 2006).

Although university students are usually taught through the medium of the national language in many countries, English language texts are required reading in many fields of study. The amount of reading and the type of reading material used in higher learning puts a particularly high demand on the students’ vocabulary knowledge. This has prompted a number of studies to investigate the size of English vocabulary knowledge of university students who are L2 users.

There has been a focus on learners’ knowledge of basic vocabulary and a learning goal in the middle of the 3000-5000 interval of 4000 words has been the point of reference in the investigation of whether university students know enough vocabulary to cope with university reading. A number of studies have obtained this goal from the national curriculum of the respective country in which the study is carried out (e.g. Nurweni and Read 1999; Barrow, Nakanishi and Ishino 1999; Tschirner 2004).

In order to investigate whether the average Indonesian student knows enough basic English words for independent reading of English academic texts at the university level, Nurweni and Read (1999) tested a large sample of Indonesian university students (N = 324) on the 2000 most frequent words in English in addition to the 800 words in the UWL (cf. section 2.4.1.2). The
rationale for setting the limit to only 2800 words is that these two lists cover around 92% of the running words in any academic text, the knowledge of which would yield a rough indicator of the informants’ degree of comprehension of academic texts used at the university (ibid.). Size of vocabulary knowledge was tested through a translation test in which the test takers translated the target word presented in an English sentence into their L1. The estimate of vocabulary size shows that on average the students had a vocabulary size of 1,226 words, thus falling short of the 4000 goal set by the national curriculum for upper-secondary school (ibid.).

Another study focusing on Asian university students’ knowledge of high-frequency vocabulary is Barrow et al (1999). The informants were Japanese college students and they were tested on the 4000 most frequent words sampled from the Japan Association of College Teachers (JACET) ‘4000 Basic Words’ word list. The JACET list has word form entries, which means that different forms of the same word have their own entry, e.g. *play, played*. Consequently, they used a different definition of a word than the word family one. In this study a Y/N test was used. The first year students showed a 75% familiarity with the test items and the second year students were familiar with 87% of the tested vocabulary. Accordingly, there seems to be an increase in vocabulary size over one year of study. The informants’ estimated knowledge of basic vocabulary was about 2300 “words”. By comparing the scores of the Y/N test and a meaning recall test of the same sample, Barrow et al (1999) found that the informants tended to overestimate their knowledge on the self-check test. Thus, suggesting that the average vocabulary size of 2300 “words” might in reality be slightly less. In both studies the vocabulary sizes are smaller than the authors had expected, which seems to suggest that regular testing of learners’ vocabulary size would be a good idea from a pedagogical point of view.

In a study conducted in a European context, Tschirner (2004) tested the vocabulary size of 142 German first term students of English Language and Literature at the University of Leipzig. Unlike Nurweni and Read (1999) and Barrow et al (1999) she also looked at their productive vocabulary size. The learning goals for upper secondary school graduates in Germany are 3000 words for productive vocabulary and as mentioned above 4000 words for receptive vocabulary size. Tschirner (2004) used the RVLT and the PVLT. With a passing rate set at 90% correct answers on the different levels, only 30% of the students were found to master the 5,000 word level in the

---

5Nurweni and Read (1999) do not specify what unit of counting words they used. However since they relate their findings to the 3,000-5,000 level, which marks the number of word families a learner must know in order to reach an adequate comprehension of general academic texts, I assume that they use the same unit of counting words, namely the unit of word family.

6 Tschirner (2004) does not specify the unit of counting.
receptive test and only 22% passed the 2,000 word level in the productive test. Accordingly the majority of the informants did not reach the vocabulary size goals required for university studies at a German university.

A study that investigated the receptive vocabulary size of students of English beyond the first year of study is that of Pérez Basanta (2005) who found that her students at the final year of English studies had an average estimated receptive size of 5500 word families. On average the students did not reach the cut-off point of mastery of the 5K level.

A study that assessed the vocabulary size of first-year students of English in the Scandinavian context is Albrechtsen et al (2008), who found that only 10 out of 19 Danish university students of English mastered the 5K level. They stress the “importance of having a well established sight vocabulary base of at least 5000 “words” for being able to operate in an L2” (Albrechtsen, et al 2008: 61). In addition, a majority of the first year students in Staehr Jensen (2005) were found not to master the 5K level. It is quite surprising that a majority of the advanced informants who are students of English did not master the 5K level. One would expect that adult learners in low exposure contexts such as those of Nurweni and Read (1999) and Barrow et al (1999) do not reach the 5K level, but not Danish adult learners who study English at the university level.

If we look at vocabulary size estimates of students of English in the Swedish context, evidence indicates that a higher vocabulary size is attained. Gyllstad (2007) reports on average his first-term students of English (N = 163) reached the cut-off point of mastery of the 5K level. The only level that the average student of English did not master was the 10K level even by the second (N = 49) and third term (N = 35) of study. He does not report the proportion of students who master the 5K level. However, the low standard deviation figures reported suggest that a majority of these mastered the 5K level already by the first term of study. Since the tests were administered at the end of each term, we have no way of knowing whether the first-term students of English mastered the 5K level already when entering university.

All the above mentioned studies were of foreign language learners at the tertiary level and the findings seem to suggest that many of the learners have limited receptive vocabulary knowledge. In contrast, advanced students in a second language context have been shown to reach at least near-native levels in terms of vocabulary size. Cervatiuc (2008) cites a study of 20 highly proficient non-native university students in Canada who were estimated to have an average vocabulary size of 16,512 basewords.

While it should come as no surprise that advanced learners in a second language context might reach high levels of vocabulary size, there is still insufficient data showing whether advanced EFL learners can be expected to achieve receptive mastery of the 5K level and beyond in an FL environment.

As noted above a number of studies have found that the vocabulary sizes of learners at the university level fall short of relevant vocabulary learning
goals (cf. Schmitt forthcoming). There are studies that have reported high vocabulary sizes among EFL learners, such as Zareva, Schwanenflugel and Nikolova (2005) who found very large vocabulary sizes among Bulgarian adult EFL learners enrolled on advanced and intermediate certificate preparation courses in English. No statistically significant differences were found between the advanced EFL learners (N = 17) and L1 graduate students (N = 30). The vocabulary sizes of the advanced group was estimated to about 8500 word families, whereas the intermediate group had an estimated vocabulary size of about 6000 word families. However, the paper would appear to be over ambitious in its claims. They used a somewhat unorthodox way of measuring vocabulary size which could explain the very high figures. In addition to using the VKS format to measure size, which is a common measure of depth of vocabulary knowledge, the test was administered as a take-home test. The inflated size figures might very well be an artifact of the test design.

It is also important to point out that a learner who has a high estimated total vocabulary size may not master all frequency levels within the frequency range of known vocabulary. A case in point is Cameron (2002), in which, although the informants had an overall vocabulary size of almost 7000 word families, they still did not master the 5K level. The results show that even ESL learners have gaps in their knowledge of high-frequency vocabulary compared to native-speaking peers. No significant differences were found between the two groups in their knowledge of very basic vocabulary represented by the 2K level and in academic and low frequency vocabulary. However, comparisons of mean scores using t-tests revealed significant differences at the 3K and 5K levels. Cameron (2002: 168) concludes that “both groups need attention to academic and low frequency vocabulary, but they bring to this ‘advanced language development’ different profiles of everyday vocabulary knowledge, and thus require different teaching strategies”.

In Table 3, we can see the estimated vocabulary sizes of different types of well-educated post-puberty learners found in the studies reviewed above.

Based on Table 3 one can make a rough categorization of the respective countries in terms of the degree of exposure to English. In order to be able to compare overall vocabulary sizes of learners of similar age, we need a framework for being able to make meaningful comparisons. For instance it would make no sense to compare the vocabulary sizes of Indonesian university students of English to that of Swedish university students. In order to establish a rough framework for making meaningful comparisons and for being able to interpret the size estimates of Swedish university students examined here, I follow Kachru’s (1985) concentric circles model in which countries are divided into the following three categories on the basis of the

---

7 The informants in Cameron (2002) were adolescent second language learners of English.
type of spread, the patterns of acquisition and the functional allocation of English: the inner circle which represents countries in which English is used as the L1, the outer circle consists of countries which use English as an institutionalized non-native variety and the expanding circle which is made up of countries in which there is a restricted use of English. The importance of English in the wake of globalization has increased the use of English in many countries that according to Kachru’s (1985) could be categorized as expanding circle countries. Moreover, certain groups of learners, although residing in an expanding circle country, might be categorized as second language speakers rather than foreign language speakers due to extensive exposure to English.

In view of Kachru’s model I will focus on the extent of exposure to English and the context of English learning. The following three categories which provide a further sub-categorization of Kachru’s (1985) expanding circle countries will be used: Low exposure learning macro-contexts are expanding circle countries where English is introduced fairly late in formal education and where there is relatively low amount of exposure outside the school context. They are characterized by learners almost exclusively coming into contact with English through formal learning. Middle exposure learning macro-contexts are countries where English is introduced fairly early in school and where learners have a relatively high receptive contact with English outside school from the media and popular culture. Another significant feature of middle-exposure countries is that foreign language programmes and films are dubbed into the national language.

Countries with a small national language and with a relatively higher exposure to English can be characterized as High-middle exposure learning macro-contexts, such as the Scandinavian countries and the Netherlands. It should be pointed out that this categorization of different learning contexts focuses on non-disadvantaged groups.

In this regard, countries such as Indonesia and Japan can be categorized as Low exposure learning macro-contexts, Germany along with other Western European countries, particularly such countries with a large national language, can be categorized as Middle exposure learning macro-contexts. High-middle exposure macro-learning contexts can be found in countries, such as Denmark, Sweden and the Netherlands.

As can be seen in Table 3, the estimated vocabulary sizes of the informants seem to vary according to the amount of English language exposure in the learners’ respective countries. The vocabulary size data from Zareva et al (2005) are not included in Table 3, because the estimates seem to be unreliable.
Table 3. Estimated vocabulary size of EFL learners.

<table>
<thead>
<tr>
<th>Country</th>
<th>Learning context</th>
<th>Estimated total vocabulary size/word families</th>
<th>Mastery of the 5K level</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesian first-term university students of various subjects</td>
<td>Low exposure</td>
<td>1226</td>
<td>NO</td>
<td>Nurweni and Read 1999</td>
</tr>
<tr>
<td>Japanese first-year students of various subjects</td>
<td>Low exposure</td>
<td>2304*</td>
<td>NO</td>
<td>Barrow et al 1999</td>
</tr>
<tr>
<td>German first-term students of English</td>
<td>Middle exposure</td>
<td>-</td>
<td>NO</td>
<td>Tschirner 2004</td>
</tr>
<tr>
<td>Spanish final-year students of English</td>
<td>Middle exposure</td>
<td>5500</td>
<td>NO</td>
<td>Pérez Basanta 2005</td>
</tr>
<tr>
<td>Danish first-year students of English</td>
<td>High-middle exposure</td>
<td>-</td>
<td>NO</td>
<td>Albrechtsen et al 2008; Staehr Jensen (2005)</td>
</tr>
<tr>
<td>Swedish first-year university students of English</td>
<td>High-middle exposure</td>
<td>8295**</td>
<td>YES</td>
<td>Gyllstad 2007</td>
</tr>
<tr>
<td>Swedish second-term university students of English</td>
<td>High-middle exposure</td>
<td>8570**</td>
<td>YES</td>
<td>Gyllstad 2007</td>
</tr>
<tr>
<td>Swedish third-year university students of English</td>
<td>High-middle exposure</td>
<td>9040**</td>
<td>YES</td>
<td>Gyllstad 2007</td>
</tr>
</tbody>
</table>

* The unit of counting employed in Barrow et al 1999 is the word type.
** The estimate of total vocabulary size is based on the reported mean scores on the 3K, 5K and 10K levels.

Summing up, as shown in Table 3, in Nurweni and Read (1999) and Barrow, Nakanishi and Ishino (1999) Asian adult EFL learners were all found to have relatively small vocabulary sizes below 3000 word families. Also Tschirner (2004) found that a majority of her informants did not master the vocabulary of the 5K level. Her conclusion is that the majority of the informants do not seem to reach the vocabulary size goals required for university studies at a German university.

There are unclear data regarding whether university students of English should be expected to master the 5K level at the onset of their studies in a High-middle exposure learning context. There is a lack of studies that provide a detailed picture of the vocabulary size of Swedish university students of English. According to David (2008: 168) data of this kind “are essential for our understanding of lexical development of L2 learners as they are the basis for further work looking to inform teaching practices through researching incidental or explicit learning”.

These findings suggest that since many learners fail to reach even moderate learning goals, we cannot take for granted that an adequate
vocabulary size will automatically be reached (Schmitt forthcoming). It is thus of importance not to assume that learners at a certain educational level have reached an adequate vocabulary size level. One important rationale for assessing learners’ vocabulary size in the present study is that this information forms a crucial backdrop for the understanding of their overall vocabulary ability.

Based on this review the following research question will be addressed:

9. What frequency levels do advanced Swedish learners master receptively at different points in their university studies?

There is no clear data regarding whether we should expect that a majority of the students of English examined here should be expected to master the 5K level already at the onset of their studies. However, on the basis of Gyllstad’s (2007) findings we should probably anticipate that the average total onset size of the students should not be below 7000 word families. Moreover, his findings showed that even at the third term of study, the students did not on average master the 10K level. A similar result is expected for the sample investigated here in regard to their lexical profile at the end of term two.

3.3.1.2 Learners’ size of productive vocabulary knowledge

The previous studies have mainly focused on the receptive end of the receptive/productive continuum of size of vocabulary knowledge. Few studies have investigated learners’ size of productive vocabulary knowledge. One study that did look at this is Zimmerman (2004). In this study of the role of vocabulary size in assessing second language proficiency, the productive vocabulary of 173 EFL exchange students at a preparatory English course for university studies in the US was examined. The students were placed in five proficiency levels according to their test results on the placement test, which assessed their reading, writing, speaking, listening and grammar skills. Level-one students are at a beginning level, whereas level-five students are at a high intermediate level of English proficiency. In Table 4 we can see the average productive vocabulary sizes found among the informants at the five levels:

As is evident from Table 4 the productive vocabulary sizes are very low for the two first levels and increase quite sharply with increasing language proficiency. However, Zimmerman (2004) found substantial gaps among level 5 students in their basic vocabulary. On average they knew only about 1300 words out of the 2000 most frequent words. The remainder of their productive size of vocabulary knowledge consisted of 2233 word families from the higher frequency levels.
Looking at all the students’ productive vocabulary, more than half of their productive vocabulary size consists of words less frequent than the first 2000 words. This seems to suggest that one should expect gaps in learners’ productive vocabulary as well as in their receptive vocabulary as noted above. Moreover, Zimmerman (2004) found that size of productive knowledge was a valid discriminator between learners at varying institutional placement levels.

Another study examining productive size is Tschirner (2004) who reports that among German first-term students of English only 22% master the 2K level.

With the exception of Zimmerman (2004) and Tschirner (2004) the vocabulary knowledge of adult learners in an academic context has mainly been assessed at the receptive end due to its perceived role in reading comprehension. In view of the important role of size of productive vocabulary knowledge as a placement tool in an academic setting indicated by Zimmerman (2004) it is of relevance to also investigate learners’ vocabulary size at the productive end among advanced Swedish learners. To this end the following research question is raised:

10. What frequency levels do advanced Swedish learners master productively at different points in their university studies?

In light of the higher exposure to English in the Swedish context than in the German context, we should expect to find that a majority of the students master the 2K level productively already at the onset of their studies.

Having looked at studies that have assessed learners’ size of vocabulary with the aim of determining whether they reach important vocabulary size goals for being able to manage university studies through the medium of English, the next section reviews studies that attempt to determine how much vocabulary is needed to carry out language tasks that are important for academic achievement.
3.3.1.3 Studies on the vocabulary size needed for different tasks

Whereas the point of departure of the above mentioned studies was mainly learners’ size of receptive knowledge, and they were carried out in order to draw conclusions about learners’ ability to reach certain vocabulary goals, the following review consists of studies that approach size of vocabulary knowledge from the aspect of how large a vocabulary a learner needs to adequately carry out a specific task. In order to establish relevant vocabulary learning goals, a great number of studies have attempted to determine how many words learners need to know in order to be able to comprehend different text types. This can be examined either as the percentage of lexical coverage needed for comprehension of spoken and written language or the amount of vocabulary a learner must know in order to comprehend spoken and written texts.

A great deal of research has been conducted to determine what kind of lexical coverage a learner needs to obtain sufficient comprehension of a text. In the research literature regarding coverage two amounts are often mentioned, 95% and 98% text coverage. According to Webb and Rodgers (2009) the two estimates can be said to represent two different degrees of comprehension, reasonable and ideal. In the research literature there are mainly two overall types of texts that have been investigated. One is non-academic texts representing everyday language use, such as reading a novel or watching a TV-programme. The other is academic texts representing the type of texts a learner might encounter when studying English or through the medium of English at the university. There are two main methods for determining an adequate threshold for reading comprehension. One is through corpus analysis. Different estimates within this method arise out of the coverage percentage, the type of text and range of texts. The second method is through relating size tests to comprehension tests. Here the threshold is determined by how one sets the pass/fail cut-off score on the reading test.

In regard to understanding everyday spoken English, evidence suggests that a vocabulary size of between 4000 word types is required for a 95% coverage (Adolphs and Schmitt 2004) and between 6000-7000 word families for a 98% coverage (Nation 2006). In order for learners to gain 95% coverage of English television programs knowledge of the 3000 most frequent word families is needed and knowledge of the 7000 most frequent word families provides 98% coverage (Webb and Rodgers 2009). Although a few researchers (e.g. Schmitt forthcoming; Nation 2006) imply that these figures somehow reflect productive mastery, it should be pointed out that there is not enough evidence showing how these figures relate to productive use. The vocabulary size estimates for being able to comprehend a wide range of written texts are slightly higher. Based on 98% lexical coverage Nation (2006) suggests that a vocabulary size of between 8000-9000 word
families is needed to comprehend a variety of authentic texts (e.g. novels and newspapers).

In addition to the pedagogical implications of research on lexical coverage and reading and listening comprehension these figures are to a great extent used as a frame of reference for interpreting size figures generated by vocabulary size tests. Although these types of studies provide valuable insights mainly for pedagogy, the question is to what extent these thresholds reflect the minimum required vocabulary size needed for comprehending university reading materials in an L2. One study that aimed to answer this question is Hazenberg and Hulstijn (1996). Before, thresholds ranging from 3000-5000 basewords were used as yardsticks even for advanced learners learning English at the tertiary level. The text material at the university is of a mainly academic character and contains a high proportion of low-frequency vocabulary. The question is whether a threshold in the range of 3000-5000 basewords is a realistic goal for second language learners at the tertiary-level to be able to manage university studies.

In an attempt to answer this question Hazenberg and Hulstijn (1996) administered a vocabulary size and a reading comprehension test to a large sample of first-year and prospective students at a Dutch university. The informants consisted of three groups of students: native speakers entering university as freshmen (N = 28), non-native graduate students (N = 41) and non-native prospective students taking a Dutch language entry examination battery test (N = 137). By looking at the mean scores on the vocabulary of the test takers who passed the reading test, they suggest that a non-native speaker of Dutch needs to know about 10000 “words” in order to be able to manage university studies.

As noted above the sample consisted also of non-native prospective students and they found that the difference between these and the two other groups was apparent in all four frequency levels examined, but the disadvantage of the non-native graduate students in comparison to the native speakers does not become evident until the 5K frequency level. This leads the authors to state that these findings highlight the importance of the 5,000 most frequent words as a learning goal for non-native university students, irrespective of field of study (Hazenberg and Hulstijn 1996). Relating this to the findings in Cameron (2002) the 5K level seems to constitute a key level which differentiates between native and non-native speakers.

Accordingly, whereas the 10000 level might be regarded as a lower limit for non-native speakers to be able to manage university studies, the 5,000 level can be regarded as a lower limit for students preparing for university studies (Hazenberg and Hulstijn 1996). According to Hazenberg and Hulstijn (1996) “the relationship between word frequency and word knowledge is not a straightforward one, it depends on proficiency level and frequency range” (158). Accordingly it would have been helpful to have had access to data showing whether mastering any of the four frequency levels produced a
significant correspondance with passing the reading test. This might have lowered the threshold for acceptable text comprehension. It might be the case that learners who have reached a certain minimum vocabulary size threshold can transfer L1 higher-order skills and processes which enable reading comprehension without having reached the 10K level. They actually say that vocabulary selection for L2 instruction is, and should remain ultimately, a subjective affair. They do however emphasize that the 5K level should be obligatory for inclusion.

In addition to reading comprehension, listening comprehension is perceived as a crucial skill in order to be able successfully to manage university studies. Staehr Jensen (2005) found that the 5K level constitutes a minimum vocabulary size threshold for adequate listening comprehension in an academic setting. The informants consisted of 100 Danish EFL learners who were all first-year university students of English. It was shown that vocabulary size correlated highly with listening comprehension and that a vocabulary size of 5000 basewords constitutes a minimum threshold for adequate listening comprehension for the majority of the informants. One should keep in mind that in reality the threshold might be less since in face to face communication different cues can be relied on to help the listener to make sense of the message. Moreover, in many lectures there is room for asking for clarification. As noted above, interestingly enough, the 5K level seems to be a crucial level also in studies of differences between native speakers and non-native speakers of English and what constitutes a minimal receptive vocabulary goal for pre-university students studying in an L2, such as Hazenberg and Hulstijn (1996) and Cameron (2002). Knowledge of other levels has also been found to correlate with language proficiency. For instance, Cameron (2002) found modest correlations between scores on the 2K and AWL and English language proficiency assessment scores. Whereas the 5K and 10K levels seem to be crucial levels in an academic setting, it is also of importance to know how large a receptive vocabulary is required for typical everyday use, since language development is dependent on a varied input.

In Table 5 different vocabulary size thresholds found for different language tasks are displayed.

**Table 5. Vocabulary size thresholds for different language tasks.**

<table>
<thead>
<tr>
<th>Language task</th>
<th>Threshold in number of basewords</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coping with university reading</td>
<td>5000-10000</td>
<td>Hazenberg and Hulstijn 1996</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>8000-9000</td>
<td>Nation 2006</td>
</tr>
<tr>
<td>Basic reading comprehension</td>
<td>3000-5000</td>
<td>Laufer 1995</td>
</tr>
</tbody>
</table>
As can be seen in Table 5 the threshold for adequate comprehension seems to go from basic comprehension via listening comprehension of academic texts to reading academic texts in terms of the required size of vocabulary knowledge.

Although there is a great deal of evidence for a strong relationship between vocabulary knowledge and language skills that are important to master in order to be able to manage university studies, it is important to keep in mind that academic success is presumably a function of many different abilities both linguistic and extra-linguistic in addition to reading and listening comprehension. It is thus of importance to investigate the relationship between vocabulary size and academic achievement in a more direct manner. This kind of information is useful for determining to what extent students’ performance standards can be monitored and diagnosed by continuous size assessments.

Studies that have examined the relationship between size and grades, have mainly looked at the relationship among secondary-school learners. Milton (2006) found that among secondary-school learners of French in the UK, learners with large vocabularies are likely to obtain a high ‘A’ level grade, while learners with low vocabularies are more likely to obtain a low grade. In a similar study by Richards, Malvern and Graham (2008) a modest significant correlation (.43) between receptive vocabulary size and GCSE grades was found among year-12 learners of French in UK schools. A study that specifically looked at the relationship among university students is that of David (2008). However, unlike the other two studies, she did not find a significant relationship between the receptive size and the academic achievement as reflected by ‘A’ level first-year university students of French at a UK university. She suggests that the lack of relationship found might be due to the exams primarily measuring productive knowledge. She calls for further studies to investigate the relationship between vocabulary size and academic performance.

Studies that have investigated the link between learners’ vocabulary size and grades have found a relationship among secondary-school learners, but no relationship among university students (David 2008; Milton 2006; Richards, Malvern and Graham 2008). David (2008) suggests that among more advanced learners a relationship might be found between productive size and academic achievement, because measures of academic achievement for advanced learners probably include relatively more emphasis on productive knowledge than mere receptive knowledge of words.

As suggested by the review the relationship of vocabulary size to academic achievement has mainly been investigated indirectly by relating vocabulary size to different language tasks some of which are crucial for managing university studies. In view of this the following research question will be addressed:
11. How is size of vocabulary knowledge related to academic achievement?
12. How does the relationship between vocabulary knowledge and academic achievement change over time?

Although there are some contradictory findings regarding the relationship between vocabulary size and school and academic performance, the vast circumstantial evidence of a relationship between vocabulary size and overall grade suggests we should expect to find a positive relationship between size and academic achievement. If, as suggested by David (2008) grades at higher levels of education incorporate productive knowledge of vocabulary knowledge to a higher extent, we should also anticipate a stronger link of productive size and productive depth with course grade.

Having looked at studies that dealt with size of vocabulary knowledge, we will now turn to studies that have dealt with the relationship between depth of vocabulary knowledge and different language skills.

3.3.2 Depth of vocabulary knowledge and academic achievement

It is acknowledged that vocabulary knowledge is multi-facetted and that in addition to a large vocabulary size, learners also need to have qualitative knowledge of lexical items to be able to function in a language. Consequently, a number of researchers have attempted to examine the relationship between vocabulary knowledge and different language skills, beyond just vocabulary size. As with size of vocabulary knowledge it seems that mostly receptive skills have been investigated in regard to their relationship with depth, specifically reading comprehension. In order to fully understand the importance of vocabulary knowledge in learners’ L2 learning process depth must also be examined from the point of view of its relation to ‘productive’ vocabulary use.

In view of the strong links found between vocabulary size and depth on the one hand and between size and reading comprehension on the other, a number of studies have attempted to explore the extent to which depth in comparison to size is related to reading comprehension. Among adolescent and adult ESL learners a high correlation was found between vocabulary size, depth and reading comprehension test scores (Qian 1999; Qian 2002; Qian and Schedl 2004). In regard to the relative importance of size and depth in predicting the performance on reading tasks, both are found to be equally good predictors of reading performance since scores on the depth and size of vocabulary knowledge measures are both capable of explaining a considerable portion (over 50%) of the variance in reading comprehension scores (Qian 2002). Moreover depth explained 11% (1999) and 13% (2002) percent of the variance in the reading comprehension scores beyond the variance already explained by vocabulary size. Accordingly, depth of vocabulary knowledge
knowledge seems to be a significant factor in reading performance, given that one has a minimum vocabulary size of 3000 word families, which was the minimum requirement for inclusion in the study (Qian 1999).

Another skill that has been examined from the point of view of its relationship with depth of vocabulary knowledge is listening comprehension. Staehr-Jensen found a high correlation between listening comprehension and depth conceptualized according to the components approach and a modest correlation with depth conceptualized as network knowledge. Examining the relationship of depth of vocabulary knowledge and grammatical knowledge with reading and listening comprehension among foreign language learners of Spanish, Mecartty (2000) found that only depth of vocabulary knowledge explained the variance in both listening and reading comprehension.

The question is whether such a strong relationship may also be found between depth of vocabulary knowledge and writing ability. This question is especially relevant in light of the complex nature of writing ability, which draws on a wide range of skills. Cumming (1989) states in reference to writing ability “that writing expertise [is] a central cognitive ability—with second language proficiency aiding it, facilitating it in a new domain, and possibly adding to it”. The only study that has examined the relationship between depth and writing ability appears to be Albrechtsen, Haastrop and Henriksen (2008); (cf. section 3.1.1). Looking at the relationship between depth of vocabulary knowledge conceptualized as network knowledge and writing ability, they found no significant correlation between scores on the productive depth test and scores for writing product quality (see below) among the advanced university learners. A combination of two scoring systems was used: the TOEFL scoring guide and the Michigan Writing Assessment Scoring Guide. The assessment of the learners’ essays included three features: ideas and arguments, rhetorical features and language control. As mentioned in section 3.3.1 they attribute the lack of significant correlations to the possible lack of sensitivity of the two depth tests. At any rate, these results ran counter to their prediction that writing ability would be closely related to depth of vocabulary knowledge since “the quantity and quality of the lexical links tapped by the network measure would function as an important declarative knowledge base for complex processing in writing by providing more access routes to lexical items, and thus freeing mental capacity for important higher-order processing” (Albrechtsen et al 2008: 165). In order for a significant correlation to emerge they suggest a depth measure based on a wider range of word classes (they only sampled adjectives) and less frequent vocabulary (ibid.). Another possible explanation for the lack of significant correlations might be that since the score of writing product quality represents many different sub-skills it might be difficult to find a significant correlation with any one individual score of these sub-skills. 
As indicated by this review there is are few studies that look directly at the relationship between depth of vocabulary knowledge and academic achievement in L2 based subjects. Accordingly on the basis of the current state of knowledge of how depth of vocabulary knowledge is related to academic achievement the following two research questions will be addressed here:

13. How is depth related to academic achievement English?
14. How does the relationship between depth and academic achievement change over time?

In view of the circumstantial evidence of a strong relationship of depth to mainly receptive skills, we should anticipate a positive relationship of depth to academic achievement.

In the next section studies examining the relationship between lexical richness and measures of overall written text quality will be outlined.

3.3.3 The relationship of lexical richness to text quality and academic achievement

Section 3.1.1 dealt with lexical richness in learners’ written output from the perspective of its relationship to size of vocabulary knowledge as measured by discrete-item tests. The question arises as to what extent lexical richness is related to the overall quality of written texts. This has mainly been investigated from two perspectives in the research literature. The first one is in the form of the importance teachers and learners perceive lexical richness to have for the quality of written texts.

From the point of view of students’ and teachers’ experiences, studies such as Santos (1988) and Leki and Carson (1994) suggest that lexical richness is closely related to the overall quality of essays. From the point of view of teachers, comprehensibility is judged as most important and lexical errors are considered most serious, since they impinge on comprehension (Santos 1988). The importance of lexis is, accordingly, stressed by those responsible for the teaching and evaluation of students, due to the close relationship between content and lexis. It does not seem possible to grade content independently from lexical errors, since “it is precisely with this type of error that language impinges directly on content; when the wrong word is used, the meaning is very likely to be obscured” (Santos 1988:84).

Also from the point of students’ perspective this connection is emphasized (Leki and Carson 1994). Lack of control of vocabulary is perceived as something holding the students back. They required more help in increasing their vocabulary, and more teaching of sophisticated vocabulary (ibid.). The following concerns were among the ones raised by
students regarding the lack of appropriate vocabulary knowledge: that it takes a long time to access appropriate lexical forms, and that there is imprecision in their writing due to them being forced to express themselves with words that are only semi-fitting with their thoughts (Leki and Carson 1994). The importance of vocabulary is thus recognized both by learners and teachers.

The second approach deals directly with learners’ written output from the point of view of the extent to which objective lexical measures are related to overall text quality. Two early studies using word-list-free approaches to measure lexical richness that have found a positive relationship between lexical richness and overall text quality are Linnarud (1986) and Engber (1995). The material analyzed in Linnarud (1986) consisted of 54 compositions written by 42 Swedish learners of English at the upper secondary level and by 12 NS of English of the same age. The only lexical measure for which she found a significant moderate correlation (0.47) with composition grades was lexical individuality\(^8\). Engber (1995) found a significant moderate correlation between lexical variation (see section 2.6.1 for a description of different measures of lexical richness) and scores of overall text quality of 0.57. She also had a count of lexical variation that included lexical errors but obtained a lower correlation score than the one without lexical errors.

In both Linnarud (1986) and Engber (1995) the essays were holistically scored. In view of the fact that holistic scoring represents a compromise of many different considerations, such as content, organization and language control, a modest correlation between lexical richness scores and essay quality scores obtained in the two studies suggests a strong relationship between lexical richness and overall text quality.

However, the problem is that different studies obtain different correlations for different lexical richness measures. From the point of view of comparing results from different studies a universal measure of lexical richness would be ideal. The LFP has gained currency in later studies and is a measure that today comes closest to a standard analysis of lexical sophistication (cf. section 2.6.1.2.2).

Studies employing the LFP to measure lexical sophistication have found a weak to modest relationship between lexical sophistication of timed compositions and overall text quality (i.e. Coniam 1999; Goodfellow et al 2002; Li 1997). Moreover, East 2009, who used the LFP in a study evaluating the reliability of a new holistic scoring rubric for assessing foreign language writing, found significant differences in lexical sophistication between timed compositions according to their text quality among adolescent intermediate learners of German. However, none of these

\(^{8}\) Lexical individuality reflects the “[r]atio of lexical words that are unique to the writer to the total number of words in the composition” (Engber 1995: 142)
studies examined this relationship in an authentic learning context in which the writing task and the text quality scoring is part of an academic examination.

One study that has looked at the relationship between lexical richness and academic achievement in the context of upper-secondary education, found a modest relationship between the lexical diversity of timed compositions and GSCE grades among English year-12 learners of French (Richards, Malvern and Graham 2008).

Another study which attempts to shed light on the relationship between lexical sophistication in student writing and academic success is Morris and Cobb (2004). They examined the LFP as a predictor of academic performance among 122 bilingual TESL trainees and found that a significant low correlation (.37) holds between the proportion of academic words in the informants’ texts and course grades (CG). According to Morris and Cobb (2004) this correlation is too low to warrant the use of LFP as the only assessment instrument of potential TESL candidates.

However, they stress the weight of the correlation on the grounds of the pattern of statistical significance in the word categories, i.e. CG correlated positively with the proportion of academic words, and negatively with the proportion of function words and high frequency words. They also found that the LFP was able to discriminate between texts from highly advanced nonnative speakers and native speakers. On the basis of analysis of which LFP scores systematically resulted in academic success, they present the following levels: K1 score of less than 85% and an AWL score of over 5%. Relating this standard to the LFP scores of the informants in Laufer and Nation (1995), among the three proficiency groups only the low proficiency group does not reach this standard. This lends some support to the validity of the standard.

As shown in previous studies there seems to be at least a weak relationship between lexical sophistication as measured by the LFP and overall text quality. However, we know relatively little about whether lexical richness of take-home essays is related to academic success both in terms of essay grade and the overall course grade. There is a fundamental difference, on the one hand, between the writing tasks of timed compositions and take-home essays and, on the other hand, between text quality assessed in an inauthentic low-stakes study context and text quality assessed in a high-stakes academic examination context. Accordingly, it is of interest to investigate how lexical sophistication of learners’ written output relates to authentic measures of text quality and academic success. It is also important to shed further light on the feasibility of employing the LFP as a measure of text quality in the academic context.

All the studies mentioned above that have examined learner texts from the point of view of the relationship between lexical richness and text quality have only analyzed timed compositions. The relationship that holds between
lexical richness of timed compositions and text quality may be different from the relationship between lexical richness of take-home essays and text quality, since take-home essays are written under different time and cognitive constraints. There are also the differences in terms of access to different writing aids. With reference to the differences between the writing tasks of timed compositions and take-home essays Muncie (2002: 232) states:

if the timed composition is a measurement of their normal vocabulary range, the composition using the process approach shows not just their everyday range of vocabulary, but also the extra work and extra resources that the students have been able to employ during its production.

For instance, Kenworthy (2006) found that there were significant differences between the timed compositions and at-home essays of Chinese applicants to an American university leading him to conclude that the at-home essay with its benefits of additional time and access to aids positively affect overall textual quality.

Muncie (2002), on the other hand, looked specifically at lexical richness as measured by the LFP when comparing timed compositions with a first and final draft of a take-home essay on the same subject. These were written by 30 Japanese EFL learners enrolled on an English composition course. She found that LFP scores did not improve significantly from the first draft to the last and that the last draft did not contain more beyond-2000 vocabulary than the timed composition. However, when excluding final draft essays that bear no resemblance with the first draft, final drafts have an average beyond-2000 score of 11.74% as opposed to 8.02% in the timed compositions. As is evident from Kenworthy (2006) and Muncie (2002) take-home essays differ from timed composition in terms of grammatical and lexical richness being written under different circumstances. In addition, East (2006) found that timed compositions written by adult learners of German who had access to support resources had greater lexical richness than timed composition written without access to such aids. Moreover, no significant differences were found in lexical richness between lower ability informants who had access to dictionaries and higher proficiency informants due a proportionately higher increase in the use of low frequency vocabulary among the former group.

In regard to reading comprehension Mcmillion and Shaw (2009) found that L2 learners’ reading where time is not an issue can compensate for weakness in low-level processing (e.g. word recognition) by utilizing higher-level processing. Accordingly, these findings suggest that learners with a relatively lower vocabulary knowledge can produce equally lexically rich texts as stronger learners under certain conditions by utilizing compensatory strategies.

67
In summary, there is a lack of studies that have investigated the relationship between the lexical richness of take-home essays and text quality. As indicated by Muncie (2002) and Kenworthy (2006) we cannot immediately assume that the relationship between lexical richness of timed compositions and text quality also holds for lexical richness of take-home essays and text quality. In order to both evaluate the feasibility of using the LFP as a pedagogical tool for assessing essays in the academic context and also investigate whether it is possible to detect a relationship between lexical richness of take-home essays as measured by the LFP and text quality the following research questions will be addressed:

15. To what extent is lexical richness of take-home essays related to essay grade and course grade?
16. How does the relationship of lexical richness to essay grade and course grade change over time?

On the basis of previous findings to the effect of a weak to modest relationship of lexical richness of learners’ timed compositions and overall text quality, a similar relationship is anticipated here. In addition evidence indicates that lexical richness of learners’ timed compositions is related to school and academic grades. A weak to modest relationship between the lexical richness of students’ take-home essays and course grade can be expected. Despite clear data about how the relationship develops in other contexts, a stronger relationship of lexical richness to essay and course grade over time is expected.

In the previous three sections we looked at studies investigating the relationship between vocabulary knowledge and different language skills and lexical richness and overall text quality.

The next section will outline studies that have dealt with the development of learners’ vocabulary knowledge and use over time.

3.4 The development of learners’ vocabulary knowledge and use

An insight into how vocabulary knowledge develops over time in a learning context without any particular focus on vocabulary is a central part in a theory of second language vocabulary acquisition, because it can help us predict vocabulary learning outcomes for different language learning contexts and different groups of language learners and can facilitate the formulation of appropriate vocabulary learning goals. This section will review studies that have dealt with the development of learners’ vocabulary
knowledge and use in a learning context without specific emphasis on vocabulary.

3.4.1 The development of vocabulary size over time

The development of size of vocabulary knowledge is affected by many different factors. Two factors that are easily identified are age and the learning context. In cases where an additional language is acquired from a young age one would expect a plateau in the increase of vocabulary size when the learner reaches adulthood, because the bulk of the vocabulary needed to function in the L2 would typically already have been acquired. Since the present study is concerned with advanced learners, the following review will be made up of research on secondary and tertiary level learners, particularly learners who are within the proficiency range of intermediate to advanced.

The second crucial factor is the extent to which particular focus is placed on learning vocabulary in the particular learning context. Insight into how much vocabulary a learner will typically learn by mere exposure is important for further work aiming to inform teaching practices.

In order to understand how advanced learners’ size of vocabulary knowledge develops at the university level, it is necessary to first consider findings pertaining to the previous school level namely that of the secondary level.

To pursue the idea that the process of learning a second language can be described as progress along the Interlanguage continuum, Laufer (1998) examined the vocabulary development of secondary-level EFL learners.

The pedagogical implication of studying vocabulary growth of EFL learners is that we get an idea of how much instruction is needed to reach certain vocabulary size goals. The theoretical implications are that by investigating development of vocabulary size among second language learners we can shed light on the relationship between different learning inputs such as level of education and learning context, second or foreign, and the differences between the development of productive and receptive vocabulary. Laufer (1998) argues that vocabulary development, both productive and receptive, does not seem to progress in a linear fashion. She found that despite the learning context being foreign there seems to be a burst of increase in later education levels. She examined the development of three types of vocabulary knowledge among 26 EFL 10th graders and 22 11th graders. The study was cross-sectional, and receptive and productive knowledge was tested using the VLT. She found an increase of 1600 word families in the receptive vocabulary, which corresponds to an increase of 84% over one year. The productive vocabulary increased by 850 word families which translates into a 50% increase over the same period.
The results make sense in the light of the higher exposure to English due to increasing exposure to popular culture as the learner becomes adolescent and adult. Moreover, the different expectations and tasks in secondary-level and tertiary-level studies might very well encourage growth. However, as pointed out by Waring (1999), one can question whether these results reflect a real increase since nothing is known about these groups other than that they differed by a single school year.

Laufer (1998) states that since the raw scores of the 10th graders correspond to a vocabulary size of 1900 word families, it would mean that it took them 6 years to learn 1,900 word families and only one school year to learn 850 word families. According to her this might suggest that either vocabulary learning is non-linear in nature or that teachers in earlier years underestimate the students’ capacity to learn vocabulary and do not require enough. If there is indeed a burst of vocabulary growth among high school learners it might not necessarily be enough in terms of having a sufficient vocabulary size to be able to manage university studies. Although English is quite widely used in Israel and they start relatively early, the fourth grade or in some cases earlier, the typological differences between Hebrew and English create difficulties for the students. This might be a factor behind the low onset vocabulary size in that there is a delayed L2 language development. As pointed out by Laufer (1998) the increase in expectations on the part of the teachers might be due to matriculation examinations.

With respect to the growth profile among the informants, the greatest growth for receptive size occurred at the 3K, 5K and the UWL level. For productive size the increase was evenly spread along all the four levels assessed. Laufer (1998) concludes that a large number of words can be learned in a normal classroom setting without extensive exposure to the target language outside the classroom.

If one can expect a large increase in receptive and productive size among EFL learners in a secondary school context without particular focus on vocabulary, the question arises whether a similar development might be found among university students in an unfocused vocabulary learning context. Evidence suggests that Israeli university students learn 300 word families over only one term of study (Laufer 1995). The slightly lower increase than that found among upper-secondary students is perhaps because they were less motivated than the upper-secondary students who were preparing for matriculation exams (Laufer 1998).

In view of Schmitt and Meara’s 1997:24 statement, “the larger vocabulary, the less room for growth”, the question is rather whether one might expect similar growth among more advanced learners with a larger onset vocabulary.

For instance, Lessard-Clouton (2006) examined the size of specialized vocabulary knowledge of 5 non-native speakers and 7 native speakers of English enrolled on a graduate school of theology. The aim was twofold, to
investigate how well the informants knew a selection of specialized theological vocabulary and how their knowledge of these developed over one term. He found that the non-native speakers increased their vocabulary size by 10%, whereas the native speakers increased their vocabulary by 3%, closing the gap between them. This seems to indicate that lower proficiency students make relatively more progress in order to reach departmental requirements. However, the relatively lower increase among the NSs might be due to a ceiling effect in the measure employed. Although this study supports the idea that growth can take place among advanced learners with a relatively high onset vocabulary, it might be difficult to relate the learning of technical terms to the learning of general vocabulary.

If we look at other studies that have found a relatively high increase of size of vocabulary they all have in common with Laufer (1998) that the onset vocabulary is quite far below the 5K mark which has been found to be an important learning goal and a level that discriminates between native and non-native speaking peers (e.g. Hazenberg and Hulstijn 1996; Cameron 2002). For instance, in a longitudinal study of 88 Japanese learners of whom a majority were freshmen and sophomore students of English with an onset average vocabulary size of 3900 word families, Schmitt and Meara (1997) found an average increase of 330 word families over one term. Although this does not seem to be an impressive figure the increase is fairly high according to the authors, based on the minimal input of English outside of the classroom. One of the limitations with this study is that we are not provided with information about the growth of vocabulary size according to how many years the students had studied English. The sample consisted of informants from three different levels of education. In regard to how their vocabulary increases according to frequency levels, as in Laufer (1998), there was an average significant increase on all five levels. However, a slight difference in contrast to Laufer (1998) is that the greatest increase took place on the 2K level and 10K level, which makes them conclude that the subjects were filling in gaps in their knowledge of high-frequency vocabulary as well as gaining new lower-frequency words.

This indicates a slightly different learning profile than that found in Laufer (1998). Both Schmitt and Meara (1997) and Laufer (1998) suggest that word frequency is not a very reliable index to the probability of a word being known, since the informants do not seem to learn high-frequency words before learning words at lower-frequency levels (for a discussion of word frequency see section 2.4.12).

Another study indicating that a below-5K onset vocabulary is a prerequisite for large growth is Bennett (2006), in which 17 Japanese learners enrolled on an intensive university preparatory English language

---

9 Laufer (1995; 1998) and Schmitt and Meara (1997) used the receptive Vocabulary Levels Test.
programme in Japan were tested on among other things their size of English vocabulary knowledge using the VLT. Since Bennett (2006) only tested two levels namely, the 2K and 3K, in a pre-test and post-test fashion, he used a somewhat elaborate method of estimating the informants’ overall vocabulary increase over three months. By assuming a similar ratio as in Laufer (1998) a rough estimate of vocabulary size was obtained. It showed that the informants initial mean vocabulary size of 1718 grew to 2690 which is an increase of 972 word families or an overall increase of approx. 57%. Although this is probably an overestimate, the increase is in line with the findings of the above mentioned studies. If we disregard other possible factors, the higher increase in Laufer (1998) and Bennett (2006) in comparison to Laufer (1995) and Schmitt and Meara (1997) can be ascribed to the learners’ relatively lower onset vocabulary size i.e. below the 5K mark. It should be pointed out that the learning context in Bennett (2006) was an intensive English language programme which might explain the relatively high increase, but might not be representative of most university English programmes.

All the above mentioned studies have looked at learners of English, a study which investigated the vocabulary gains among English university students of French with an onset size of about 2000 lemmas is Milton (2008) who found an increase of about 600 lemmas over two terms of study. Although, Milton (2008) expresses some concern about what he sees as a low growth in these learners receptive size, his findings accord with other studies who have found increase in receptive size among learners with a relatively low onset level and who have not received any explicit vocabulary instruction. Milton (2008) did also find evidence of vocabulary growth being related to onset level. Little or no progress was found among students who entered university with relatively large vocabularies, while students who had a relatively low onset size made quite substantial progress, on average about 1000 lemmas over two terms of study.

There is some evidence of a relatively high increase of learners’ receptive vocabulary knowledge in a foreign language setting without specific focus on vocabulary (e.g. Laufer 1995; Schmitt and Meara 1997). This seems also to be the case for university-level learners in an L2-context. Although Milton and Meara (1995) found a high average increase in the vocabulary size of exchange students enrolled at a British university, the largest increase was found in low-level learners. In this paper 53 European exchange students were tested using a Y/N-vocabulary test administered twice within a six-month interval. Milton and Meara (1995) found an average increase of 1326 lemmas (26%) over a period of six months, which, according to them, would correspond to an annual average increase of 2600 lemmas. They also calculated the informants’ half–year vocabulary increase before visiting the UK, comparing it with the growth during the six-month stay in Britain and found an average increase of 275 basewords during a six-month period
before coming to Britain. This corresponds to only approx. 20% of the increase while in UK. Low-level students accounted for the greatest increase. Although the general size was about 5800 lemmas, they found that the highest increase was among learners with low onset vocabularies. This single factor explained about half of the total variation in the growth rates.

A vocabulary size of 5000 basewords is not particularly high when talking about university-level learners. Nevertheless, disregarding other possible factors, the studies reviewed suggest that this level seems to be a factor behind vocabulary growth, in that we can identify a relatively sharp increase of vocabulary size among learners who have not acquired an overall receptive size of more than 5000 word families.

There is some contradictory evidence of small or no growth among university learners of English with a low onset vocabulary size. For instance Cobb and Horst (1999) found no significant increase in receptive size among Hong Kong students of English over a six-month period. A similar finding is reported in Clark and Ishida (2005) who examined the receptive size of foreign exchange students at an English for Academic Purposes course at an Hawaiian University. However, there are some limitations of these studies which make it difficult to generalize from them. The main one is that only a few of the levels were examined, which means that increase might have taken place in the levels that were not included in the analysis. In the case of Cobb and Horst (1999) only three levels were used for assessing the informants’ vocabulary which are probably too few for obtaining a reliably estimate of their total vocabulary size and also for capturing growth. The results from the two studies seem to be unreliable in terms of vocabulary growth of university students of English in a setting with no particular focus on vocabulary, because the studies were conducted in a L2 context where one would have expected at least some increase from mere exposure. Nevertheless, the findings of these two studies are not applicable to the foreign learning context examined here.

It is important to point out that even though we can observe an increase in learners’ vocabulary knowledge from mere exposure, there are probably numerous variables which can compensate for a lack of focus on vocabulary. The absence of focus on vocabulary in a specific learning context does not imply that in other respects such contexts are equal in terms of the opportunities for vocabulary acquisition offered. Accordingly we need to be clear about what we mean when speaking about formal learning contexts without specific focus on vocabulary. In the current study a formal learning context without focus on vocabulary, such as an English language programme at the university refers to language courses that do not include an intentional vocabulary learning component such as teaching of vocabulary learning strategies, seminars or workshops devoted to offering opportunities.
for intentional learning of vocabulary or dedicated to teaching vocabulary. Table 6 displays a summary of the results in the reviewed studies\(^{10}\).

Table 6. *The increase of learners’ vocabulary size.*

<table>
<thead>
<tr>
<th>Type of learners</th>
<th>Onset receptive size</th>
<th>Onset productive size</th>
<th>Increase in receptive size</th>
<th>Increase in productive size</th>
<th>Time</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese pre-university students of English</td>
<td>1718 w.f.</td>
<td>-</td>
<td>972 w.f.</td>
<td>-</td>
<td>3 months</td>
<td>Bennett 2006</td>
</tr>
<tr>
<td>Israeli upper-secondary school students of English</td>
<td>1900 w.f.</td>
<td>1700 w.f.</td>
<td>1600 w.f.</td>
<td>850 w.f.</td>
<td>1 year</td>
<td>Laufer 1998</td>
</tr>
<tr>
<td>Israeli university students of English</td>
<td>3000 w.f.</td>
<td>-</td>
<td>300 w.f.</td>
<td>-</td>
<td>1 term</td>
<td>Laufer 1995</td>
</tr>
<tr>
<td>Exchange EAP students</td>
<td>4678 w.f.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1 term</td>
<td>Clark and Ishida 2005</td>
</tr>
<tr>
<td>Exchange first-year students of various subjects</td>
<td>5772</td>
<td>-</td>
<td>1325 lem.</td>
<td>-</td>
<td>6 months</td>
<td>Milton and Meara 1995</td>
</tr>
<tr>
<td>EFL students</td>
<td>-</td>
<td>-</td>
<td>275 lem.</td>
<td>-</td>
<td>6 months</td>
<td>Milton and Meara 1995</td>
</tr>
</tbody>
</table>

Lemmas = lem.; Word families = w.f.

Studies are ordered according to the estimated onset vocabulary size of the learners. As is clear from Table 6 two main patterns can be discerned. First of all very few studies have investigated development of productive size of vocabulary knowledge. The studies that have used a longitudinal design have only examined the development over one term. In some cases conclusions on annual development have been drawn from this shorter term data (e.g. Milton and Meara 1995). In others, due to the proficiency level of the informants not all levels were used, which obfuscates the picture of how the development takes place for low frequency vocabulary (e.g. Laufer 1998; Lessard-Clouston (2006) is not presented in the table, since his study involved technical vocabulary rather than general vocabulary, Cobb and Horst’s (1999) study is not included in the table since it is not applicable to foreign language learners.

\(^{10}\)
Bennett 2006; Clark and Ishida 2005). Three of the studies used a cross-sectional or a mixed design to investigate the development, which unfortunately does not provide a clear picture of how vocabulary develops for the same individual over time. A case in point is Laufer (1998) where use of cross-sectional data generates possibly invalid results.

In the present study a longitudinal design will be applied over two terms in which the informants’ receptive and productive vocabulary size will be assessed on all five frequency levels of the Levels Test. Two main points emerge from the review that would benefit from being further examined longitudinally. First, in an EFL context a large increase was found among learners with an onset overall vocabulary size which is below the 5K mark. The highest increase was found among adolescent learners followed by adult tertiary-level learners with a below-5K vocabulary. Even among the informants in Milton and Meara (1995), who on average had an onset vocabulary size of more than 5000 lemmas, there is a clear tendency for students with low onset vocabulary to make a great deal of progress, while those with larger vocabularies show smaller gains. It should be mentioned that the estimate would be considerably lower if expressed in terms of the number of word families known. In view of these findings it is of interest to further explore whether it is possible to detect a development of vocabulary size among learners in an EFL context who have a larger onset vocabulary than 5000 word families. Second, if we look at the development of the individual levels, we see that the greatest growth for receptive size occurred at the 3K, 5K and the academic word level. For productive size the increase was evenly spread along all four levels assessed (Laufer 1998). This will be further examined by investigating the development at the specific levels among the sample in question.

These two points will be addressed by the following research question:

17. How does advanced university EFL learners’ size of vocabulary knowledge change over time?

If the onset size is an important determiner of the expected growth of vocabulary in a learning context without particular focus on vocabulary, then we should probably not expect to find a significant average increase of the students’ receptive size over the period of time examined. However, we should anticipate significant growth among learners with a below-average onset receptive size. A potential growth in receptive size should be limited to the lower-frequency vocabulary levels, more specifically the 10K level. If the onset level is of relevance then we should find a relatively larger and more rapid increase in productive size. The increase is expected to be more spread out over the levels than for receptive size.

To date most research on the development of learners’ vocabulary knowledge has been concerned with size of vocabulary knowledge. The
following section will look at studies that have examined the development of learners’ depth of vocabulary knowledge.

3.4.2 Studies on the development of learners’ depth of vocabulary knowledge

In comparison with vocabulary size, one might expect that depth develops more slowly, since increasing one’s depth of vocabulary knowledge requires more effort and exposure than merely adding new vocabulary of which one has a superficial knowledge, especially in formal learning contexts where no specific emphasis is placed on learning vocabulary. As has been shown, estimates from vocabulary size measures can often be compared across studies. This is less possible with measures of depth, since in many cases they do not share a common reference point due to the different ways depth has been conceptualized and assessed. Nevertheless, knowledge of the development of depth at different stages in learners’ L2 acquisition is important in order to understand how qualitative knowledge of vocabulary grows in learning contexts where qualitative knowledge of words is acquired mainly through incidental learning. The development of depth of vocabulary knowledge has been approached from two main perspectives. The first concerns whether one can see an increase through mere exposure in a formal learning context, and the second relates to whether different components of depth form a developmental hierarchy.

One of the first attempts to operationalize the components approach both receptively and productively was made by Schmitt and Meara (1997) (cf. section 3.4.1). The aim was among other things to see how two components of depth, word associations and grammatical suffix knowledge, developed over one term of study. According to Schmitt and Meara (1997) the subjects’ onset depth of vocabulary knowledge of both components was rather poor. Their suffix knowledge increased productively by 5% and receptively by 4%, hence, an almost identical development in both knowledge types. As regards word associations, there was no significant increase in receptive associations and a 16% increase in the scores for productive associations, respectively. They found no clear pattern in the data that could explain the non-parallel growth of productive and receptive association knowledge.

Another study that conceptualized depth as components is Schmitt (1998), but unlike the subjects in Schmitt and Meara (1997) the informants in his study were advanced learners of English who had just begun their postgraduate studies at a British university. In an attempt to shed light on how individual words are acquired, he conducted a detailed case study of 3 adult learners tracking the incremental acquisition of 11 words over the course of one year. Four components were investigated productively,
knowledge of spelling, associations, grammatical information and meaning\textsuperscript{11}. There was a lack of parallel progression between the components. Two of the three informants progressed steadily in their knowledge of meaning senses and their associations became more and more native-like over time. As for spelling and knowledge of word class derivations no significant progress was made due to the students’ spelling and grammatical knowledge already being quite advanced. As noted in section 2.5.3, this study is a good example of the problems with operationalizing the components approach (since the more components assessed on one occasion the fewer words can in a practical way be tested).

Both Schmitt and Meara (1997) and Schmitt (1998) found that in appropriate circumstances depth of vocabulary knowledge develops over a relatively short time period, one term and one year, respectively, even among advanced learners. In both studies depth of vocabulary knowledge was conceptualized according to the components approach. This enabled them to track the development of individual components and they found that learners’ knowledge of components did not progress in a parallel fashion. As shown in Schmitt and Meara (1997) this lack of parallel development applied equally to the development of productive and receptive depth of vocabulary knowledge.

In the research literature depth of vocabulary knowledge has been conceptualized to a greater extent as precision of knowledge and network knowledge than as components. For instance, in his case study Lessard-Clouston (2006), in addition to examining learners’ development of size of vocabulary knowledge of technical theological vocabulary (see section 3.4.1), also tracked the development of depth of this vocabulary knowledge using an adapted version of the Vocabulary Knowledge Scale which is a depth test. He found a 6% increase of the score on the depth test among the NES over a period of one term of study. However, the results of this study are difficult to generalize to a learning context without focus on vocabulary, since technical vocabulary receives particular focus.

Two studies that have examined the development of depth conceptualized as network knowledge are Greidanus and Nienhuis (2001) and Greidanus, Bekx and Wakely (2005) who found a growth in depth of vocabulary knowledge test scores of 3.9% and 11%, respectively, among Dutch first year university students of French over a three to four year period of study. It is difficult to ascertain whether the findings of the two studies reflect a real growth within the same learner over time, since they employed a cross-sectional design.

As mentioned in the introduction of this section the development of depth of vocabulary knowledge has also been examined with respect to whether

\textsuperscript{11} Knowledge of meaning was the only component that was measured both receptively and productively
the different components form a developmental hierarchy. While Schmitt (1998) did not find any evidence of a developmental hierarchy of the four examined word knowledge types, Greidanus et al. (2005) found tentative evidence suggesting that the three components they examined form a developmental hierarchy with knowledge of paradigmatic relationships of a given word being acquired before knowledge of syntagmatic relationships which in turn is acquired before knowledge of analytic relationships (cf. section 2.5.1.2).

By way of summary the studies reviewed have found that learners’ depth of vocabulary knowledge develops both over a relatively short time period and over a longer time period in a university setting with no particular focus on vocabulary. Moreover, as shown in Schmitt and Meara (1997) and Schmitt (1998) different components seem to show different developmental patterns.

In view of a lack of studies employing a longitudinal approach on a larger sample than a couple of subjects that have tracked the development of productive depth of vocabulary knowledge, the following research question will be addressed in the present study:

18. How does learners’ productive depth of vocabulary knowledge develop over time?

On the basis of the review of studies that have investigated the development of learners depth of vocabulary, we should expect an overall development of depth in the context examined here, especially since depth was operationalized according to the context of university studies by sampling academic words. Since both Schmitt and Meara (1997) and Schmitt (1998), have found that components of depth develop in a non-parallel fashion, a similar pattern is expected among the Swedish sample examined here. Moreover, in light of Greidanus and Nienhuis’ (2001) and Greidanus et al.’s (2005) findings that different aspects of word knowledge are acquired at different rates, one should expect that the components of depth develop differently.

Having looked at studies that dealt with the development of the vocabulary knowledge dimension of vocabulary ability, we will now review research that has investigated the development of the lexical richness of learners’ written texts.

3.4.3 The development of lexical richness in learners’ written output

In sections 3.4.1 and 3.4.2 it was reported that studies investigating the development of size and depth among learners in a formal learning context without specific emphasis on vocabulary in general had found a significant
growth over time. A relevant question is whether there is increase in the proportion of advanced vocabulary used in a similar learning context as well or whether development of the sophistication of vocabulary use can typically only be expected in formal learning contexts where particular focus is placed on vocabulary.

One study which investigated the development of lexical richness in written texts is Laufer (1991: 441) who examined Israeli university students of English who were not “explicitly taught vocabulary”. The data consisted of timed compositions written by the learners, which were subjected to a lexical analysis measuring lexical variation, number of lexical words, lexical originality (lexical individuality) and lexical sophistication defined as the proportion of words found in the UWL (not to be confused with the Beyond 2000 measure outlined below). At the beginning of the term the students (N = 47) were required to write an entrance exam in the form of a free composition and at the end of the first term 22 learners were asked to write the same kind of composition. At the end of the second term, the remaining 25 learners were asked to write a composition on the same topic as the entrance exam.

Looking at the development of these learners, the following pattern emerged: there was no overall progress by group after one semester of study. However, after two semesters significant progress was made in lexical sophistication. Apart from lexical sophistication, there was no average progress in the two groups. Laufer (1991) concluded anyway that the increase of the proportion of academic words seems to be a predictor of increasing writing proficiency. Moreover, she noticed that the individual pattern of development showed that those students who were below average made progress in order to reach the average level in accordance with the departmental requirements. As noted in section 3.4.1 growth in learners’ vocabulary size seems to typically occur among learners with a relatively low onset vocabulary size relative to their peers. It is contended that lexical development is determined by the needs of the learners in that below average learners seem to adapt to the course requirements by increasing their lexical richness in writing and that learners who are safe from the point of view of fulfilling the course requirements do not make any progress in this respect (ibid.). Laufer’s (1991) conclusion as regards development is that there seems to be a threshold after which no significant progress in lexical richness takes place.

Laufer (1995) suggested that a possible measure of the active lexicon could be the percentage of non-basic words used in a sample of the learner’s language, namely, words that are not among the 2000 most frequent words in the language, hence the name of the instrument Beyond 2000 (B2000). Laufer (1995), using the B2000 profile on timed compositions written by 48 Israeli first year students of English, found that there was an increase of about three percentage points after one term and oddly enough only about
1.5 points after one whole year of study. All 48 students wrote a first timed composition at the beginning of term of whom 23 students wrote an additional composition at the end of term and 25 students wrote one at the end of the second term. So we are dealing with two different samples and this is probably why there is a greater increase in the beyond-2000 vocabulary after only one term as opposed to after two terms. However, in another study by Laufer (1998), in which high school students were tested on the same vocabulary knowledge types as in Laufer and Paribakht (1998), it was shown that there was no growth in lexical richness.

The development of learners’ vocabulary use has mainly been examined in inauthentic language tasks that have no or limited bearing on the students’ academic achievement. This approach to investigating learners’ vocabulary use is of limited benefit to increasing our knowledge of real language use. It might be the case that a different developmental pattern can be observed for the vocabulary use in a high-stakes writing task because of motivation.

Moreover the increase in lexical richness that has been found is only from cross-sectional data or from mixed longitudinal/cross-sectional data, which warrants a longitudinal approach in order to examine the development.

It might be the case that no development takes place among advanced learners at the university level, since they might already have acquired enough vocabulary for meeting the task demands or they might have reached a level of proficiency where the sophistication of their vocabulary use has reached a plateau. This should be further investigated. The following research question will thus be addressed:

19. How does the lexical richness of student essays develop over time?

If previous findings of a development of lexical richness in students’ written texts are applicable to the context examined here, we should anticipate a significant growth between first- and second-term take-home essays. Moreover as indicated in Laufer (1991) this growth should be limited to below-average students.

The following chapter describes how the theoretical framework has been operationalized.
4. Method

In this chapter the research design and the operationalization of the theoretical framework will be presented. This will comprise information about statistical concepts pertaining to test development, statistical tools employed in the analysis of the data, the informants, research instruments, data collection and data analysis. In addition, the development of the depth of vocabulary knowledge test including both native speakers and non-native speakers followed by the findings of a pilot study will be outlined. In the next section some basic statistical concepts that provide the necessary foundations for understanding test effectiveness and their application in the development of the productive depth of vocabulary test will be outlined.

4.1 Statistical concepts

In this section an outline of the standard statistical concepts pertaining to test development and how they are applied in the present study is provided. Moreover, this section presents and discusses the statistical inferencing tools employed.

4.1.1 Test usefulness

The principal consideration in the development and use of language tests is that of test usefulness, which according to Bachman and Palmer (1996) is defined as subsuming the following test qualities: reliability, construct validity, authenticity, interactiveness, impact and practicality. When demonstrating test usefulness, Bachman and Palmer (1996: 18) state “that test developers need to find an appropriate balance among these qualities, and that this will vary from one testing situation to another.” The specific test developed here is only for research purposes and not for student or program assessment, and therefore the demonstration of test usefulness will mainly include issues relating to the following two qualities: construct validity and reliability. Of course practicality has also affected the design.

In order to demonstrate construct validity “we need to provide evidence that the test scores reflects the area(s) of language ability we want to measure, and very little else” (Bachman and Palmer 1996). This was
achieved in the current study through a discussion and outline of the construct of depth of vocabulary knowledge (see section 2.5).

The second determinant of test usefulness is reliability, which refers to consistency of measurement across similar testing conditions (Bachman and Palmer 1996). There are many different factors that might have an effect on the reliability of a test. Bachman and Palmer (1996) provide the following framework which “can be used to guide both the design and development of a test, as well as to help us identify potential sources of inconsistency across test tasks” (Bachman 2004: 156):

- the setting
- the rubric
- the input
- the expected response
- the relationship between input and response

These factors were considered when developing the depth of vocabulary knowledge test. Another measure taken to obtain an acceptable level of test reliability was administration of questionnaires eliciting test takers’ view on factors such as clarity of instruction, time allotment and degree of difficulty (cf. 4.2.3.1). Furthermore, statistical analyses were employed estimating internal consistency of the test (cf. 4.1.2.1), which is a measure estimating “how consistent test takers’ performance on different parts of the test are with each other” (Bachman 1990:172).

The internal consistency of a test involves the extent to which test takers’ performance on the different parts of a test are consistent and is often measured in terms of a reliability coefficient computed with the help of formulas such as KR-20, KR-21 or Cronbach’s Alpha.

In the present study, the method of Cronbach’s Alpha was used, since in contrast to Kuder-Richardson formulae (KR-20 and KR-21) it can be employed to measure the reliability of test items that are polychotomously scored. It indicates how well a set of test items measures a single, unidimensional latent construct. The output of this method is a coefficient value somewhere at the interval of 0 to 1. The higher the value, the higher internal consistency a given test can be said to have. This value is affected by the number of test items and the size of the test-taker sample (Bonett 2002).

It must be stressed that sampling error arising from a small number of subjects might affect the reliability estimates. To minimize this risk Nunnally and Bernstein (1994) recommend a representative sample size as high as 300 persons. Accordingly, the reliability estimates in the present study must be interpreted in light of the relatively small sample. There is no universally accepted minimal cut-off point for what constitutes a satisfactory reliability coefficient. However, as a rule of thumb, a minimal adequate level

82
is often reported to be 0.70 or in some cases as low as 0.60 (Nunnally 1978 as cited in Bryman and Cramer 2001). DeVellis (1991: 85) provides the following useful benchmarks for interpreting Cronbach’s Alpha reliability coefficients:

<table>
<thead>
<tr>
<th>Range of Values</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below .60</td>
<td>Unacceptable</td>
</tr>
<tr>
<td>Between .60 and .65</td>
<td>Undesirable</td>
</tr>
<tr>
<td>Between .65 and .70</td>
<td>Minimally acceptable</td>
</tr>
<tr>
<td>Between .70 and .80</td>
<td>Respectable</td>
</tr>
<tr>
<td>Between .80 and .90</td>
<td>Very good</td>
</tr>
<tr>
<td>Above .90</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

The level of the alpha value of a given measure is to a great degree dependent on the characteristics of the group of test-takers. Reliability figures obtained among test-takers of widely-ranging proficiency are typically higher than figures obtained from more homogenous groups in terms of level of proficiency (Jones 2001). Reliability figures should accordingly be interpreted in light of factors such as the degree of variation in ability among the test-takers (ibid.).

Having discussed basic statistical concepts underpinning test development and how they are implemented in the development of the productive depth of vocabulary knowledge tests, in the next section the statistical inferencing analyses used for evaluating test reliability and for drawing inferences about the data will be discussed.

4.1.2 Statistical analyses

Three main statistical inferencing analyses are employed in the present study. The first one pertains to investigating test reliability and was outlined in section 4.1.1, whereas the other two are used for making statistical inferences about the sample. Depending on the data type, there are two main choices as to which statistical tests to employ. These are parametric or nonparametric tests. In the present study both types are used depending on the scale of measurement.

4.1.2.1 Correlation analysis

The Pearson Product-Moments correlation test measures the degree of a supposed linear association between two variables. It produces a correlation coefficient, the Pearson correlation (r), with a value between -1 to +1. A negative correlation will produce a negative coefficient, and a positive correlation will produce a positive coefficient. If one illustrates the degree of association between two variables on a scatterplot, the larger the absolute value (irrespective of it being negative or positive), the narrower the ellipse and the closer to the regression line the points in the scatterplot will fall. In
the case of no association between the variables the point will be scattered in a roughly circular shape in the scatterplot producing a coefficient which is zero or close to zero (Kinnear and Gray 2000).

With reference to interpreting correlation coefficients one has to rely on rules of thumbs. The following guideline for interpreting the strength of correlations is provided by Cohen and Holliday (1982 cited in Bryman and Cramer 2001: 174):

Table 8. Range of values for interpreting the strength of correlations.

<table>
<thead>
<tr>
<th>Range of values</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.19 and below:</td>
<td>Very low correlation</td>
</tr>
<tr>
<td>0.20-0.39:</td>
<td>Low correlation</td>
</tr>
<tr>
<td>0.40-0.69:</td>
<td>Modest correlation</td>
</tr>
<tr>
<td>0.70-0.89:</td>
<td>High correlation</td>
</tr>
<tr>
<td>0.90 and above:</td>
<td>Very high correlation</td>
</tr>
</tbody>
</table>

According to Kinnear and Gray 2000 it is important to inspect the scatterplot representing the relationship between two variables before one is able to draw any conclusions from the value of a Pearson correlation. This is due to two main reasons: first, data giving a zero or low Pearson correlation may show a very strong non-linear association on the scatterplot. Second, a single outlier that is far from the point of average can have a large effect on the correlation coefficient. Accordingly, the Pearson correlation is a measure of supposed linear relationship between two variables which must be confirmed by examination of the scatterplot.

The Pearson Product-Moments correlation test, which is a parametric test will be used in the present study only when all variables in the data set are measured on at least an interval scale. When one of the variables in the data set is measured on an ordinal scale the non-parametric equivalent of the Pearson correlation test, the Spearman R test, is employed.

In addition to correlation analysis, the association between variables is examined by regression analysis. Regression analysis is employed to shed light on the predictive value of an independent variable on the dependent variable.

4.1.2.2 T-test analysis

In the present study, the dependent-samples t-test to determine whether there is a statistically significant change in the mean scores between two test occasions, will be employed. This test will be used in order to make inferences about whether there is a statistically significant change in the mean score over time.

Having outlined the statistical analysis tools employed in the present study, in the next section information about the test instruments will be presented.
4.2 Instruments

In this section the three discrete-item tests of vocabulary knowledge employed in the present study will be described. These tests include two standard tests of receptive and productive size of vocabulary knowledge and a test of productive depth of vocabulary knowledge developed for the purpose of the study. In addition, the tool employed to measure the lexical richness of the learners’ written output, the LFP, will be outlined in this section.

4.2.1 The Vocabulary Levels Tests

There are currently two types of the VLT, a receptive and a productive (cf. section 2.4.3.1.2 and 2.4.3.2.1). Although measuring different aspects of vocabulary knowledge they are both based on a common framework. The VLT derives its name from the five levels reflecting separate frequency levels\(^{12}\) and which are supposed to gauge learners’ vocabulary knowledge at each level. In this way the tests are able to provide a profile of a learner’s degree of mastery of different frequency levels in addition to providing an estimate of overall vocabulary size (Schmitt, Schmitt and Clapham 2001).

The tests have been used in the research literature to calculate learners’ total vocabulary size. There are two principal ways of calculating the total vocabulary size based on the VLT scores. Schmitt and Meara (1997) assume that the proportion of correct answers at each level is the proportion of total words known at that level. By adding up the totals of each level we obtain an overall size estimate. Laufer (1998) has a more complicated calculation to obtain an estimate of the total vocabulary size in which the 1K and 2K level are assumed to have the same score, 4K level is taken as an average of the 3K and 5K levels. Since she did not use the 10K level, the sum is then multiplied by 5000. There is no information about how to calculate the overall size when using the 10K level.

However, according to Cameron (2002) due to there being a lack of comparability between the way size is calculated researchers should avoid working out an overall size. Her way of presenting the VLT scores as a reflection of learners’ size is to calculate the extent of mastery of each level. Mastery of a level is defined as obtaining a minimum of 89% of the maximum possible score for that level. However, lack of comparability is mainly due to the fact that a large number of studies utilizing the VLT do not in any clear way describe the method of calculating overall vocabulary size (e.g. Laufer 1996; Laufer and Yano 2001). By clearly stating which method is used the comparability problem can easily be avoided.

---

\(^{12}\) It should be pointed out that the AWL level does not represent a frequency band, since the words in the AWL are sampled from different frequency bands.
In the present study in order to describe the profile and development of the informants’ vocabulary size both the mastery and the overall vocabulary size approach will be employed. In accordance with Read (1988 cited in Read 2000) and Schmitt et al (2001) a score of 26 out of 30 (87%) for the RVLT, and 16 out of 18 (89%) for the PVLT is set as the cut-off point for mastering a particular level. The calculation of overall vocabulary size will be based on Schmitt and Meara’s (1997) method.

4.2.1.1 The receptive version

The first attempt to validate the test was carried out by Read (1988) who found the test to be reliable and found that the scores on the frequency levels tended to fall into an implicational scale, which lends support to the underlying idea that words are acquired in accordance with their frequency of occurrence. Further developments of the test were carried out by Schmitt in 1993, who revised the first version (Version A) and designed three additional versions (Versions B, C and D). However, no validation study was run on them at the time. Instead, Beglar and Hunt (1999) took the first steps to validate the four versions, in which they revised the 2K and UWL levels and found that they were essentially gauging the same underlying construct. They also obtained concurrent validity scores by correlating VLT scores with TOEFL scores of the same learners (Schmitt et al. 2001).

The two versions (version 1 and version 2) of the receptive VLT (RVLT) employed in the present study were developed and validated by Schmitt et al (2001). These two tests differ in two major ways from versions A-D. First instead of 6 clusters with three items in each level in the older versions, the revised versions contain 10 clusters with three items per frequency level. This adds up to a total of 90 ((6 clusters x 3 items) x 5 levels) and 150 items ((10 clusters x 3 items) x 5 levels), respectively. The decision to add additional items in the revised extended versions was based on a Cronbach alpha analysis which suggested that a minimum of 10 clusters are needed per level to produce reliability figures above .90. In addition, Schmitt et al (2001) replaced the academic sections based on the outdated UWL (cf. 2.4.1.2) with sections based on the Academic Word List (AWL; Coxhead 2000).

According to the authors, the test is highly practical since it can be completed in a reasonable amount of time, an average of 31 minutes, it is easily administered and scored, and “gives a more complete picture of a learner’s vocabulary knowledge than most other tests” (Schmitt et al 2001:72). Summarizing the results of a number of validation analyses, the authors conclude that the two revised versions “…can provide valid results and produce similar, if not truly equivalent scores.” (2001: 79).
4.2.1.2 The productive version

Unlike the RVLT, the PVLT has not been subjected to careful validation study. However, Laufer and Nation (1999) have carried out a small validation study of the PVLT. Moreover, two forms of the test with a satisfactory degree of equivalence were developed (Laufer and Nation 1999). With respect to their claims regarding the test’s practicality and reliability, although it is claimed to be easy to score, it differs from the receptive version since it involves a subjective dimension in the scoring in terms of deciding what should be considered minor spelling mistakes, especially when they are combined with grammatical mistakes. Accordingly, one would expect this test to be less reliable than the receptive version since scoring will probably differ somewhat between different raters.

For the purpose of the present study further changes were made to the two versions since they shared items with the RVLT versions. Each PVLT version shared target items with each RVLT version. Moreover, since the PVLT versions’ academic levels are based on the now obsolete UWL, it was decided to develop a new academic level consisting of items found in the AWL to achieve equivalence between the PVLT and RVLT in terms of ascertaining that the two tests are based on the same level definitions. In order to have one receptive version and one productive version that do not share any items, items were swapped from the two productive versions as well as from four original versions (A-D). In Table 9 the number of items swapped and the version they originated from is displayed.

<table>
<thead>
<tr>
<th>Levels</th>
<th>No. of items transferred from version, A, B, C or D to version 1</th>
<th>No. of items transferred from version, A, B, C or D to version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2K</td>
<td>8B</td>
<td>8C</td>
</tr>
<tr>
<td>3K</td>
<td>13D</td>
<td>13C</td>
</tr>
<tr>
<td>5K</td>
<td>12A, 1B</td>
<td>12C, 2B</td>
</tr>
<tr>
<td>AWL</td>
<td>5B, 1A, 2D</td>
<td>3A, 2C</td>
</tr>
<tr>
<td>10K</td>
<td>3D</td>
<td>3C</td>
</tr>
</tbody>
</table>

The numbers in the two columns to the right represent number of items swapped from each frequency level and the letter to the right of each number shows from what versions the items originated.

Since the swapping of items between the different levels in order to limit the number of shared items in the receptive and productive versions, and since the creation of a level only consisting of AWL words will affect the degree of difficulty of the two productive versions, an assumed different degree of difficulty is counterbalanced. In order to ascertain that the modified versions have an acceptable level of internal consistency an Alpha reliability analysis was carried out (cf. section 5.1.1). As mentioned in
section 2.4.3.2.1, the PVLT should be regarded as primarily a form-recall test. Accordingly, in the present study, size of productive vocabulary knowledge is operationalized as form-recall.

4.2.2 LFP/B2000

In the present study the condensed profile, the B2000, is used as a measure of lexical richness (cf. section 2.6.1.2.2 for a detailed description of the LFP). The profile is attained by adding up the number of word families belonging to the AWL list and not-in-the-lists list and dividing it by the total number of word families identified by Range in the text analyzed.

In regard to the degree of reliability and validity of the measure, Laufer and Nation (1995: 319) found that it is able to discriminate between different proficiency levels and that it produces scores which are stable between two compositions written by the same learner within a one-week period. Some additional evidence of its validity was provided by Daller and Xue (2007) who showed that the B-2000 measure was able to discriminate between two proficiency levels. Moreover, they obtained evidence of concurrent validity; B-2000 scores were found to correlate highly with scores of other measures of lexical richness.

In order for the programme to yield sensible data one has to edit the texts before running them through the programme. In the present study the text editing procedure used by Laufer and Nation (1995) was adopted. The following method was employed: all words that were clearly used incorrectly in terms of meaning were omitted. In addition, proper names were omitted. As regards misspellings they were corrected and wrong derivatives of a word were kept. In the present study also nouns and adjectives denoting nationality were omitted since frequent use of less frequent nationality words might skew the profile. Since the texts in the present study consist of take-home essays for which the subjects might make use of citations, I also decided to omit all text passages that were deemed to be taken verbatim from another source. Apart from these nothing else was omitted.

Some manual post-processing of the LFP was carried out. Since the LFP is not able to categorize not-in-the-list words (i.e. words that do not appear in the 1K, 2K or AWL list) into word-families, all words that were not screened out according to the above mentioned criteria were assigned to word families manually. Not-in-the-list words were arranged into word families according to level 3 on Bauer and Nation’s (1993) scale.

Laufer and Nation (1995) perceive the LFP as a measure of the vocabulary size that can be used at free will. Referring to the inherent weakness of previous measures of lexical richness, they state that:

There are many factors besides vocabulary size that could affect lexical richness in writing. These could include familiarity with the topic, skill in
writing, and communicative competence. This means, for example, that a change of topic could result in marked change in lexical richness (Laufer and Nation 1995: 308).

The authors’ argument is that the LFP provides an objective picture of a learners’ vocabulary size. In other words, two learners with different vocabulary sizes who write under the same conditions will produce texts with different lexical frequency profiles. Although one can argue that the lexically richer texts a learner writes the larger vocabulary s/he has, too many factors are at play that affect the lexical richness of a text that are not necessarily associated with vocabulary knowledge, such as higher-order processing skills. Accordingly, I would not go so far as claiming that the LFP is immune to these factors. Instead, in the present study, the LFP will be employed as a measure of vocabulary use as opposed to a direct measure of vocabulary size. Whereas vocabulary knowledge is more static, vocabulary use varies according to context.

4.2.3 Depth of vocabulary test

Since there are no standard tests available measuring productive depth of vocabulary knowledge conceptualized according to the components approach, the depth of vocabulary knowledge test was the only test developed specifically for this study. Three components were selected for testing based on considerations of their importance in academic tasks and practicality to include in a test. 20 target words were randomly selected for each of the two versions of the test, two from each of the ten sub-lists of AWL (i.e. 40 test items altogether). All twenty target words in each of the two versions were tested for the informants’ knowledge of their collocational preferences, word class derivations and synonyms.

Words collocate in many different ways and to different degrees and according to Nation (2001) this complex nature of collocations is reflected in the many different definitions and criteria for classifying words that frequently co-occur in units available in the research literature. The most effective way to set up criteria for defining collocations is through the use of a set of scales (Nation 2001). The large number of scales needed is indicative of the wide range of different types of collocations. He presents ten such scales used in the research literature to define collocations. The two Nation (2001) terms frequency of occurrence and adjacency which reflect how frequently items occur together and how closely together they occur were used when selecting collocates to the test items.

Table 10 displays an extract of the first version of the depth test. Knowledge of the target words’ collocational preferences was gauged by presenting a sentence including the test item and a blank.
Table 10. Extract of the first version of the depth of vocabulary knowledge test.

1 a) What exactly the author meant by that statement is _________ to interpretation. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What word class forms can interpretation be realized as in addition to the one in the sentence? (Write one correct word form per slot)

<table>
<thead>
<tr>
<th>Noun</th>
<th>Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjective</td>
<td>Adverb</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c) Provide two synonyms or near-synonyms of interpretation.
1. ____________________2. ____________________

The majority of the collocates (38) were chosen from the Longman Dictionary (LD) phrase bank which lists a given word’s most common collocates. The LD phrase bank also contains example sentences in which the test item and its collocate occur. The collocates were selected in a random fashion from the phrase bank. However, collocations (the collocate and the test item selected from the AWL) that were deemed to be rare in terms of the extent to which an adult Swedish learner might encounter it, were discarded. Collocations that occurred in example sentences that were too opaque or contained too difficult words were discarded, since such sentence items could unnecessarily impede the elicitation of the intended target item. The remaining two collocates were found by searching for the collocational profiles of the test words in the BNC in which the 50 most frequent collocates are displayed. The window span for the search was set at the default setting, i.e. a window span of 5 words. Out of these fifty most common collocates, the most common lexical collocate was selected (as opposed to function words, such as of, the, in etc.). By selecting the collocate, the concordance of the collocation can be accessed and thus an appropriate sentence for the test can be extracted. Adjacency was taken into account by only selecting collocates which are separated by prepositions or particles as opposed to collocates that are separated by lexical words or longer stretches of words, (e.g. parliament passed a highly controversial offshore drilling bill).

Since the target words for which to select appropriate collocates had already been selected. I could not choose target words that occur in highly lexicalized units and which could be tested in isolation due to there being few or no collocational alternatives. Instead, after having selected the target words, I had to find appropriate collocates and since the degree of fixedness of the collocations varies, they had to be placed in the context of a sentence that limits the alternatives. Finding appropriate sentences that were clear
enough and that restrict the number of alternatives also affected the selection of collocates in that I was sometimes forced to select new test items.

A collocate of the prompt word was judged as correct if it agreed with the context and if the combination was listed in the LD phrase bank. For example, instead of *adopt an approach, take an approach*, was judged as correct since it agreed with the sentence context provided in the test and it was listed in the LD phrase bank as an appropriate collocate of the prompt word, *approach*. As for the two collocations taken from the BNC, in addition to the two above-mentioned criteria, the specific collocate to the prompt word found in the BNC if supplied by the test-taker it was judged as correct. It must be stressed that no claims are made that the scoring criteria adopted can be used to identify all possible collocates in the context of the sentence supplied or that all of the responses scored as correct would be judged as correct in relation to other standards of scoring. This reservation also applies to the scoring criteria adopted for the other two sub-tests described below.

Knowledge of derivations was tested by supplying four blanks in which the test takers were supposed to write a form related to the test item with a different given word class (cf. table 10). Baseline data for judging correct answers were the two following English dictionaries: Collins Cobuild English Dictionary for Advanced Learners, and Concise Oxford English Dictionary.

Knowledge of synonyms was tested by supplying two blanks in which the test takers were prompted to supply two synonyms or near-synonyms of the target word. The baseline for judging an answer as correct was the *Oxford Thesaurus of English*.

4.2.3.1 Validation of the first version of the depth test and the development of a second 'parallel' version

There were two main considerations that guided the process of developing a depth of vocabulary test. First I wanted to have a representative sample of words from the AWL that reflected the informants’ overall knowledge of the academic vocabulary. The rationale for sampling test items from the AWL was that knowledge of academic words is commonly regarded as particularly important in order to successfully manage academic studies (e.g. Corson 1997, Coxhead 2000 and Nation 2001). This was carried out by selecting a large enough sample of words which at the same time would be practical to test. The number of words that were deemed practical to assess was 20. Moreover, as noted above, I sampled two words randomly from each sub-list. The second consideration guiding the process of test design was that I wanted to avoid the risk of the same words reappearing in the pre- and post-tests. This consideration was influenced by the testing methodology adopted in Laufer (1998) and Laufer and Paribakht (1998), who used two parallel versions of the receptive and productive VLT in the pre-test and post-test sessions. In other words, the rationale for using two parallel tests was to
avoid the risk of a ‘practice effect’. For example, there could be a risk that the informants actively studied or discussed the test items among each other which could have led to the test items being retained in their memory. This led me to design two parallel versions of the depth of vocabulary test with 20 test items in each version.

According to Bachman (2004), parallel tests must meet three conditions in order for the two sets of scores generated to be regarded as parallel. First they must have equal means. Second their variances must be equal and thirdly they must be equally correlated with a third measure of the same ability. In addition, each test form should be assembled according to the same test blueprint. This consideration was met in the study by developing a second version based on the same test blueprint as the first version.

In the development of the two versions of the depth test only the collocation sections were developed based on the second consideration. The main reasons for this were lack of time in that the longitudinal design of the present study limited the amount of time that could be spent on test development, and, most importantly, difficulty finding informants on whom I could pilot the complete two versions. Informants were highly reluctant to complete all the sub-sections (i.e. collocation, derivation, and synonyms) of the two versions.

The first step of the development and validation process of the depth test, involved the pre-testing of the collocation part on eight native speakers. It was deemed necessary to obtain native-speaker norms and use the native-speaker responses in the process of developing and refining the sentences. The native-speaker responses were mainly used in order to identify items or sentences that were too difficult to elicit an appropriate collocate and also in order to identify blanks that attracted too many high frequency words that are open-ended in terms of their collocational profile, such as certain function words (e.g. of, the, in, etc.) certain verbs (e.g. have, got, etc.) and certain adjectives (e.g. great, small, wide etc.).

Based on the native-speaker responses, I changed items for which a majority of the eight native speakers could not provide the intended target collocate (8 out of 20 items). The changes ranged from minor changes of the wording of the sentences to selecting entirely new test items, which was the case with two items, and were made mainly in terms of providing a context to the collocates that minimized the number of potential correct answers. The pre-testing on native speakers and non-native speakers (cf. phase 2 below) enabled me to improve the context sentences along these lines, but also made me realize that it is very difficult to entirely eliminate the possibility of test takers providing alternative answers that might be idiomatic in addition to the intended one (see section 4.2.3 for details about the baseline data used for judging the answers). Therefore, in order not to risk excluding possible collocates all answers were judged as correct if they are listed as common according to the employed baseline data. Of course one way to limit the
responses to the intended one is to supply a varied number of initial letters as the case is in the productive VLT. The reason that method was not adopted here is that I wanted the test to be as much as possible in the productive end of the continuum (cf. section 2.2). Moreover, a second aspect pertaining to practicality was that in order to find the appropriate number of initial letters I had to repeatedly pre-test the test on native speakers in order to obtain the exact number of letters needed for the native speakers to provide the correct collocate. Since I had no way to offer them financial compensation for their time, I could only rely on their help on one occasion.

In the second phase a complete first version with a collocation section, a word class derivation section and a synonym section was piloted on 7 non-native speakers enrolled on the first or second term of the English literature and language course at Stockholm University. They were given 60 minutes to complete the test. After finishing the test they were given a questionnaire asking them to assess the test format in terms of clarity of the test instruction, the amount of time allotted to complete the test and the test’s degree of difficulty\(^\text{13}\).

In Table 11, the respondents’ assessment of the amount of time afforded to complete the test is displayed.

<table>
<thead>
<tr>
<th>Far too little</th>
<th>Too little</th>
<th>Appropriate amount of time</th>
<th>Too much time</th>
<th>Far too much time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you have enough time to complete the test?</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All the respondents stated that they had either too little or far too little time allotted to finish the test. Since, the informants in the present study have to do two tests apart from the depth of vocabulary test, there was no room for adding more time to complete the test. However, out of the 34 who did the test in the main study only about five used up the whole 60 minutes, which suggests in retrospect that the amount of time allotted for the depth of vocabulary test was sufficient for the vast majority of the test takers.

In regard to the question whether the instructions were unclear, only two out of eight answered yes. In accordance with the suggestions of the two who perceived the instructions as unclear, the instructions were improved.

The third question concerned the test takers’ experienced level of difficulty of the depth of vocabulary test. Two test takers stated that the test was very difficult and six that the test was a little difficult. It must be kept in

\(^{13}\) The questionnaire employed to this end was adapted from Staehr Jensen 2005
mind that, since the words tested are academic words with which even native speakers might have difficulties in terms of depth of knowledge, it is quite natural that the nature of the tested target words contributes to the level of difficulty of the test (cf. Macmillion and Shaw 2009). The intrinsic difficulty of academic words and low-frequency words can be expected to be higher than that of high-frequency words, since learners probably have a deeper knowledge of high-frequency words than of low frequency words. Moreover, we are dealing with a measure of productive knowledge, which also contributes to the level of difficulty of the test. The rationale for developing a relatively difficult test was to investigate whether the ability to find appropriate collocates and to produce word class derivations and synonyms of academic words were related to lexical richness of take-home essays, or whether the nature of the conditions under which the essays were produced evened out differences in depth of vocabulary knowledge due to the time allotted to complete the essay and the accessibility of different aids such as dictionaries, spell-checker and thesauruses.

Based on the responses from the non-native test takers, further alterations to the test were made. The alterations consisted of replacing problematic items that were too difficult or too easy and as a consequence would not discriminate high-performance learners from low-performance learners.

Step three involved the development of a second version. Due to the difficulty of obtaining two sets of 20 academic words in a random selection that have a similar degree of difficulty and a similar number of word class derivations, only the collocation section was developed into two parallel versions. However, as noted, the principal reason for not being able to pilot the complete test was a practical one, namely I could not find enough participants who were willing to spend time on the complete test. Accordingly, I had to pilot the section that would benefit from being piloted the most. The word class derivation and synonym section consisted only of a prompt to supply the correct word class derivation of synonym on a blank, whereas in the collocation section, the test item is presented in the context of a sentence that could be refined through piloting. Thus, the collocation sections involved more error sources, which made it more important to pilot.

Developing the second version involved the following procedure: 40 new items were selected and together with the 20 items from the first version they were administered to 22 non-native speakers who were all university students in other subjects than English. The non-native speakers were divided into two groups and both groups were administered twenty items each which are different from the ones in version one. I then calculated the mean, standard deviation and variance for all three groups.

In Table 12 we see the mean score, standard deviation and variance of the 20 items in the first version of the depth of vocabulary test and twenty items per group pre-tested for inclusion in the second version.
Table 12. *Mean score and standard deviation scores on collocations test part.*

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version 1 group</td>
<td>22</td>
<td>9.00</td>
<td>4.25</td>
<td>18.05</td>
</tr>
<tr>
<td>Version 2 group 1</td>
<td>10</td>
<td>9.80</td>
<td>5.63</td>
<td>31.73</td>
</tr>
<tr>
<td>Version 2 group 2</td>
<td>12</td>
<td>8.33</td>
<td>3.39</td>
<td>11.51</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>9.44</td>
<td>3.96</td>
<td></td>
</tr>
</tbody>
</table>

The version-2-group-2 items were subsequently selected since the mean score, standard deviation and variance came closest to the version one items.

Moreover, the covariance scores on both sets of test items are satisfactory, being above .90. It must be stressed that the process of obtaining two parallel measures in the present study, as far as the collocation section is concerned, is very limited. For instance, Schmitt, et al. (2001) when validating and developing two parallel versions of the receptive VLT not only obtained much closer scores for the mean and the variance, but also had access to as many as 801 subjects on whom the two tests were piloted. They also employed more statistical analyses, such as factor analysis, in order to determine whether the test items in the two versions had similar degrees of difficulty. However, they do stress that it has always been difficult to develop different versions of a test that are truly equivalent.

Since it was not possible to pilot the complete tests, it became evident at a later stage that the word derivation sections in the two parallel versions had different possible maximum scores. This led me to abandon the parallel version approach in regard to the depth test for the second group of informants who started their first term of study after the first group of students had finished their first term. Accordingly, the second group of students were administered the same version, viz. the second version, in the pre-and post-tests. As regards the first group of students, they were administered the first version at the beginning of term one, the second version at the end of term one and again the first version at the end of term two for the ten informants in this group who proceeded to a second term. In order to be able to run statistical tests on the whole sample, the scores on the two versions of the word derivation sub-test were equated by transforming the scores into percentages. When comparing the mean scores between the first and second data point only the sub-sample who were administered the same versions at both data points were included in the analysis. However comparisons of the mean scores on the depth tests between the first and the third data point were run on the whole sample since the administration of the depth test was counter-balanced across the first and third data point.
4.3 Informants

In this section information about the informants will be given, specifying the English learning context, the selection criteria adopted, the recruitment procedure and a presentation of the profile of the informants.

4.3.1 The role of English in Sweden

In order to fully make sense of the results pertaining to the informants’ vocabulary knowledge and use, we must first understand the wider learning context in which they acquire English.

With respect to the role of English in Sweden, English is typically introduced as a school subject as early as the third school year, i.e. when the children are nine years old. There is a high exposure to spoken English due to easy access to different forms of English media and due to most television programs and films being subtitled and not dubbed.

English as a subject in university studies is studied as a foreign language. At the time of the current study (i.e. the beginning of the Fall term of 06 – the end of the Fall term 07) the English course objectives at Stockholm University emphasized literature and linguistics, in particular contrastive grammar and phonetics and on developing their written and to a lesser extent spoken proficiency. There was, however, a lack of focus on vocabulary development presumably due to the view among curriculum developers that as a result of the high exposure to English the students are expected to have highly developed vocabulary knowledge. As stated in the introduction, one of the rationales for examining vocabulary knowledge among Swedish university students is this inherent tension between high expectations of their vocabulary knowledge and the lack of focus on vocabulary development in English university courses. In view of this tension, it is of interest to further investigate university students’ English vocabulary knowledge.

In the present study degree of academic achievement is defined as the course grade obtained (CGRADE), and text quality is defined as essay grade (EGRADE). The grades at the time of the first-term data collection were set on a three-point scale: Fail (F) which is given to students who either had not completed all required assignments or who despite having completed a given assignment failed to reach the minimum criteria for the next grade on the scale, Pass (P). The highest possible grade a student could obtain for the entire course or a single assignment was Pass with Distinction (PwD.). However, the grading system at the department was subsequently changed in the middle of the study to a six-point scale. In order to be able to compare the informants on the basis of grade, the new six-point grading scale was transformed into a three-point scale according to established convention.
4.3.2 Selection criteria
The informants were required to be non-native speakers of English, since this is a study of second language acquisition. However, apart from this criterion, no language background criteria were adopted. The rationale for this was the inclusion of as large and thus as representative a sample as possible. Second, in order to make sure that the informants included had a similar amount of studies of English at the university level, the subjects should have entered an English course at the university level for the first time. Moreover, only full-time students were selected since one of the rationales for this study is to investigate if vocabulary knowledge of advanced learners of English develops over a one-term and two-term period of full time study in an English language university programme.

On the basis of these criteria, 34 first term students of English at Stockholm University volunteered to participate in the study and actually did so out of whom 16 proceeded immediately after their first-term to a second term of study.

4.3.3 Recruitment procedure and feedback
Recruitment of the informants was carried out on two occasions, i.e. at the beginning of the fall term of 2006, and the spring term of 2007. In connection to one of the first lectures of the beginner level at the English language course, the students were informed about the study and the selection criteria and those interested in participating wrote their contact details on a list. About 80 students indicated initial interest in participating in the study, but out of these only 34 (i.e. 17 in 06 and 17 in 07) participated in the study throughout the entire term. Most of those who showed interest in participating in the study dropped out before the first test administration.

The informants studied English either as a separate course or as part of a teacher training programme. Three incentives for securing as many volunteers as possible in the study were offered: two cinema tokens given after each testing session, an offer on my part to help students with linguistics assignments and an offer of individual feedback on their results on the size of vocabulary tests.

4.3.4 Profile of the informants
A questionnaire was administered which collected demographic details and a variety of information on language background and use. The language use material is not reported here. The sample consisted of 27 female informants and 7 male informants. This distribution reflects the gender distribution among the student population enrolled on the first term of the English
Language and Literature programme at the University of Stockholm. In Figure 4, the age range of the sample is displayed.

![Bar chart showing age ranges for informants (N = 34).]

**Figure 4.** Age ranges for the informants (N = 34).

As we can see in Figure 4, 54% (19) of the informants were between 20-25 years old. Another large group were student who are 19 or under compromising 22% (N = 7) of the sample. Only four of the informants were between 26 and 30 years old. All in all only 14% (N = 4) of the informants were above 30.

Information about the informants’ place of birth was collected and showed that 30% of the sample stated that they were born in another country. Out of these, 3 informants stated that they had another mother tongue than Swedish, whereas for the ones that stated they were born in Sweden, one person reported having a different mother tongue than Swedish. All in all almost 90% stated that their mother tongue was Swedish.

In Figure 5 we can see the age ranges of the participants when they started studying English formally at school.

![Bar chart showing age range when informants started studying English in school (N = 34).]

**Figure 5.** The age range when the informants started studying English in school (N = 34).

None of the participants started studying English at school after age eleven, which corresponds to the fourth year at the Swedish elementary school. In the Swedish school system English must be introduced in the fourth school
year at the latest and the variation found among the informants reflects the Swedish school system, in which English can be introduced at different years up to the fourth year.

In summary the sample represents the population in terms of age and ethnic background. As regards the mean age at which students taking the English Language and Literature programme start studying English at school, no data exists with which a comparison with the sample can be made.

4.4 Data collection

This section outlines the data collection procedure and comprises details about test administration and the collection of student essays. A pilot study was also carried out. The details of pilot data collection and the results are presented in section 4.5.

4.4.1 Test administration

As stated in connection to the description of the depth test a parallel-test design was adopted. To minimize any practice effect parallel versions of all three discrete tests of vocabulary were employed in the current study. Since, no two tests can be said to be completely equivalent, a counterbalanced approach was adopted as a way of balancing differences in degree of difficulty.

As noted above, two versions of the RVLT, the PVLT and the depth test (DEPTH) were used, henceforth version 1 and 2.

As shown in Table 13 for the first group of informants (N = 17) who started their first term in the fall of 06 the following test versions were given at the beginning of their first term: the RVLT 2, PVLT 1 and DEPTH 1:

Table 13. The sequence of test administration at the beginning of the informants’ first term (Fall 06: N = 17; Spring 07: N = 17).

<table>
<thead>
<tr>
<th>Beginning of first term</th>
<th>RVLT Version</th>
<th>PVLT Version</th>
<th>DEPTH Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 06</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Spring 07</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The informants who started their first term in spring term of 07 were given the 1st version of the RVLT and the second versions of the PVLT and DEPTH.
Table 14 presents the order in which the tests were administered at the end of the informants’ first term.

**Table 14. The sequence of test administration at the end of the informants’ first term (Fall 06: N = 17; Spring 07: N = 17).**

<table>
<thead>
<tr>
<th>End of the first term</th>
<th>RVLT Version:</th>
<th>PVLT Version:</th>
<th>DEPTH Version:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 06</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Spring 07</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Since the parallel version approach for the administration of DEPTH was abandoned for the second group of informants (i.e. the spring-07 group), they were administered the same version of DEPTH at the end of their first term. Since the cancellation of the counterbalanced approach in the administration of DEPTH was carried out after the first testing period (i.e. fall of 06), the DEPTH test was administered to the fall-06 group according to the initial test administration procedure (i.e. beginning of the first term DEPTH 1, end of the first term: DEPTH 2, end of the second term: DEPTH 1) (cf. 4.2.3.1). However, the counterbalanced procedure was maintained for the two vocabulary size tests for both groups. The order of test administration in the second term of study differed as well between the two groups as shown in Table 15.

**Table 15. The sequence of test administration at the end of the informants’ second term (Spring 07: N = 6; Fall 07: N = 10).**

<table>
<thead>
<tr>
<th>End of the second term</th>
<th>RVLT Version:</th>
<th>PVLT Version:</th>
<th>DEPTH Version:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring 07</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fall 07</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

In order to be able to run correlation analyses on the derivation sub-test scores for the whole sample and in order to be able to run paired-samples t-tests between the beginning of term one and end of term two scores, the derivation sub-test scores were transformed into percentage. That way I was able to obtain sub-test scores with the same resolution.

The analyses of the relationship of vocabulary knowledge to lexical richness and course grade and the relationship of lexical richness to essay and course grade are based on the two end-of-term correlations since lexical richness and grades are here perceived as primarily end of term measures.

In section 2.6 three comprehensive frameworks of vocabulary knowledge and ability were outlined. Chapelle’s (1994) model of vocabulary ability will be used as a basis for providing an attempt at explicating what facets of
vocabulary knowledge and ability the measures employed here are indicators of. Table 16 displays how the different vocabulary abilities were measured.

Table 16. *Test specification of the discrete vocabulary tests: RVLT, PVLT and DEPTH.*

<table>
<thead>
<tr>
<th>Components of vocabulary knowledge</th>
<th>Measure</th>
<th>Type of knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vocabulary size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word recognition</td>
<td>RVLT</td>
<td>Receptive/</td>
</tr>
<tr>
<td>Form recall</td>
<td>PVLT</td>
<td>Meaning recognition</td>
</tr>
<tr>
<td><strong>Knowledge of word</strong></td>
<td></td>
<td>Productive/</td>
</tr>
<tr>
<td>characteristics/Depth</td>
<td></td>
<td>Form recall</td>
</tr>
<tr>
<td>Ability to supply the correct collocate to an academic word in context</td>
<td>Collocational knowledge sub-test</td>
<td>Productive/ Form recall</td>
</tr>
<tr>
<td>Ability to supply appropriate word class derivation forms of academic words/metalinguistic knowledge</td>
<td>Word derivations knowledge sub-test</td>
<td>Productive/ Form recall</td>
</tr>
<tr>
<td>Ability to supply two synonyms to academic words/metalinguistic knowledge</td>
<td>Knowledge of synonyms sub-test</td>
<td>Productive/ Form recall</td>
</tr>
</tbody>
</table>

The first column contains a specification of the dimension of vocabulary knowledge the respective tests tap into. The term “knowledge of word characteristics” is taken from Chapelle’s (1994) model of vocabulary ability. It should be pointed out that the cover term for this dimension of vocabulary knowledge employed here is depth.

The third column displays what stage of the R/P continuum the discrete-item tests tap into (cf. section 2.2). A distinction between the degrees of productive mastery tapped by the PVLT and the depth test should be made. Although both come closest to a form-recall test, it should be pointed out that fewer cues are provided in the depth test than in the PVLT which makes it a more productive test.

In order to guard against the possibility that a possible increase in vocabulary knowledge might be due to exposure to English during the longer summer break, the informants belonging to the spring 07 group were asked if they spent time learning English during the summer break. All except one stated that they did not study any English or communicate in English during the summer break. One student stated that he had read some English books and done a few vocabulary exercises in a vocabulary book. This amount of exposure to English was not deemed as a sufficient ground for the exclusion of this particular informant.
4.4.2 Collection of student texts

All informants were asked to hand in an electronic copy of all essays they had written as part of their continuous-assessment examination in their first and second term of studies. The first group of informants who studied at the beginner level in the fall of 06 had written three literature “close readings” and one linguistics essay.

When grading the essays, the teachers adopted different grading procedures. The majority of the informants had to hand in three close readings of literature belonging to three different genres: poetry, fiction and drama. A fourth close reading that was graded was handed in at the end of the term. The students were allowed to choose one of the earlier three close readings as their fourth close reading to be graded, and they could either leave it as it was or improve it. Three of the subjects had all their four essays graded, since their teacher worked according to a different system.

The second group of informants who started their first term in the spring of 07 got all their three close readings graded except for three informants whose teacher still worked by the previous system of only marking a fourth close reading.

As for the linguistics essays, they were all graded. 16 of the informants proceeded to a second term and the number and type of essays they had to write varied according to which courses they had chosen. However, all students in their second term were required to write a term paper on a linguistic topic of their choosing called a B-essay. For different reasons not all informants completed their essay assignments which led to missing data for some students. Consequently, for the reasons mentioned above and because information on the grade on some essays could not be retrieved, for some students essays grade information is missing.

In Table 17 the number of essays and grades available for each informant is displayed. A capital C signifies that a copy was available and a capital C with an asterisk signifies that the grade for that particular essay was available.

As outlined in Table 17 informants 1-17 were enrolled at the beginner level in the fall of 06 and informants 17-34 were enrolled on the beginner level in the spring of 07.

For 15 out of the first 17 informants, there is a full set of essays and grades available including one informant who wrote three close readings that were all graded. For two informants, no. 11 and no. 14 some data is missing. For no. 11, the linguistic essay is not available, and for no. 14 two close readings are not available.
Table 17. The number of essays handed in and graded during the informants' first term of study.

<table>
<thead>
<tr>
<th>Informants</th>
<th>1st close reading</th>
<th>2nd close reading</th>
<th>3rd close reading</th>
<th>4th close reading</th>
<th>Linguistics essay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>2.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>3.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>4.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>5.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>6.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>7.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>8.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>9.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>10.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>11.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>13.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>14.</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>15.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>16.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>17.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>18.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>19.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>20.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>21.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>22.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>23.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>24.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>25.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>26.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>27.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>28.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>29.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C*</td>
<td>C*</td>
</tr>
<tr>
<td>30.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>31.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>32.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>33.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
<tr>
<td>34.</td>
<td>C*</td>
<td>C*</td>
<td>C*</td>
<td>-</td>
<td>C*</td>
</tr>
</tbody>
</table>

For the second group of informants complete essay and grade data is available for all 17 informants including three informants, nos. 26, 28 and 29, who only had their fourth close reading graded. The next section reports on the results of a pilot study.
4.5 Pilot study

In order to refine data collection procedures, instruments and research design, a pilot study was conducted involving 16 learners of English as a foreign language enrolled on the English Language and Literature course at Stockholm University. The pilot study was conducted over one term of study in the fall term of 2005.

Both the receptive and the productive VLT were used as well as two types of depth tests that were subsequently replaced by the depth test employed in the main study. The two parallel versions of each of the two types of the VLT were employed. All three tests were administered at the beginning and the end of the term. In addition one literary and one linguistics essay from each of the informants were collected and subjected to the B-2000 analysis. Two different depth of vocabulary test formats were used before settling for the current format. The first format was in the form of an adapted version of the VKS. The target word was presented in the context of a sentence in Swedish. The test taker was prompted to tick yes or no to the question whether he/she knew the English translation of the target word and its collocate. Depending on the test takers’ degree of knowledge of the target word, if the yes box was ticked, he/she could supply an English translation of the target word or the complete noun phrase it occurred in. The test taker was also asked to supply a derivation and synonym of the target words. Points were awarded according to how much information about the target word the learners could provide.

This test format was abandoned for the following reasons. First, conceiving depth of vocabulary knowledge in the form of a neat scale is problematic (cf. Waring 1999). Second, providing the Swedish translation of an intended target word does not mean that the test takers have the same target word in mind. Although, I tried to limit the responses to only the target word by supplying initial letters many of the test takers either ignored the cue or did not notice it. Since many of the target words have a great number of synonyms that almost equally correspond to the Swedish word, one might question what words are tested, the intended target word or another word in the mind of the test taker. A third problem involved finding suitable Swedish translations of the target word and its collocate. In some cases there was no one to one correspondence with the English target word and the Swedish translation. This problem is connected to the question of what is tested, vocabulary knowledge or translation skills, since the test takers in such a format would be expected to know when to translate the Swedish word in a literal manner or a figurative one.

In the development of the second test, the VKS format and the translation section were abandoned in favor of a format that attempts to operationalize the components approach. Table 18 displays a section from the second type of depth of vocabulary knowledge test employed in the pilot study.
Table 18. A section of the second pilot test format based on Nation’s (1990; 2001) components framework.

| 1. a) ____________ vision
  | blurred | defective | distinct 
  | incorrect | indistinct | imperfect 

b) What other word class forms can this word be realized as (write the correct word form)?

Noun__________  Verb__________
Adjective__________  Adverbial__________

c) Provide two synonyms or near-synonyms of this word

1. ________________________ (degree of similarity from 1 (most) to 5 (least) _____
2. ________________________ (degree of similarity from 1 (most) to 5 (least) _____

d) Can you think of two meanings that the word has (give the Swedish translation or a short definition)?

1. ____________________
2. ____________________

In this depth test (PILOTDEPTH) four components were operationalized collocational knowledge (COLL), knowledge of word class derivations (WD), knowledge of synonyms (SYN) and knowledge of polysemy (POLY). The test items in this test consist of 20 academic words randomly sampled from the AWL.

In the first part, the target word is presented with a gap indicating where a collocation should be inserted. The test takers are presented with five distractors and one correct alternative.

In the second part the test takers are supposed to provide all the possible word class derivations of the target word, apart from the word class form in which the test item is presented in the test.

In part three two synonyms are requested and in part four the test takers are prompted to supply two meanings the word might have in the form of a Swedish translation or a short definition in English. This format was subsequently adapted into the depth of vocabulary knowledge test format employed in the main study. I decided to condense the test by discarding part four of the test and the part eliciting collocational knowledge was changed by making it more productive in nature. This resulted in the third depth of vocabulary knowledge test format described in section 4.2.3. In the next section the results of the pilot study will be presented and discussed.
measures employed are as follows: the R- and PVLT, the PILOTDEPTH described in this section and the B2000 measure of lexical richness.

4.5.1. Analyses and results of the pilot study
As mentioned the pilot study was conducted in order to explore three methodological aspects i.e. collection procedures, instruments and research design. In addition the intention was to obtain tentative indications as to the development of vocabulary knowledge and relationships between the main variables of the study.

4.5.1.1 Step 1: Descriptive statistics of the test scores
The first step\(^\text{14}\) of the presentation of the results of the pilot study consists of the descriptive statistics summarizing the informants’ performance on the tests. The results will not be discussed in detail. Only two points will be dealt with, the degree of difficulty of the tests and the score distribution. In Table 19 the descriptive statistics of the test scores are shown.

Table 19. (PILOT): Descriptive statistics of the pilot discrete vocabulary test scores: Means, standard deviations (SD) and score ranges.

<table>
<thead>
<tr>
<th>Tests</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>MPS</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT 1</td>
<td>16</td>
<td>81</td>
<td>117</td>
<td>120</td>
<td>106.56</td>
<td>8.55</td>
</tr>
<tr>
<td>RVLT 2</td>
<td>16</td>
<td>89</td>
<td>120</td>
<td>120</td>
<td>109.94</td>
<td>8.31</td>
</tr>
<tr>
<td>PVLT 1</td>
<td>16</td>
<td>37</td>
<td>74</td>
<td>90</td>
<td>53.25</td>
<td>8.23</td>
</tr>
<tr>
<td>PVLT 2</td>
<td>16</td>
<td>38</td>
<td>69</td>
<td>90</td>
<td>54.94</td>
<td>7.20</td>
</tr>
<tr>
<td>PILOTDEPTH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>18</td>
<td>27</td>
<td>36</td>
<td>22.14</td>
<td>3.28</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>6</td>
<td>20</td>
<td>40</td>
<td>14.54</td>
<td>4.17</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>5</td>
<td>18</td>
<td>36</td>
<td>11.43</td>
<td>3.93</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>8</td>
<td>23</td>
<td>28</td>
<td>17.18</td>
<td>3.47</td>
</tr>
<tr>
<td>TOTAL</td>
<td>14</td>
<td>47</td>
<td>82</td>
<td>140</td>
<td>65.29</td>
<td>10.23</td>
</tr>
</tbody>
</table>

MPS = Maximum possible score.

The descriptive statistics provide a general profile of the pilot sample. It is worth noting that from Table 19 we can infer that the tests form a hierarchy in terms of degree of difficulty, both between the tests and between the components of PILOTDEPTH.

The mean score on the RVLT pre- and post-test is approx. 90% of the MPS (maximum possible score) making it the least difficult of the tests. The PVLT pre- and post-tests occupy a middle position with a mean score of approx. 60% of the MPS. PILOTDEPTH has a relatively low mean score, approx. 47% making it the most difficult test. In regard to the degree of difficulty of the components of PILOTDEPTH, COLL and POLY have the lowest degree of difficulty (62% and 61% of the MPS). The WD and SYN

\(^{14}\) The organization of the presentation of results has been adapted from Staehr Jensen (2005).
components have the highest degree of difficulty (36% and 32% of the MPS).

Turning now to the question of the shape of the distribution of the scores on the tests, two statistics that provide this information are the indices of skewness and kurtosis\(^\text{15}\). A rule of thumb for identifying a reasonably normal distribution are values for skewness and kurtosis somewhere between -2 and 2 (Bachman 2004). Skewness figures indicate the degree of symmetry of the distribution with a value of 0 indicating a perfectly symmetrical distribution, while negative and positive values indicate negative and positive skewness. The kurtosis statistics signify the degree of acuteness of the distribution with a value of 0 indicating a normal distribution, while the distribution will be flat for negative values and peaked for positive (ibid.).

All except two main tests and one sub-test have values within an acceptable range in terms of skewness and kurtosis. The distribution curve for the RVLT is highly peaked, which is probably a reflection of its relatively low degree of difficulty. Also, the first version of the PVLT shows a slightly peaked distribution curve.

4.5.1.2 Step 2: The relationship between size, depth of vocabulary knowledge and lexical richness

In the next step the results pertaining to the relationship among the measures of vocabulary knowledge and the LFP of the essays are outlined.

Table 20 displays the Pearson product-moment correlations between the scores on the tests.

It is apparent from Table 20 that very few of the variables correlate significantly with each other. There is a modest significant correlation between the two size test scores (RVLT and PVLT) and the proportion of B2000 words in the essays. Somewhat surprising the RVLT and PVLT scores only correlate modestly with each other. Moreover, no significant correlations are found between the PILOTDEPTH scores and the size test scores. The results also show that there is a significant negative correlation between the component WD scores and the B2000 scores. This is probably due to unclear test instructions for the WD section (cf. Table 18). A number of the informants misunderstood the instructions and supplied the word derivations of the target word in the collocation section instead of the test item.

\(^{15}\) The kurtosis and skewness figures for the pilot tests are not displayed in the table.

<table>
<thead>
<tr>
<th>Tests</th>
<th>RVLT</th>
<th>PVLT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PILOTDEPTH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>(16)</td>
<td></td>
</tr>
<tr>
<td>PILOT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>(14)</td>
<td></td>
</tr>
<tr>
<td>COLL</td>
<td>.40</td>
<td>-.21</td>
</tr>
<tr>
<td>N</td>
<td>(14)</td>
<td></td>
</tr>
<tr>
<td>WD</td>
<td>-.40</td>
<td>-.26</td>
</tr>
<tr>
<td>N</td>
<td>(14)</td>
<td>(14)</td>
</tr>
<tr>
<td>SYN</td>
<td>-.38</td>
<td>.03</td>
</tr>
<tr>
<td>N</td>
<td>(14)</td>
<td>(14)</td>
</tr>
<tr>
<td>POLY</td>
<td>.14</td>
<td>.04</td>
</tr>
<tr>
<td>N</td>
<td>(14)</td>
<td>(14)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>-.13</td>
<td>-.15</td>
</tr>
<tr>
<td>N</td>
<td>(14)</td>
<td>(14)</td>
</tr>
<tr>
<td>B2000</td>
<td>.63*</td>
<td>.62*</td>
</tr>
<tr>
<td>N</td>
<td>(13)</td>
<td>(13)</td>
</tr>
</tbody>
</table>

*significant at p<.05

4.5.1.3 Step 3: The development of vocabulary size

In step three the development of the informants’ receptive and productive vocabulary knowledge is examined. Since two different formats for measuring depth of vocabulary were trialed as part of developing a suitable depth test, it was not possible to examine the development of depth of vocabulary knowledge in the pilot study.

Table 21 shows the change in the total scores of the R- and PVLT over one term.

Table 21. (PILOT): Changes in RVLT and PVLT after one term (N = 16).

<table>
<thead>
<tr>
<th>Tests</th>
<th>Beg. of term</th>
<th>End of term</th>
<th>Change in mean score</th>
<th>Change in %</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT</td>
<td>106.56</td>
<td>109.94</td>
<td>3.38</td>
<td>3.17%</td>
<td>8.55</td>
</tr>
<tr>
<td>PVLT</td>
<td>53.25</td>
<td>54.94</td>
<td>1.69</td>
<td>3.17% (sic)</td>
<td>8.23</td>
</tr>
</tbody>
</table>

There is no statistically significant increase in the total mean score of the two size tests over one term. Since the degree of difficulty of the two sets of parallel versions was not counterbalanced in the pilot study, one cannot rule out the possibility that the lack of a significant increase is due to a difference in the degree of difficulty between the pre- and posttests. In other words, the lack of a difference in the mean scores between the beginning of term tests and end of term tests may simply be due to the pre-tests having a higher degree of difficulty than the post-tests. This possibility led me to introduce a counterbalanced procedure in the main study.
4.5.1.4 Step 4: The estimated overall and frequency band vocabulary size at the beginning and end of term one

In step four the development of the informants’ overall vocabulary size and proportion of words mastered at each frequency level is presented.

Figure 6 shows the subjects’ size of receptive vocabulary profile at the beginning of term and end of term.

*Figure 6. (PILOT):* The average receptive vocabulary size at the four levels (3K, AWL, 5K and 10K) at the beginning of term and end of term one (N = 16). White area = no. of words not known; Blue area = no. of words known.

The blue sections of the bars represent the number of words known at the specific frequency level. On average the informants seem to master the 5K level, already at the onset. Only 133 words are not mastered at this level. At the 10K level there is as expected a quite substantial gap.

Figure 7 shows the informants’ size of productive vocabulary knowledge profile at the beginning of term and end of term. At the productive end only the 2000 most frequent words seem to be sufficiently mastered (cf. section 4.2.1). Only about 64% (639/1000) percent of the vocabulary at the 3K frequency level is known. The proportion is even lower for academic vocabulary with only about 55% (311/570) known.
Figure 7. (PILOT): The average productive vocabulary size at the five levels (2K, 3K, AWL, 5K and 10K) at the beginning of term and end of term one (N = 16). White area = no. of words not known; Blue area = no. of words known.

As displayed in Figure 7 the largest gap can be found at the 10K level in which only 34% (1701/5000) of the vocabulary is known. The only level the students seem to master on average is the 2K level.

4.5.2. Discussion of the findings from the pilot study

There seems to be no improvement over one term in receptive and productive vocabulary knowledge. This lack of improvement may be due to flaws in the test design in that the two parallel versions employed may have different degrees of difficulty. However, the main study, which comprises a larger sample and in which data was collected also after one year of study, should give a clearer picture of the development of vocabulary knowledge not only over one term but also over one year of study.

Correlational analysis provided a somewhat confusing picture of how the different variables are related. The PILOTDEPTH scores did not correlate significantly with size of vocabulary knowledge test scores or lexical richness scores. This lack of concurrent validity can be a result of the depth test employed rather than due to a lack of relationship between size and depth of vocabulary knowledge in the sample. Correlational analysis between the RVLT and PVLT scores showed a modest correlation. This might be due to a ceiling effect in the RVLT which might decrease the strength of correlation. Laufer (1998) found a correlation of .89 between the RVLT and PVLT scores of EFL learners and a relatively lower strength of correlation for ESL learners. Accordingly, either the relationship between
receptive and productive vocabulary knowledge decreases with increasing proficiency or a ceiling effect sets in on this test for more advanced learners.

In the pilot study different versions of the RVLT and PVLT were administered at the pre- and post-tests. The results from the pilot study suggest that a counterbalanced design should be employed for the VLT tests in order to mitigate the risk of measurement error. As mentioned in section 4.2.3 this approach was initially employed for the depth test as well but was abandoned.

The correlation profile of the scores of the four components of depth seems to suggest that the test measures unrelated constructs. A number of sources of error were identified. First the collocation part was identified as defective and was therefore changed into a productive sub-section, hence ensuring that all components are tested productively. A second source of error identified was the instructions for the sub-sections. The instructions for the word derivation and synonym sub-sections were made clearer by supplying the test item for which the test taker should provide word derivations and synonyms of (cf. Tables 10 in section 4.2.3 and 18 in section 4.5). Another change made to make the sub-sections more uniform was to discard the part in the synonym sub-section asking the test takers to report the degree of similarity between the test item and the supplied synonym (cf. Tables 10 in section 4.2.3 and 18 in section 4.5). As shown in Table 19 in section 4.5.1.1, the depth test had a relatively high degree of difficulty. This together with fatigue on the part of the test takers might be a source of measurement error, which might explain the irregular correlation profile of the scores on the components sections. It was, thus, deemed necessary to condense the test by discarding the sub-section testing knowledge of polysemy.

4.6 Analysis procedure of the data from the main study

The data from the main study was analyzed in 4 main steps (i.e. steps 2-5 outlined below), each step examining different aspects of the variables of the study and addressing different research questions.

In the Theoretical framework the following 19 research questions (RQ) were stated expanding the five main research questions from 1.2, addressing relationships between different theoretical constructs of the study and the development of vocabulary knowledge and vocabulary use:

---

16 The organization of the presentation of results has been adapted from Staehr Jensen (2005).
1. What is the relationship between learners’ receptive and productive size?
2. How does the relationship between learners’ receptive and productive size change over time?
3. How are advanced adult EFL learners’ size and depth of vocabulary knowledge related?
4. How does the relationship between advanced EFL learners’ size and depth of vocabulary knowledge develop over time?
5. To what extent are different components of depth interrelated?
6. How does the interrelationship between components of depth change over time?
7. To what extent are vocabulary knowledge (receptive and productive size and productive depth) and lexical richness of take-home essays related?
8. How does the relationship between vocabulary knowledge (receptive and productive size and productive depth) and lexical richness change over time?
9. What frequency levels do advanced Swedish learners master receptively at different points in their university studies?
10. What frequency levels do advanced Swedish learners master productively at different points in their university studies?
11. How is size of vocabulary knowledge related to academic achievement?
12. How does the relationship between vocabulary knowledge and academic achievement change over time?
13. How is depth related to academic achievement English?
14. How does the relationship between depth and academic achievement change over time?
15. To what extent is lexical richness of take-home essays related to essay grade and overall academic achievement?
16. How does the relationship of lexical richness to essay grade and overall academic achievement change over time?
17. How does advanced university EFL learners’ size of vocabulary knowledge change over time?
18. How does learners’ productive depth of vocabulary knowledge develop over time?
19. How does the lexical richness of student essays develop over time?

These 19 RQs will be reordered according to topic and rephrased in order to obtain a basis for statistical inference and will be termed operationalized research questions (ORQs). Table 22 displays an overview of the variables of the study.
Table 22. *Overview of the variables of the study.*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptive vocabulary size</td>
<td>RVLT</td>
</tr>
<tr>
<td>Productive vocabulary size</td>
<td>PVLT</td>
</tr>
<tr>
<td>Depth of vocabulary knowledge</td>
<td>DEPTH</td>
</tr>
<tr>
<td>Knowledge of collocations</td>
<td>COLL</td>
</tr>
<tr>
<td>Knowledge of word derivations</td>
<td>WD</td>
</tr>
<tr>
<td>Knowledge of synonyms</td>
<td>SYN</td>
</tr>
<tr>
<td>Lexical richness</td>
<td>B2000</td>
</tr>
<tr>
<td>Coursegrade</td>
<td>CGRADE</td>
</tr>
<tr>
<td>Essaygrade</td>
<td>EGRADE</td>
</tr>
</tbody>
</table>

The presentation of analysis and results dealt with in the next section will proceed in five steps.

**Step 1** of the data analysis presents the descriptive statistics of the test results in order to provide an overview of the informants’ performance on the discrete vocabulary tests of the study. Moreover, it gives the Cronbach’s Alpha reliability coefficients estimating the internal consistency of the employed discrete-items tests. **Step 2** explores the relationship between the variables RVLT, PVLT, DEPTH, and B2000 and how it changes over the three data points. Since grades and lexical richness are predominantly end of term measures, these will only be looked at according to how they are related to the end of term scores on the three discrete tests. The following three ORQs are addressed:

1. To what extent do RVLT, PVLT, DEPTH and B2000 correlate with each other? (RQ: 1-4, 7-8)
2. To what extent do the components of DEPTH correlate with RVLT, PVLT and B2000? (RQ: 3, 4, 7, 8)
3. To what extent do the independent variables RVLT, PVLT, and DEPTH explain the variance in the dependent variable B2000? (RQ: 7, 8)

In **step 3** the interrelationship of the investigated components of depth of vocabulary knowledge is examined and addresses the following two ORQs:

4. To what extent do the components of DEPTH correlate with each other? (RQ: 5 and 6)
5. To what extent do the COLL, WD and SYN, respectively, explain the variance in WD+SYN, COLL+SYN and COLL+WD, respectively? (RQ: 5 and 6)
Step 4 of the data analysis addresses the relationship between PVLT, RVLT, DEPTH, B2000 and GRADE (both course grade and essay grade) and how it changes over the three data points. In step four of the data analysis, the following three ORQs will be dealt with:

6. To what extent do RVLT, PVLT and DEPTH correlate with GRADE? (RQ: 11, 12 and 13, 14)
7. To what extent does B2000 correlate with GRADE? (RQ: 15, 16)
8. How much emphasis do faculty teachers at the English department at Stockholm University report placing on vocabulary features when grading student essays in the first and second term of study? (RQ: 15)

In step 5 the development of RVLT, PPVLT, DEPTH and B2000 over time is examined and addresses the following five ORQs:

9. Is there a difference between the mean score of RVLT at the beginning of term one, end of term one and end of term two? (RQ: 17)
10. Is there a difference between the mean score of PVLT at the beginning of term one, end of term one and end of term two? (RQ: 17)
11. What is the estimated overall and frequency band vocabulary size at the beginning of term one, end of term one and end of term two? (RQ 9 and 10)
12. Is there a difference between the mean score of DEPTH at the beginning of term one, end of term one and end of term two? (RQ: 18)
13. Is there a difference in the mean-B2000 score of all essays between term one and two? (RQ 19)

The results of the five steps of the data analysis will be reported in the next chapter.
5 Results

In this chapter the results of the analyses conducted will be presented. The purpose of the empirical study is to explore on the one hand the relationship between size of vocabulary knowledge, depth of vocabulary knowledge, lexical richness of student essays and academic achievement as reflected in course grade, and on the other hand, the relationship between lexical richness and text quality as reflected in essay grade. Moreover, the aim is to track the development of these relationships, vocabulary knowledge and lexical richness over two terms of university studies (cf. section 1.2). As outlined in the previous chapter the data was analyzed in five steps. In step one the descriptive statistics are presented and in steps 2-5 the different research questions were addressed.

5.1 Step 1: Descriptive statistics of the discrete vocabulary test scores

The first step of the study provides an overview of the informants’ performances on the tests employed in order to shed light on the shape of the score curve and the degree of difficulty of the tests. Furthermore, the internal consistency of the individual tests and sub-tests measured by Cronbach’s Alpha is reported.

Table 23 shows the mean, standard deviation, score ranges and skewness and kurtosis values of the tests administered at the beginning of term one. All test scores are within the -2 and +2 skewness and kurtosis range, which indicates a reasonably normal distribution (Bachman 2004). As shown in Table 23 the mean scores of version 1 and 2 of the RVLT are fairly high (129.53 and 134.65) amounting to 86% and 90% of the maximum possible score. This seems to indicate that a large number of the participants had quite substantial receptive vocabularies. However, the SDs of 8.78 and 7.23 indicate a reasonable spread in the scores. In order to test whether the test scores on each of the 36 test and sub-test scores are normally distributed a Shapiro-Wilks Normality test was run. It showed that all test scores apart from the scores on the RVLT had a normal distribution. In regard to the RVLT test scores this might very well be due to a ceiling effect.
Table 23. Descriptive statistics of the main study discrete vocabulary test scores: Means, standard deviations (SD) and scores ranges of tests administered at the beginning of term one (both groups 06 and 07).

<table>
<thead>
<tr>
<th>Tests</th>
<th>N</th>
<th>Term</th>
<th>Min.</th>
<th>Max.</th>
<th>MPS</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT 1 v. 1</td>
<td>17</td>
<td>1</td>
<td>110</td>
<td>146</td>
<td>150</td>
<td>129.53</td>
<td>8.78</td>
<td>-0.39</td>
<td>0.55</td>
</tr>
<tr>
<td>RVLT 1 v. 2</td>
<td>17</td>
<td>1</td>
<td>121</td>
<td>146</td>
<td>150</td>
<td>134.65</td>
<td>7.23</td>
<td>-0.10</td>
<td>-0.87</td>
</tr>
<tr>
<td>PVLT 1 v. 1</td>
<td>17</td>
<td>1</td>
<td>46</td>
<td>74</td>
<td>90</td>
<td>54.82</td>
<td>8.63</td>
<td>0.89</td>
<td>0.03</td>
</tr>
<tr>
<td>PVLT 1 v. 2</td>
<td>17</td>
<td>1</td>
<td>38</td>
<td>75</td>
<td>90</td>
<td>86.00</td>
<td>9.87</td>
<td>-0.34</td>
<td>0.03</td>
</tr>
<tr>
<td>COLL</td>
<td>17</td>
<td>1</td>
<td>4.5</td>
<td>16</td>
<td>20</td>
<td>9.03</td>
<td>3.22</td>
<td>0.89</td>
<td>0.03</td>
</tr>
<tr>
<td>WD</td>
<td>16</td>
<td>1</td>
<td>4</td>
<td>24.5</td>
<td>40</td>
<td>11.25</td>
<td>5.71</td>
<td>0.79</td>
<td>0.16</td>
</tr>
<tr>
<td>SYN</td>
<td>17</td>
<td>1</td>
<td>4</td>
<td>15</td>
<td>40</td>
<td>8.68</td>
<td>2.57</td>
<td>0.26</td>
<td>1.84</td>
</tr>
<tr>
<td>TOTAL</td>
<td>16</td>
<td>1</td>
<td>16</td>
<td>47</td>
<td>100</td>
<td>28.59</td>
<td>9.47</td>
<td>0.60</td>
<td>0.63</td>
</tr>
<tr>
<td>COLL</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>20</td>
<td>7.41</td>
<td>3.36</td>
<td>-0.59</td>
<td>0.14</td>
</tr>
<tr>
<td>WD</td>
<td>16</td>
<td>1</td>
<td>3</td>
<td>12.5</td>
<td>35</td>
<td>8.88</td>
<td>2.73</td>
<td>-0.72</td>
<td>0.30</td>
</tr>
<tr>
<td>SYN</td>
<td>17</td>
<td>1</td>
<td>3</td>
<td>14.5</td>
<td>40</td>
<td>8.12</td>
<td>3.47</td>
<td>0.23</td>
<td>-0.92</td>
</tr>
<tr>
<td>TOTAL</td>
<td>16</td>
<td>1</td>
<td>13.5</td>
<td>36.5</td>
<td>95</td>
<td>24.91</td>
<td>6.41</td>
<td>-0.14</td>
<td>-0.51</td>
</tr>
</tbody>
</table>

MPS = Maximum possible score

The mean scores on both versions of the PVLT amount to about 60% of the maximum possible score. In addition, the two SDs of 8.63 and 9.87 suggest a good spread.

In contrast to the two size tests the mean score for the two versions of DEPTH amount to 29% and 26% of the maximum possible score, which reveals that these tests were relatively difficult. Moreover, the SDs of the two versions exhibit a quite large spread in the scores. If we examine the scores on the individual sub-tests, the mean scores on COLL amount to 45% and 37%, respectively, of the maximum possible scores, making it the easiest of the three sub-tests. The second sub-test in terms of degree of difficulty is WD, the mean scores of which amount to 28% and 25% of the maximum possible score. The SYN sub-test is the most difficult with mean scores of both the versions amounting to 22% and 20% of the MPS.

An independent t-test was carried out between the total test scores on the first and second group of informants’ receptive and productive tests. The difference between the means and the SD was non-significant, which suggests that the two versions of the receptive and the productive tests can be said to be equivalent forms.

5.1.1 Reliability of the tests

In this section the internal consistency of the two PVLT versions and the two DEPTH versions is estimated by using the method of Cronbach’s Alpha.

In Table 24 the alpha coefficients for all tests except the RVLT which is already fully standardized (cf. Schmitt et al. 2001) are presented.
Table 24. Cronbach’s Alpha reliability coefficients for PVLT and DEPTH administered at the beginning of term one.

<table>
<thead>
<tr>
<th>Tests</th>
<th>$N$</th>
<th>Cronbach’s Alpha coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVLT 1</td>
<td>17</td>
<td>.850</td>
</tr>
<tr>
<td>PVLT 2</td>
<td>17</td>
<td>.890</td>
</tr>
<tr>
<td>DEPTH 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL</td>
<td>17</td>
<td>.744</td>
</tr>
<tr>
<td>WD</td>
<td>16</td>
<td>.830</td>
</tr>
<tr>
<td>SYN</td>
<td>19</td>
<td>.500</td>
</tr>
<tr>
<td>TOTAL</td>
<td>16</td>
<td>.846</td>
</tr>
<tr>
<td>DEPTH 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL</td>
<td>17</td>
<td>.679</td>
</tr>
<tr>
<td>WD</td>
<td>16</td>
<td>.617</td>
</tr>
<tr>
<td>SYN</td>
<td>17</td>
<td>.536</td>
</tr>
<tr>
<td>TOTAL</td>
<td>16</td>
<td>.774</td>
</tr>
</tbody>
</table>

If one adopts the minimum acceptable reliability coefficient of 0.60 suggested by DeVellis (1991) all tests and sub-tests except SYN in both the depth test versions display acceptable reliability coefficients. One plausible reason for the low alpha coefficients for SYN might be that it has a skewed distribution due to a floor effect. In other words, a relatively high degree of difficulty might contribute to a skewed distribution which in turn might affect the degree of reliability. Another plausible reason is that this particular sub-test taps multiple different skills. As shown in Table 24 the sample size for the SYN in DEPTH 1 has been increased by two additional informants. This was done because the alpha coefficient obtained for a sample size of 17 test takers was very low. The larger the sample size, the more accurate or precise population alpha coefficient estimates are obtained (e.g. Charter 1999; Nunnally and Bernstein 1994). As noted by Charter (1999) with small sample sizes alpha coefficients can be unstable. Moreover, as displayed in Table 23 the standard deviation of the scores of SYN is relatively low. By increasing the sample, a higher standard deviation was obtained which generated a higher alpha coefficient.

5.2 Step 2 of the data analysis: The relationship between RVLT, PVLT, DEPTH and B2000

In this section the relationship between the main variables of the study (RVLT, PVLT, DEPTH and B2000) are explored by addressing the following ORQs:

---

17 These two informants were part of the main study but were excluded due to a test administration error. This error did not affect their test performance on DEPTH 1.
1. To what extent do RPVLT, PVLT, DEPTH and B2000 correlate with each other?
2. To what extent do the components of DEPTH correlate with RVLT, PVLT and B2000?
3. To what extent do the independent variables RVLT, PVLT, and DEPTH explain the variance in the dependent variable B2000?

Research questions 1-3 will be addressed according to the three data elicitation points of time, the beginning of term one, the end of term one and the end of term two. This is done in order to explore whether there is a difference in the correlational profile of the examined variables at different points in time. The approach for determining whether there is a difference in the strength of correlation\(^{18}\) between two same sets of variables is that the two coefficients must differ in terms of strength modality i.e. the strength levels proposed by Cohen and Holliday (1982; cf. section 4.1.3).

5.2.1 To what extent do RVLT, PVLT, DEPTH and B2000 correlate with each other?

5.2.1.1 The beginning of the first term of study

The first set of analyses examined the relationship between the variables, RVLT, PVLT and DEPTH at the beginning of the informants’ first term of study.

Pearson product-moment correlations between RVLT, PVLT, DEPTH were computed, the results of which are displayed in Table 25.

### Table 25. Beginning-of-term-one Pearson correlations among RVLT, PVLT, DEPTH.

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>RVLT</th>
<th>PVLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PVLT</td>
<td>34</td>
<td>.60**</td>
<td></td>
</tr>
<tr>
<td>DEPTH</td>
<td>32</td>
<td>.51**</td>
<td>.69**</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); ** significant at the 0.01 (two-tailed); \(P\)-value = .086

As shown in Table 25 all variables are statistically significantly correlated. There is a significant modest correlation of RVLT and PVLT to DEPTH.

\(^{18}\) The r-to-z-Fisher transformation formula can only be used for comparing the correlation coefficients of independent samples (Sheskin 2000). Since there are no standard computations for comparing the correlation coefficients from dependent samples in any of the standard statistics packages (e.g. SPSS, SAS) a non-statistical approach of comparing correlation coefficients is adopted in the present analysis.
The highest correlation in the data is that of PVLT to DEPTH, which is modest. The correlation between PVLT and RVLT is slightly lower and modest. The weakest correlation in the data is between RVLT and DEPTH.

In summary, a modest relationship was found between having a large vocabulary size receptively and productive depth of vocabulary knowledge of academic words. In addition, there is a modest relationship in the sample between receptive and productive vocabulary size. Accordingly, the three dimensions of vocabulary knowledge seem to be equally interrelated. This is a somewhat surprising result, since one would have expected a relatively stronger relationship between the learners’ receptive and productive size on the one hand and between their productive size and productive depth on the other than between their receptive size and depth which was measured productively.

In the next section the correlations between the variables RVLT, PVLT, DEPTH and B2000 at the end of the first term are presented.

5.2.1.2 The end of the first term of study

As a way of shedding light on whether the correlations between the investigated variables change longitudinally a correlation analysis was performed on the test scores of post-tests at the end of term one and two. The B2000ALL variable reflects the average proportion of B2000 words in all essays (i.e. not divided by discipline) handed in during the informants’ first term of study. The B2000LIT and –LING variables represent the proportion of beyond-2000 vocabulary in literature and linguistics essays, respectively.

Table 26 reports on the results of correlation between RVLT, PVLT, DEPTH and B2000ALL, -LIT and -LING at the end of term one among the sample.

Table 26. End-of-term-one Pearson correlations among RVLT, PVLT, DEPTH and B2000ALL, -LIT and -LING.

<table>
<thead>
<tr>
<th>Variables</th>
<th>B2000ALL</th>
<th>B2000LIT</th>
<th>B2000LING</th>
<th>RVLT</th>
<th>PVLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PVLT</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>.76**</td>
<td></td>
</tr>
<tr>
<td>DEPTH</td>
<td>.41*</td>
<td>.39*</td>
<td>.31†</td>
<td>.65**</td>
<td>.72**</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>34</td>
<td>31</td>
<td>34</td>
<td>34</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); **significant at p<0.01 (two-tailed), †P-value = .087

As shown in Table 26, overall there seem to be stronger significant correlations between the variables at the end of term one than at the beginning of the term. There is a high statistically significant correlation between RVLT and PVLT. Whereas the strength of correlation between
RVLT and DEPTH remained static, the strength of correlation between PVLT and DEPTH increased from modest to high. However, it should be noted that the difference between the beginning of term and the end of term correlation coefficients between the two variables is only .03 points. A more substantial development in the strength of correlation was found between RVLT and PVLT which increased from modest to high (i.e. .16 points). Moreover, Table 26 shows that there is a modest correlation between the independent variable DEPTH and the dependent variable B2000ALL. A slightly lower weak correlation holds between DEPTH and B2000 LIT. The two independent variables RVLT and PVLT did not display a significant correlation with the dependent variables B2000/ALL, -LIT and -LING.

In order to determine the predictive power of RVLT and PVLT on DEPTH a regression analysis was carried out. Employing the forced entry procedure, PVLT was entered first into the equation, since it showed the strongest correlation with DEPTH.

Table 27. End-of-term-one regression results with PVLT and RVLT as the predictor variables and DEPTH as the dependent variable (N = 34).

<table>
<thead>
<tr>
<th>Step</th>
<th>Procedure</th>
<th>Status $r^2$</th>
<th>$r^2$ change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>PVLT</td>
<td>.512*</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>RVLT</td>
<td>.537*</td>
<td>.025</td>
</tr>
</tbody>
</table>

PVLT explains about 51% ($F = 33.61, p< .001$) of the variance in DEPTH. The entry of RVLT at the second step of the equation did not yield a significant increase of the variance already explained by PVLT ($F$ change = 1.63, not significant at $p = .05$). These results indicate that productive size is a better index of productive depth than receptive size. There seems to be a stronger relationship between size and depth according to type of knowledge.

To sum up, a stronger relationship was found between the students’ receptive and productive vocabulary size operationalized as form/recall (cf. section 2.4.3.2.1) at the end of term than at the beginning of term one. Moreover, the results showed a modest relationship between depth of vocabulary knowledge and vocabulary use operationalized as the proportion of beyond-2000 vocabulary in all take-home essays. This is mainly due to the lexical richness of literary essays; the lexical richness of linguistics essays accounts for a minimal share of this relationship ($r = 0.2$). However, no relationship was found between size of vocabulary knowledge and lexical richness either in all essays or by essay type, i.e. literature and linguistics.

In regard to the relationship between size and productive depth, productive size was found to be a better predictor of productive depth than receptive size. This seems to suggest that with reference to productive size
and productive depth, learners make an appeal to similar types of knowledge.

Having presented the results of the correlations between the variables (RVLT, PVLT, DEPTH, B2000ALL, -LIT and -LING) at the end of term one, we will now look at the relationship between the variables at the end of the second term.

5.2.1.3 The end of the second term of study

In Table 28 the correlations between RVLT, PVLT, DEPTH and B2000 are displayed.

Table 28. End-of-term-two Pearson correlations among RVLT, PVLT, DEPTH and B2000ALL, -LIT and -LING.

<table>
<thead>
<tr>
<th>Variables</th>
<th>B2000ALL</th>
<th>B2000LIT</th>
<th>B2000LING</th>
<th>RVLT</th>
<th>PVLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>.75**</td>
<td></td>
</tr>
<tr>
<td>PVLT</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEPTH</td>
<td>.56¹</td>
<td>n.s</td>
<td>n.s</td>
<td>.65**</td>
<td>.80**</td>
</tr>
</tbody>
</table>

As can be seen from Table 28 there is no statistically significant correlation between the independent variable DEPTH and the dependent variables B2000ALL, -LIT and -LING.

The strength of correlation between RVLT and DEPTH remained stable between the end of the first and the second term. The weaker correlation between the RVLT and DEPTH than between PVLT and DEPTH may be due a ceiling effect in the RVLT scores, which might deflate the correlation of RVLT with DEPTH. However, a potential ceiling effect should also affect the correlations of RVLT with PVLT. Instead there is an increase in the strength of correlation between these two scores over time (i.e. between the beginning and end of term one). The strength of association between PVLT and DEPTH was found to be higher than at the beginning of term, but static over the end of term one and two.

Having examined the relationship between RVLT, PVLT, DEPTH and B2000 and how this relationship changes over the course of the informants’ English studies, in the next section we look at how RVLT, PVLT and B2000ALL, LIT, LING are related to each of the components of depth of vocabulary knowledge.
5.2.2 To what extent do the components of DEPTH correlate with RVLT, PVLT and B2000?

In this section we will examine the extent to which the three components of depth are related to vocabulary size and lexical richness and the extent to which there is a change over time in the relationships.

5.2.2.1 The beginning of the first term of study
Table 29 presents the correlations between the set of variable components, COLL, WD and SYN and the two size variables RVLT, PVLT at the beginning of term one.

Table 29. Beginning-of-term-one Pearson correlations among COLL, WD, SYN, RVLT and PVLT.

<table>
<thead>
<tr>
<th>Components</th>
<th>RVLT</th>
<th>PVLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLL</td>
<td>.67**</td>
<td></td>
</tr>
<tr>
<td>WD</td>
<td>.31^1</td>
<td>.48**</td>
</tr>
<tr>
<td>SYN</td>
<td>n.s</td>
<td>.30^2</td>
</tr>
<tr>
<td>N</td>
<td>34</td>
<td>34</td>
</tr>
</tbody>
</table>

**significant at p<0.01 (two-tailed); ^1P-value = .056; ^2P-value = .083

From Table 29 we can see that the strongest correlation between any two variables in the data is between PVLT and COLL, exhibiting a correlation coefficient of .78. The correlation between RVLT and COLL generates a lower but still significant coefficient of .67. Moreover, there is a modest correlation between PVLT and WD. The only component that does not show a statistically significant correlation with either RVLT or PVLT is SYN.

To summarize this analysis shows that knowledge of collocational preference of academic words is closely related to productive vocabulary size and moderately related to receptive vocabulary size. Whereas receptive vocabulary size is only related to one of the components comprising depth of vocabulary knowledge, i.e. knowledge of collocations, productive vocabulary size is associated with two components, namely knowledge of the collocational preference of academic words and the word derivations of academic words. No relationship was found between knowledge of synonyms of academic words and vocabulary size, receptive or productive.

5.2.2.2 The end of the first term of study
Table 30 reports on the results of the correlation between the independent variables COLL, WD and SYN and the dependent variables, RVLT, PVLT and B2000ALL, -LIT and -LING for the whole sample at the end of the first term of study.

<table>
<thead>
<tr>
<th>Components</th>
<th>RVLT</th>
<th>PVLT</th>
<th>B2000ALL</th>
<th>B2000LIT</th>
<th>B2000LING</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLL</td>
<td>.65**</td>
<td>.73**</td>
<td>.33†</td>
<td>.36*</td>
<td>n.s</td>
</tr>
<tr>
<td>WD</td>
<td>.49**</td>
<td>.63**</td>
<td>.40*</td>
<td>.42*</td>
<td>n.s</td>
</tr>
<tr>
<td>SYN</td>
<td>.44**</td>
<td>.42*</td>
<td>n.s</td>
<td>n.s</td>
<td>.37*</td>
</tr>
<tr>
<td>N</td>
<td>34</td>
<td>34</td>
<td>31</td>
<td>34</td>
<td>31</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); **significant at p<0.01 (two-tailed); †P-value = .075

It can be seen from the data in Table 30 that there is an overall increase in the number of statistically significant correlations compared to the beginning of term 1. In contrast to the beginning of term all components show a statistically significant correlation with both RVLT and PVLT, and COLL shows a stronger association with PVLT than with RVLT also at the end of term. All components except COLL show a similarly strong correlation with both RVLT and PVLT. In order to check whether the difference in the correlation between the scores on WD and the scores on the two size measures was due to the sub-tests of WD having different maximum possible scores at the two data points (at the beginning of term half of the sample were given the first version of the depth test in which the maximum possible score is 40 compared to the second version in which it is 35), the second version WD-sub-test scores was multiplied by 1.14 (40/35). This does not seem to be the case, however, since similar correlation coefficients were obtained with this recalculation.

At the end of term we can see a statistically significant correlation between WD and B2000 of ALLESS and LIT. While B2000LIT correlates with both COLL and WD, B2000LING correlates only with SYN.

If we look at how the lexical richness of the two essay types is related to the components of depth, we find that knowledge of synonyms is related to the proportion of advanced vocabulary used in linguistics essays, while knowledge of collocations and word derivations is related to this parameter in literature essays. Accordingly, the results seem to indicate that informants who have relatively more knowledge of synonyms of academic words tend to produce lexically richer linguistics essays, whereas those who perform relatively better on WD and COLL display a higher proportion of advanced vocabulary in their literature essays. In other words, the discipline in which the essays are written seem to be a factor behind the degree to which the components are predictive of the degree of lexical richness. As will be discussed in section 6.3.4, it might, thus, be the case that different components of depth play a role in vocabulary use depending on the writing task.
5.2.2.3 The end of the second term of study

In Table 31 we can see the correlations among the components, COLL, WD and SYN and RVLT, PVLT and B2000ALL, LIT and LING at the end of the second term of study.

Table 31. *End-of-term-two Pearson correlations among COLL, WD, SYN, RVLT, PVLT and B2000 ALL, LIT and LING among the two-terms-of-study informants.*

<table>
<thead>
<tr>
<th>Components</th>
<th>RVLT</th>
<th>PVLT</th>
<th>B2000ALL</th>
<th>B2000LIT</th>
<th>B2000LING</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLL</td>
<td>.72**</td>
<td>.78**</td>
<td>n.s</td>
<td>n.s</td>
<td>.56†</td>
</tr>
<tr>
<td>WD</td>
<td>.59*</td>
<td>.72**</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>SYN</td>
<td>n.s</td>
<td>.72*</td>
<td>.72*</td>
<td>.61*</td>
<td>n.s</td>
</tr>
<tr>
<td>N</td>
<td>16</td>
<td>16</td>
<td>11</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); **significant at p<0.01 (two-tailed)
† P-value = 0.074

Table 31 shows a similar overall pattern to other results. COLL correlates strongly with RVLT. A lower but still significant correlation was found between WD and RVLT. All three component variables correlate strongly with PVLT. A high correlation was found between SYN and B2000ALL, and SYN was found to correlate at a lower albeit significant level with B2000LIT. Moreover, a noteworthy correlation was found between COLL and B2000LING.

In summary, the findings indicate that over time more links are established between components of depth of vocabulary knowledge and productive size. Moreover, as suggested by the correlations of SYN with B2000ALL and B2000LIT, among the sub-sample consisting of students who continued to a second term of study a relationship between knowledge of synonyms and lexical richness emerges over time.

5.2.3 To what extent do the independent variables DEPTH, WD, and SYN explain the variance in the dependent variable B2000?

In Sections 5.2.1 and 5.2.2 a bivariate correlation analysis was conducted in order to establish the relationship between the variables RVLT, PVLT, DEPTH (and COLL, WD and SYN) and B2000ALL, LIT and LING. In order to determine the total amount of variance of the dependent variable B2000 explained by the variables that were found to correlate with B2000 a linear regression analysis was carried out. The analysis in this section will focus on the significant correlations between the independent variables and the dependent variable B2000 found for the whole sample.
5.2.5.1 The end of the first term of study

In Table 32 the results from a regression analysis in which DEPTH was entered into the model as the predictor variable and B2000ALL and B2000LIT as the dependent variables are presented.

Table 32. End-of-term-one regression results with DEPTH as the predictor variable and B2000ALL (N = 31) and -LIT (N = 34) as the dependent variables.

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>B2000ALL</th>
<th>B2000LIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH</td>
<td>.169*</td>
<td>149*</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed).

As displayed in Table 32 DEPTH accounts for about 17% (F = 5.89) and 15% (F = 5.58) of the variance in the B2000ALL and B2000LIT, respectively.

By way of summary, depth of vocabulary knowledge is able to predict the proportion of advanced vocabulary used in all essays, as well as in literature essays.

Table 33 provides the results of the regression analysis of the predictor variable WD and the dependent variables B2000/ALLES and B2000/LIT.

Table 33. End-of-term-one regression results with WD as the predictor variable and B2000ALL (N = 31) and B2000/LIT (N = 34) as the dependent variables.

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>B2000ALL</th>
<th>B2000LIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>WD</td>
<td>.161*</td>
<td>.176*</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed)

As can be seen in Table 33, WD alone accounts for 16% (F = 5.55) of the variance of the B2000 score of all essays, slightly less than that accounted for by the composite score reflected by DEPTH. However, WD alone accounts for a slightly higher proportion, about 18% (F = 6.85), of the variance of the B2000 score of literature essays than that explained by the composite score. Combining the B2000 scores of the literature and linguistics essays deflates the amount of variance which can be explained by WD. Accordingly, whereas DEPTH is a marginally better predictor of lexical richness of all essays than WD, WD alone accounts for a higher portion of the variance in B2000LIT.

In summary, knowledge of derivations of academic words contributes to the use of advanced vocabulary in literature essays and to a lesser degree all essays.

As was shown in Table 30 in section 5.2.2.2, SYN was found to correlate with B2000LING at .37, which is a weak correlation. Table 34 displays the results from a regression analysis in which SYN was entered as the predictor variable and B2000LING as the independent variable.
Table 34. End-of-term-one regression results with SYN as the predictor variable and B2000LING as the dependent variable (N = 31).

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>B2000LING</th>
<th>( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYN</td>
<td></td>
<td>.140*</td>
</tr>
</tbody>
</table>

*significant at p<0.05(two-tailed)

It can be seen from the data in Table 34 that SYN explains 14% (F = 4.73) of the variance of the B2000 score of linguistics essays written at the first term of study. Accordingly, knowledge of synonyms seems to contribute slightly to the use of advanced vocabulary in linguistics essays.

In summary the best predictor of the B2000 scores of all essays written in the first term of study seems to be DEPTH. If we look at the specific disciplines, WD is the best predictor of the B2000 scores of literature essays. As for linguistics the only statistically significant predictor of the B2000 scores is SYN. However, WD is the only component among the three sub-components that alone accounts for a significant portion of the variance in B2000ALL.

5.2.5.2 The end of the second term of study

In section 5.2.2.3, it was shown that SYN correlated with B2000ALL and LIT at .72 and .61, respectively. In order to determine the extent to which the independent variable SYN accounts for the variance in the B2000 score of all essays and the literature essays in the second term of study a linear regression analysis was carried out.

Table 35 provides the data of the regression analysis in which SYN was entered as the independent variable and the B2000 of ALLESS and -LIT as the dependent variables.

Table 35. End-of-term-two regression results with SYN as the predictor variable and B2000ALL (N = 11) and B2000LIT (N = 13) as the dependent variables.

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>B2000ALL</th>
<th>B2000LIT</th>
<th>( r^2 )</th>
<th>( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYN</td>
<td>.519*</td>
<td>.371*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*significant at p<0.05(two-tailed)

As shown in Table 35 SYN explains approx. 52% (F = 9.72) and 37% (F = 6.48) of the variance in the independent variables B2000ALL and B2000-LIT, respectively.

In summary the predictive value of the independent variables vis-à-vis the dependent variables seems to vary between the two terms and according to discipline. However, the inconsistent correlational profiles across the two terms might be a result of the smaller sample size in the second term. Overall, with reference to the predictive value of the components vis-à-vis
lexical richness in comparison to the predictive value of the composite measure of depth, the results suggest that individual components are as good or better predictors of lexical richness as the composite measure of depth.

5.3. Step 3: The relationship between the components of depth of vocabulary knowledge

Having examined the relationship among size, depth and lexical richness, in this section we look at the extent to which the components comprising depth of vocabulary knowledge conceptualized according to the components approach correlate with each other over three points of time. This analysis focuses mainly on the whole sample. The ORQs dealt with in this section are as follows:

4. To what extent do the components of DEPTH correlate with each other?
5. To what extent do the COLL, WD and SYN, respectively, explain the variance in WD+SYN, COLL+SYN and COLL+WD, respectively?

5.3.1 To what extent do the components of DEPTH correlate with each other?

The analysis of the interrelationship of the components of depth is presented according to the three data points.

5.3.1.1 The beginning of the first term of study

Table 36 presents the correlations between the components of DEPTH, COLL, WD and SYN.

Table 36. Beginning-of-term-one Pearson correlations among COLL, WD and SYN.

<table>
<thead>
<tr>
<th>Variables</th>
<th>DEPTH</th>
<th>COLL</th>
<th>WD</th>
</tr>
</thead>
<tbody>
<tr>
<td>WD</td>
<td>.56**</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>N</td>
<td>(32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYN</td>
<td>n.s</td>
<td></td>
<td>n.s</td>
</tr>
<tr>
<td>N</td>
<td>(34)</td>
<td></td>
<td>(32)</td>
</tr>
</tbody>
</table>

**significant at p<0.01 (two-tailed)

It is apparent from Table 36 that there is a relatively weak interrelationship between the components of DEPTH at the beginning of the first term. The only two components that correlate significantly with each other are COLL and WD which correlate at a modest level.
In view of the increase in the association between the components and the variables RVLT, PVLT and B2000, it is interesting to see whether a similar pattern can be observed in regard to the interrelationship of the sub-components.

In order to rule out the possibility that the correlation profile of the components reflects the fact that two different versions were used, a correlation analysis was carried out on a split sample. As described in section 4.4.1 two “parallel” versions of the depth test were developed and administered in a counterbalanced fashion. The sample was split according to which version the informants were administered. The results of this analysis revealed that the two versions displayed a similar correlational profile, which indicates that the correlational profile of the combined versions reflects the degree of interrelationship between sub-test scores on both versions viewed separately.

In summary, the data indicates a modest relationship between knowledge of collocations and knowledge of derivations. No relationship was found between knowledge of synonyms and the other two knowledge components. At the beginning of the students’ English language studies we see a relatively weak relationship between the components of depth of vocabulary knowledge which were investigated.

5.3.1.2 The end of the first term of study
Table 37 illustrates the correlations between the components, COLL, WD and SYN at the end of the first term of study.

Table 37. End-of-term-one Pearson correlations among COLL, WD and SYN (N = 34).

<table>
<thead>
<tr>
<th>Variables</th>
<th>DEPTH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COLL</td>
</tr>
<tr>
<td>DEPTH</td>
<td>.64**</td>
</tr>
<tr>
<td>WD</td>
<td>.44**</td>
</tr>
<tr>
<td>SYN</td>
<td></td>
</tr>
</tbody>
</table>

**significant at p<0.01 (two-tailed)

As Table 37 shows, there is a clear trend of an increasing number of associations between the components. As opposed to the beginning of term one in which only COLL and WD showed a statistically significant positive correlation, at the end of term one all three components show a statistically significant modest correlation with each other.

The results, thus, indicate that over one term of study the learners’ knowledge of the three components of depth of vocabulary knowledge becomes more integrated.
5.3.1.3 The end of the second term of study

Table 38 reports on the results of the correlation between the components of DEPTH, COLL, WD and SYN at the end of term two.

Table 38. End-of-term-two Pearson correlations among COLL, WD and SYN (N = 16).

<table>
<thead>
<tr>
<th>Variables</th>
<th>DEPTH</th>
<th>COLL</th>
<th>WD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH</td>
<td>WD</td>
<td>.76**</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SYN</td>
<td>.71**</td>
<td>.87**</td>
</tr>
</tbody>
</table>

** significant at p<0.05 (two-tailed)

As indicated in Table 38 there is an increase in the strength of correlation between the components, which together comprise DEPTH. The variables show a statistically significant high intercorrelation at the end of term 2. It is worth noting that the strength of the relationship between SYN and the other two components shows a steady increase from the beginning of term 1 to the end of term 2, whereas the association between COLL and WD increases in strength only between the end of term one and end of term two.

5.3.2 To what extent do COLL, WD and SYN, respectively, explain the variance in WD+SYN, COLL+SYN and COLL+WD, respectively?

In order to explore whether the individual components form a hierarchy in terms of the extent to which they are able to explain the total amount of variance of the score of a combination of the other two components, a regression analysis was carried out for the test scores obtained at the three different points in time. The following analyses focus on the whole sample.

5.3.2.1 The beginning of the first term of study

Table 39 displays the $r^2$ coefficients that hold between the independent variables in the first column and the respective dependent variable in columns 2-3 at the beginning of term one.

As displayed in Table 39 the only predictor variable that is not able to explain a statistically significant portion of the dependent variable is SYN. The component COLL explains approx. 22% ($F = 9.01$) of the variance in the dependent variable and WD accounts for slightly more of the variance in the dependent variable, namely approx. 34% ($F = 11.34$).
Table 39. Beginning-of-term-one regression results with COLL, WD and SYN as predictor variables and SYN + WD, COLL + SYN and WD + COLL as dependent variables (N = 32).

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>WD+SYN</th>
<th>COLL+SYN</th>
<th>COLL+WD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r²</td>
<td>r²</td>
<td>r²</td>
</tr>
<tr>
<td>COLL</td>
<td>.220*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WD</td>
<td>-</td>
<td>.339***</td>
<td>-</td>
</tr>
<tr>
<td>SYN</td>
<td>-</td>
<td>-</td>
<td>n.s</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); ***significant at p<0.001

Accordingly, the beginning-of-term data seem to indicate that knowledge of collocations or derivations is predictive of the degree of knowledge of the other two knowledge components. As for knowledge of synonyms, it was not found to be able to predict the scores of the other two components.

5.3.2.2 The end of the first term of study

Table 40 displays the amount of variance in the dependent variable explained by the individual components at the end of term one.

Table 40: End-of-term-one regression results with COLL, WD and SYN as predictor variables and SYN + WD, COLL + SYN and WD + COLL as dependent variables (N = 34).

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>WD+SYN</th>
<th>COLL+SYN</th>
<th>COLL+WD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r²</td>
<td>r²</td>
<td>r²</td>
</tr>
<tr>
<td>COLL</td>
<td>.314*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WD</td>
<td>-</td>
<td>.293*</td>
<td>-</td>
</tr>
<tr>
<td>SYN</td>
<td>-</td>
<td>-</td>
<td>n.s</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed)

As is shown in Table 40 at the end of term one two independent variables are able to explain a statistically significant proportion of the variance of their respective dependent variable. Both COLL and WD account for approx. 30% (F₁ = 13.72 and F₂ = 13.23) of the variance of their respective dependent variable. As was the case for the beginning of term scores, at the end of term SYN is not able to explain a significant portion of the variance in COLL+WD.

To sum up the data seem to indicate an implicational hierarchy in that knowledge of word derivations is the strongest predictor of knowledge of the other components. Knowledge of collocations has a slightly less predictive
power. Knowledge of synonyms does not seem to imply knowledge of the other two knowledge components.

5.3.2.3 The end of the second term of study

Table 41 shows the extent to which the individual components can explain a portion of the variance in their respective dependent variable at the end of term 2.

Table 41. *End-of-term-two regression results with COLL, WD and SYN as predictor variables and SYN + WD, COLL + SYN and WD + COLL as dependent variables at the end of term 2 (N = 16).*

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WD+SYN</td>
</tr>
<tr>
<td></td>
<td>$r^2$</td>
</tr>
<tr>
<td>COLL</td>
<td>.601**</td>
</tr>
<tr>
<td>WD</td>
<td>.775**</td>
</tr>
<tr>
<td>SYN</td>
<td>.691**</td>
</tr>
</tbody>
</table>

**significant at p<0.01(two-tailed)

All these $r^2$ values shown in Table 41 are much higher than those in term one. The highest proportion of the variance accounted for by any of the components is WD which explains as much as 78% ($F = 48.21$) of the variance in the combined score of COLL and SYN. SYN explains about 70% ($F = 31.37$) of the variance in the dependent variable COLL+WD. Among the three components COLL explains the lowest amount of variance in its independent variable, approx. 60% ($F = 21.04$).

If we look at the development of the respective components’ $r^2$ value in terms of the extent to which they are able to explain the variance in their respective independent variable, two main points are worth noting. First, we see that the amount of variance explained seems to increase over time. Second they do not seem to be equally good predictors of their respective independent variable score. On average over time the single best predictor of the performance on the other two components is WD.

In summary the results indicate an increase in the predictive value of the knowledge components over time. At the end of term two, knowledge of any one of the three knowledge components implies knowledge of any of the other two. Moreover, at the end of term the single best predictor of degree of knowledge of the other two knowledge components is knowledge of academic-word derivations. If we focus on the first term correlational profile which included a larger sample the results can be translated into an implicational hierarchy in which knowledge of collocations and word derivations imply knowledge of word derivations and synonyms and collocations and synonyms, respectively, while knowledge of synonyms does not seem to imply knowledge of the other two.
Having examined the degree to which the individual components can explain the variance in the other two, next will we look at the extent to which the independent variables DEPTH, WD and SYN can explain the variance in the scores for lexical richness.

5.4 Step 4: The relationship of RVLT, PVLT, DEPTH and B2000 to academic achievement and essay grade

It is of interest to see to what extent informants’ overall vocabulary knowledge and lexical richness in student essays are related to grades. The reason for this is twofold. First it is important to look at the extent to which lexical richness is a good measure of essay quality and overall proficiency in terms of the course grade. Second it is important to shed light on the extent to which vocabulary knowledge and the use of advanced vocabulary in essays are reflected in the grades. The following three operationalized research questions are addressed in this section:

6. To what extent do RPVLT, PVLT, and DEPTH correlate with GRADE?
7. To what extent does B2000 correlate with essay and course GRADE?
8. How much emphasis do faculty teachers at the English department at Stockholm University report placing on vocabulary features when grading student essays in the first and second term of study?

5.4.1 To what extent do RVLT, PVLT, and DEPTH correlate with GRADE?

Under this heading, the relationship between vocabulary knowledge, both size and depth and academic achievement as reflected by course grade (CGRADE) is reported. The relationship is examined in the light of end of term results, since course and essay grades are here treated as mainly end of term measures. As mentioned in section 4.1.3, since grades are measured on an ordinal scale, the non-parametric correlation test Spearman’s rank order was employed.

5.4.1.2 End of term one
In Table 42 the correlation coefficients between the independent variables and the dependent variables at the end of term one are presented.
Table 42. *End-of-term-one Spearman correlations between the independent variables RVLT, PVLT and DEPTH and the dependent variable CGRADE (N = 34).*

<table>
<thead>
<tr>
<th>Variables</th>
<th>CGRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT</td>
<td>.47**</td>
</tr>
<tr>
<td>PVLT</td>
<td>.49**</td>
</tr>
<tr>
<td>DEPTH</td>
<td></td>
</tr>
<tr>
<td>COLL</td>
<td>.50**</td>
</tr>
<tr>
<td>WD</td>
<td>.46**</td>
</tr>
<tr>
<td>SYN</td>
<td>.55**</td>
</tr>
<tr>
<td>TOTAL</td>
<td>.55**</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); **significant at p<0.01 (two-tailed); 

As can be seen in Table 42 all variables correlate moderately with CGRADE at the end of term one. These findings suggest that depth of vocabulary knowledge is an equally good index of academic achievement as vocabulary size.

If we look at how the individual levels of the two size tests correlate with CGRADE at the end of term one, we see in Table 43 that more productive levels than receptive levels are related to course grade.

Table 43. *End-of-term-one significant Spearman correlations between the levels and course grade (N = 34).*

<table>
<thead>
<tr>
<th>Variables</th>
<th>Course grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT: 5K</td>
<td>.35*</td>
</tr>
<tr>
<td>RVLT: 10K</td>
<td>.44**</td>
</tr>
<tr>
<td>PVL: 2K</td>
<td>.40*</td>
</tr>
<tr>
<td>PVL: 5K</td>
<td>.38*</td>
</tr>
<tr>
<td>PVL: 10K</td>
<td>.35*</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); **significant at p<0.01 (two-tailed) 

Moreover, productive knowledge of high frequency words seems to be related to course grade, while in the case of receptive size only knowledge of low frequency vocabulary is related to course grade. This might be due to a ceiling effect in the receptive high-frequency-vocabulary levels. Interestingly, size of academic vocabulary, neither receptive nor productive is found to be related to academic achievement.

5.4.1.3 End of term two

The results obtained from the correlation analysis at the end of term two are presented in Table 44.
Table 44. *End-of-term-two Spearman correlations between the independent variables RVLT, PVLT and DEPTH and the dependent variable CGRADE (N = 16).*

| Variables | CGRADE  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT</td>
<td>.42</td>
</tr>
<tr>
<td>PVLT</td>
<td>.58*</td>
</tr>
<tr>
<td>DEPTH</td>
<td></td>
</tr>
<tr>
<td>COLL</td>
<td>.39</td>
</tr>
<tr>
<td>WD</td>
<td>.57*</td>
</tr>
<tr>
<td>SYN</td>
<td>.54*</td>
</tr>
<tr>
<td>TOTAL</td>
<td>.52*</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed)

The only three independent variables that exhibit a significant positive correlation with the dependent variable, CGRADE, are PVLT, WD and SYN. Again it is worth pointing out that the sample size is probably too small to detect a statistically significant correlation when dealing with relatively low correlation coefficients.

As shown in Table 45 the only receptive size test levels which was found to correlate with course grade is the 5K level. In regard to the productive levels, by the second term a high correlation was found between the academic vocabulary level scores and course grade. The second-term grades seem to be more representative of productive size than receptive size.

Table 45. *End-of-term-two significant Spearman correlations between the levels and course grade (N = 16).*

| Variables | Course grade  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RVLT: 5K</td>
<td>.57**</td>
</tr>
<tr>
<td>PVLT: 2K</td>
<td>.52*</td>
</tr>
<tr>
<td>PVLT: AWL</td>
<td>.70**</td>
</tr>
<tr>
<td>PVLT: 5K</td>
<td>.67**</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed); **significant at p<0.01 (two-tailed)

On the whole across the two terms course grade seems to be representative of receptive knowledge of the 5K level words and productive knowledge of the 2K and 5K. Moreover, the grades seem to reflect productive size more than receptive size, at least in the second term in which a strong relationship was found between productive size of academic words and academic achievement.

In summary, based on the results for the end of the first term, vocabulary knowledge seems to be a fairly good predictor of academic achievement. There also seems to be a tendency towards grade being more representative of productive size than receptive size as the students’ progress in their university studies.

134
5.4.2 To what extent does B2000 correlate with GRADE?

In this section the relationship between lexical richness and grade will be examined. This relationship will be investigated both in terms of whether there is a relationship between lexical richness and essay grade (EGRADE) and whether there is a relationship between the lexical richness of all essays and course grade (CGRADE).

5.4.2.1 Term one

The B2000 scores of the individual essays CL1-LING (cf. section 4.4.2) were correlated with their respective essay grade. As for the average B2000 score of all essays, it is correlated with the dependent variable CGRADE.

Table 46. Term-one Spearman correlations between the independent variables B2000 CL1, -CL2, -CL3, -CL4, -LING and –ALL and the dependent variables EGRADE and CGRADE.

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>EGRADE/CGRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2000CL1</td>
<td>15</td>
<td>n.s</td>
</tr>
<tr>
<td>B2000CL2</td>
<td>14</td>
<td>n.s</td>
</tr>
<tr>
<td>B2000CL3</td>
<td>15</td>
<td>n.s</td>
</tr>
<tr>
<td>B2000CL4</td>
<td>15</td>
<td>n.s</td>
</tr>
<tr>
<td>B2000LING</td>
<td>28</td>
<td>n.s</td>
</tr>
<tr>
<td>B2000ALL</td>
<td>31</td>
<td>.38*</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed)

As displayed in Table 46 there are no significant correlations between the independent variables and essay grade, except in the B2000 score of all essays where there is a significant weak correlation with the dependent variable course grade. Accordingly, there seems to be a relationship, albeit weak, between the average B2000 score of all essays written in the first term and academic achievement as reflected by course grade.

In the next section the relationship between the variables in the second term of study will be examined.

5.4.2.2 Term two

Table 47 reports on the results of the correlation between the B2000 scores of essays written in the second term and essay or course grades.

Table 47 shows that there are no significant correlations between the independent and dependent variables. In contrast to the relationship between course grade and lexical richness found for the first term, there is no evidence of such a relationship in the second term. However, based on a sample size calculation, a sample size of 10 is too small to yield a statistically significant correlation coefficient of only .50.
Table 47. Term-two Spearman correlations between the independent variables B2000LIT1, -LIT2, -BESS and –ALL and the dependent variables EGRADE and CGRADE.

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>EGRADE/CGRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2000LIT1</td>
<td>13</td>
<td>.12</td>
</tr>
<tr>
<td>B2000LIT2</td>
<td>10</td>
<td>-.16</td>
</tr>
<tr>
<td>B2000BESS</td>
<td>11</td>
<td>.13</td>
</tr>
<tr>
<td>B2000ALL</td>
<td>10</td>
<td>.50</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed)

In summary the correlation analysis indicates that those informants who use relatively more advanced words in their writing are also those who tend to obtain a higher course grade. However, the general picture is that lexical richness is not a valid indicator of good quality writing with reference to the writing task of take-home essays written as part of the examination in university studies. In order to examine whether this lack of relationship can be attributed to the relative weight teacher-raters put on lexical richness, in the next section the results of a qualitative analysis of teacher-raters’ reported criteria in judging essays will be presented.

5.4.3 How much emphasis do teachers at the English department at Stockholm University report placing on vocabulary features when grading student essays in the first and second term of study?

The degree to which any language feature such as vocabulary or grammar is related to essay quality in terms of teacher ratings, depends mainly on two aspects: which features teachers focus on when assessing student writing and the extent to which the raters have adopted an agreed-upon objective standard for rating student texts. The first aspect has to do with the degree to which any one feature is focused on by the teachers in a consistent manner. In regard to the second aspect, a high degree of inconsistency in which language or content features are emphasized by the teachers will make it difficult to find a pattern in terms of the relationship between, in this case, the proportion of advanced vocabulary and grades (Leki 1995).

The following question was intended to elicit information on whether the teachers employ any written criteria when rating essays: “Do you follow any set of written criteria when rating essays?” The responses showed that only three of the ten respondents used any written criteria when rating essays. Although use of objective standards when rating essays does not guarantee
reliability in terms of intra- or inter-rater reliability, it does at least increase the degree of consistency in both these regards. Leki (1995: 24) writes:

That we share standards and expectations of ‘good writing’ is implicit in our teaching and assessment of writing. But the problem with these standards and expectations is that we cannot be certain if, or to what degree, our assumptions are shared by other constituents of our community.

Accordingly, one reason for there not being a significant correlation between the proportion of advanced vocabulary and essay grade might be due to a high degree of inconsistency in the way the teachers assess different language features in the essays.

As for how much weight the ten respondents put on content features in comparison to language features in their assessment of student essays, four of them state that they put equal weight on language and content features. Four of the respondents state that they put more weight on language features than content. Two of the respondents state that they focus more on content than language features when rating essays. In other words, a majority of the teachers put equal or more weight on language features than content. This seems to suggest that language proficiency plays an important role in how essays are graded. However, a majority of the respondents (7/10) stated that within language they put more weight on grammar than lexical features. In order to shed light on the degree to which the respondents view use of advanced vocabulary as a major vocabulary feature of a good essay, the respondents were asked to list the vocabulary features they considered as indicative of a good essay.

Figure 8 illustrates the lexical quality features that the respondents regard as important for a good essay in a multiple-choice question.

Figure 8. Answers concerning which of the five types of lexical features are focused on in the assessment of student essays.

The most frequent feature was “appropriate use of words” which is mainly concerned with depth of vocabulary knowledge, in that it is not sufficient for learners to have superficial knowledge of a word; they must also have
knowledge of a word’s range of meanings, its collocates and its register constraints in order for the use of a specific word to be considered appropriate by the rater. This specific aspect of vocabulary use is not reflected in the LFP analysis to a very great degree since only words that are clearly used incorrectly are discarded. Thus, words are retained that, although not clearly used incorrectly, might be assessed by the teachers as inappropriately used, and these words might then contribute to a negative assessment of the quality of the essay at hand. The second most frequently reported vocabulary feature, “variation”, is not reflected in the LFP analysis, since the LFP does not calculate lexical variation. The third feature, “advanced vocabulary”, on the other hand, is measured by the LFP.

Based on the responses to this question one can draw the conclusion that the two most frequently reported vocabulary features regarded as characterizing a good essay are not captured by the LFP analysis. This might be reflected in the degree to which the proportion of B2000 vocabulary is related to the teachers’ assessment and, by extension, to the grade given to a specific essay. As noted in section 5.3.2 no significant relationship of lexical richness to essay grade was found.

Let us now turn to three questions aimed at reflecting the degree to which the respondents relate grammar and vocabulary to text quality.

- Can you have a good essay with poor grammar?
- Can you have a good essay with poor vocabulary, e.g. low degree of lexical variation and high dependence on high frequency words?
- Can you have good vocabulary but a weak essay, e.g. a high degree of lexical variation and a low dependence on high frequency words?

Nine of the ten respondents answered the first question, and the majority (N = 6) answered no to it. Six of the teachers answered yes to the second question and eight of the teachers answered yes to the third question. Extrapolating from this, a majority of the respondents do not regard good vocabulary as a decisive factor in the overall quality of an essay.

The results of the survey indicate that in terms of overall quality correct grammar seems to be a more crucial factor than good vocabulary. How this is actually manifested in practice in the teachers’ ratings of essays is beyond the scope of this study. Suffice it to say that this general point of view among the teachers surveyed in the present study probably affects the degree to which a high proportion of advanced vocabulary in student essays is predicative of a higher grade, since lexical richness for the teachers does not seem to be an essential criterion in judging student essays.

Having examined the relationship in the previous section between lexical richness and grades and, in this section, teachers’ assessment practices, in the next section we will look at results pertaining to the development of the students’ vocabulary knowledge and use over the course of their studies.
5.5 Step 5: The development of vocabulary knowledge and lexical richness

In this section the development in receptive and productive size of vocabulary knowledge, depth of vocabulary knowledge and lexical richness over one and two terms of study at the university level will be examined. The ORQs (cf. 4.6) addressed are:

9. Is there a difference between the mean score on RVLT at the beginning of term one, end of term one and end of term two?
10. Is there a difference between the mean score on PVLT at the beginning of term one, end of term one and end of term two?
11. What is the estimated overall and frequency band vocabulary size at the beginning of term one, end of term one and end of term two?
12. Is there a difference between the mean score on DEPTH at the beginning of term one, end of term one and end of term two?
13. Is there a difference in the mean-B2000 score of all essays between term one and two?

5.5.1 Is there a difference between the mean score on RVLT at the beginning of term one, end of term one and end of term two?

In Table 48, we can see the change in the informants' receptive size of vocabulary knowledge over one term of study.

Table 48. Change in RVLT after one term (N = 34).

<table>
<thead>
<tr>
<th>RVLT</th>
<th>Mean</th>
<th>Mean-Change</th>
<th>Mean-Change in %</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning of term one</td>
<td>132.09</td>
<td>+1.73</td>
<td>1.31%</td>
<td>8.34</td>
</tr>
<tr>
<td>End of term one</td>
<td>133.82</td>
<td></td>
<td></td>
<td>9.53</td>
</tr>
</tbody>
</table>

As shown in Table 48, over a period of one term of study there is no statistically significant increase in the informants' receptive size of vocabulary knowledge.

Table 49 displays the mean difference of the RVLT score between the beginning of term one, end of term one and end of term two for the informants who proceeded to a second term of study.

Also for the sub-sample comprising the informants who proceeded to a second term, there is no statistically significant increase in the mean score of the RVLT over the first term of study. In regard to the development of receptive vocabulary size in term one, this sub-sample seems to show a similar pattern to the whole sample. However, there is a significant increase in the informants' receptive size of vocabulary knowledge of 5.06 points.
(1.19 + 3.88) 3.79%) over a period of two terms. This corresponds to an average increase of 337 basewords.

Table 49. Change in RVLT after one and two terms (N = 16).

<table>
<thead>
<tr>
<th>Receptive VLT</th>
<th>Mean</th>
<th>Mean-Change</th>
<th>Mean-Change in %</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning of term one</td>
<td>133.38</td>
<td>+1.19</td>
<td>+0.88</td>
<td>8.69</td>
</tr>
<tr>
<td>End of term one</td>
<td>134.56</td>
<td>+1.19</td>
<td>10.70</td>
<td></td>
</tr>
<tr>
<td>End of term two</td>
<td>138.44</td>
<td>+3.88**</td>
<td>+2.88</td>
<td>4.96</td>
</tr>
</tbody>
</table>

**significant at p<0.01 (two-tailed)

As displayed in Table 48 there is no statistically significant increase in the informants’ size of vocabulary knowledge after one term of study. An already quite large receptive size of vocabulary knowledge does not seem to increase over such a short period of time. However, after two terms a slight increase in the informants’ receptive vocabulary size can be observed.

The sample was split according to whether they scored above or below average. Among the below average group, although not statistically significant, a noteworthy increase of 3.22 points (p value = 0.052) over their first term of study can be observed. This seems to indicate that there is some development in the receptive vocabulary already at the students’ first term of study among the students with a below average receptive size. The sample consisting of two-terms student was also split along these lines and the analysis revealed that the below average students made substantially greater gains than the above average students over two terms of study. The below average students’ mean score increased by 6% percent, while among the above average students the increase was only 2.6% percent. Accordingly, there seems to be a relatively larger increase among those students with a below average onset size among the two-terms students.

Next the development in the participants’ productive size of vocabulary knowledge will be examined.

5.5.2 Is there a difference between the mean score on PVLT at the beginning of term one, end of term one and end of term two?

The results of a paired-samples t-test analysis, in which the PVLT scores at the beginning and end of term one are compared, are displayed in table 50.

Table 50. Change in PVLT after one term (N = 34).

<table>
<thead>
<tr>
<th>Productive VLT</th>
<th>Mean</th>
<th>Mean-Change</th>
<th>Mean-Change in %</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning of term one</td>
<td>55.41</td>
<td>2.47**</td>
<td>4.46</td>
<td>9.15</td>
</tr>
<tr>
<td>End of term one</td>
<td>57.88</td>
<td></td>
<td></td>
<td>8.63</td>
</tr>
</tbody>
</table>

**significant at p<0.01 (two-tailed)
It can be seen from Table 50 that there is a statistically significant increase of almost 5% in PVLVT scores after only one term of study. In order to shed light on whether students with a below average score improved more than the students with an above average score, the sample was split into two sub-samples, those who had a below average score and those with an above average score. A significant increase was only found among the informants with a below average score. On average their score increased with about 10% over one term. Table 51 presents the results for the sub-sample consisting of the informants who proceeded to a second term.

Table 51. Change in PVLVT after one and two terms (N = 16).

<table>
<thead>
<tr>
<th>Productive VLT</th>
<th>Mean</th>
<th>Mean-Change</th>
<th>Mean-Change in %</th>
<th>Std. d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning of term one</td>
<td>55.50</td>
<td>3.56*</td>
<td>6.41</td>
<td>10.28</td>
</tr>
<tr>
<td>End of term one</td>
<td>59.06</td>
<td></td>
<td></td>
<td>8.51</td>
</tr>
<tr>
<td>End of term two</td>
<td>62.56</td>
<td>3.50***</td>
<td>5.93</td>
<td>10.33</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed), ***significant at p<0.001 (two-tailed)

We can see in Table 51 that there is a slightly larger increase in the PVLVT score over one term of study among the sub-sample comprising informants who proceeded to a second term. Accordingly, there is a risk that the sub-sample is not representative of the whole sample in terms of the development of productive vocabulary knowledge. There is a linear increase over two terms of study. The increase is as much as approx. 13% over two terms of study.

Having looked at the development of the size dimension of vocabulary knowledge, next we will examine how the vocabulary changes in terms of the estimated overall vocabulary size and which frequency bands reflect this increase.

5.5.3. What is the estimated overall and frequency band vocabulary size at the beginning of term one, end of term one and end of term two?

In this section the informants’ size of vocabulary knowledge profile will be examined. Figure 9 presents the average receptive vocabulary size at the five frequency levels for the sample at the beginning of term one, end of term one and end of term two. The blue area in the columns represents the number of words known, while the white area represents the number of words not known. As mentioned in section 4.2.1.1 the criterion of mastery is set at 26 correct items out of the 30 possible per level for the RVLT and for the PVLVT it is set at 16 out of 18.
The mean receptive size of vocabulary knowledge at the beginning of term is 7769 word families (2K+3K+5K+10K). Overall there do not seem to be any significant gaps in the 2K, 3K and 5K levels. On average the informants master 90% of the words at the 5K level. At the 10K level there is a sharp drop in that a considerable proportion of the words are not known, i.e. as many as approx. 40% (1951/5000).

In terms of the proportion of the word families known at each level the AWL should be placed somewhere between the 3K and 5K level. The informants know on average 96.30% of the words at the 3K level, whereas slightly fewer are known at the AWL level (94.74%). This seems to suggest that they are relatively familiar with English academic vocabulary at the onset of their studies.

The overall vocabulary size for the sub-sample who studied for two terms is 7850 word families at the beginning of term. At the end of term two it is 8537 word families, hence a difference of almost 700 word families. As much as 91% (625/687) of this increase is at the 10K level. As illustrated by the bars in Figure 9, due to an already high vocabulary size at the below-10K levels, very small gains can be observed at these levels. Accordingly, the relatively high increase in the 10K level is attributable to the fact that it is the only level at which there is room for improvement. It should be pointed out that the two-term-of-studies group are representative of the whole sample.

---

Figure 9. The average receptive vocabulary size at the five levels (2K, 3K, AWL, 5K and 10K) at the beginning of term one (N = 34), the end of term one (N = 34) and the end of term two (N = 16). White area = no. of words not known; Blue area = no. of words known.

Among the two-terms students only the end-of-term-two size estimates are shown in the figure for simplicity.
since no statistically significant differences were found between them and
the one-term-of-study groups on any of the tests and sub-tests at the two data
points in the first term.

Table 52, displays the proportion of students who master the individual
levels of the RVLT at different points in time.

Table 52. Proportion of students reaching the cut off point for mastery of the RVLT
levels at the beginning of term one, the end of term one and the end of term two.

<table>
<thead>
<tr>
<th>Level</th>
<th>Proportion of students reaching the cut-off point for mastery at the beginning of term one</th>
<th>Proportion of students reaching the cut-off point for mastery at the end of term one</th>
<th>Proportion of students reaching the cut-off point for mastery at the end of term two</th>
</tr>
</thead>
<tbody>
<tr>
<td>2K</td>
<td>100% (34/34)</td>
<td>100% (34/34)</td>
<td>100% (16/16)</td>
</tr>
<tr>
<td>3K</td>
<td>97% (33/34)</td>
<td>100% (34/34)</td>
<td>100% (16/16)</td>
</tr>
<tr>
<td>AWL</td>
<td>91% (31/34)</td>
<td>97% (33/34)</td>
<td>100% (16/16)</td>
</tr>
<tr>
<td>5K</td>
<td>71% (24/34)</td>
<td>79% (27/34)</td>
<td>94% (15/16)</td>
</tr>
<tr>
<td>10K</td>
<td>12% (4/34)</td>
<td>6% (2/34)</td>
<td>25% (4/16)</td>
</tr>
</tbody>
</table>

It is apparent from Table 52 that a majority of the students failed to reach the
cut-off point for mastering the 10K level. The level-mastery profile seems to
be stable among a majority of the students over the two terms of studies
examined.

Figure 10 shows the students’ average productive vocabulary sizes at the
different levels at the beginning of term one, end of term one and end of term
two.

Figure 10. The average productive vocabulary size at the five levels (2K, 3K, AWL,
5K and 10K) at the beginning of term one (N = 34), the end of term one (N = 34)
and the end of term two (N = 16). White area = no. of words not known; Blue area =
o. of words known.
The average size of the informants’ productive vocabulary knowledge at the beginning of term is 4893 word families, at the end of the first term 5161 and at the end of the second term 5671. There are gaps in the basic vocabulary represented by the 2K and 3K levels. At the 2K and 3K levels 11.5% and 23.6% of the vocabulary, respectively, is not known productively.

The changes at the 2K-3K levels are not significant. Despite certain gaps in the 2K and 3K levels no significant increase takes place in these levels. At the 5K and 10K levels there is a significant average increase of 88 and 188 base words, respectively. Accordingly, after one term, the largest increase in productive vocabulary size takes place among the lower-frequency levels.

Among the students who studied English for two terms the average productive vocabulary size at the beginning of term is 4974 word families and 5669 word families at the end of the second term. There is an average increase of 696 word families in the informants’ productive vocabulary after two terms. In comparison to the learning profile of the informants’ receptive vocabulary size, their productive vocabulary size learning profile is more evenly spread out among the five frequency levels. However, they are not filling in gaps in their knowledge of 2K vocabulary. The increase of productive vocabulary is limited to the lower frequency levels with the 5K and 10K levels accounting for about 83% ((264+313)/696) of the overall increase. In contrast to the informants’ receptive vocabulary size in which there was only room for gains in low-frequency vocabulary, their productive vocabulary size had relatively large gaps in all levels. In view of this, it is somewhat surprising that they were not filling in the gaps in their knowledge of more frequent words, specifically the 3K level. Instead the gains were limited to lower-frequency words represented by the 5K and 10K levels.

Table 53 presents the overall level-mastery profile at the three data points.

<table>
<thead>
<tr>
<th>Level</th>
<th>Proportion of students reaching the cut-off point for mastery at the beginning of term one</th>
<th>Proportion of students reaching the cut-off point for mastery at the end of term one</th>
<th>Proportion of students reaching the cut-off point for mastery at the end of term two</th>
</tr>
</thead>
<tbody>
<tr>
<td>2K</td>
<td>76% (26/34)</td>
<td>74% (25/34)</td>
<td>94% (15/16)</td>
</tr>
<tr>
<td>3K</td>
<td>18% (6/34)</td>
<td>32% (11/34)</td>
<td>56% (9/16)</td>
</tr>
<tr>
<td>AWL</td>
<td>21% (7/34)</td>
<td>21% (7/34)</td>
<td>25% (4/16)</td>
</tr>
<tr>
<td>5K</td>
<td>3% (1/34)</td>
<td>0</td>
<td>19% (3/16)</td>
</tr>
<tr>
<td>10K</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Among the two-terms students only the end-of-term-two size estimates are shown in the figure for simplicity.
There is a change in the level-mastery profile between the first and second term. Compared with the level-mastery profile of the first term, a majority of the students reach the cut-off point for mastering the 3K level at the end of term two. In contrast to receptive size, two vocabulary size milestones can be identified for productive size, namely the 2K level at the beginning of the students’ university studies and the 3K level at the end of term two.

In summary, the informants’ size of receptive vocabulary knowledge at the onset of their studies is relatively large and there were no gaps in the basic vocabulary represented by the 2K and 3K levels. Only after two terms was there a significant increase. Due to the informants’ already high familiarity with high-frequency and academic vocabulary, the gains were limited to low-frequency words.

With respect to their productive vocabulary knowledge, there was a significant increase after one and two terms mainly in low-frequency vocabulary represented by the 5K and 10K levels. The informants as a group did not seem to fill the gaps in their productive knowledge of high frequency- and academic vocabulary, even after two terms.

Having looked at the development of the size dimension of vocabulary knowledge, next we will examine whether an increase can be identified in the participants’ depth of vocabulary knowledge.

5.5.4 Is there a difference between the mean score on DEPTH at the beginning of term one, end of term one and end of term two?

The first term development of depth was examined only among the sub-sample who were given the same version at the beginning and end of term.

Table 54 shows the mean change in the score for the depth test and its sub-sections for the version-2 group who were administered the same test both at the beginning and end of term one.

Table 54. Change in DEPTH after one term among the informants who were administered the second version at the beginning of term one.

<table>
<thead>
<tr>
<th>DEPTH v. 2</th>
<th>N</th>
<th>Beginning-of-term-one mean</th>
<th>End-of-term-one mean</th>
<th>Mean-Change</th>
<th>Mean-Change in %</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH</td>
<td>16</td>
<td>21.10</td>
<td>24.68</td>
<td>+3.58*</td>
<td>16.97</td>
<td>6.3</td>
</tr>
<tr>
<td>COL</td>
<td>17</td>
<td>7.41</td>
<td>7.82</td>
<td>+0.41</td>
<td>5.53</td>
<td>3.6</td>
</tr>
<tr>
<td>WD</td>
<td>16</td>
<td>5.07</td>
<td>5.43</td>
<td>+0.36</td>
<td>7.10</td>
<td>1.5</td>
</tr>
<tr>
<td>SYM</td>
<td>17</td>
<td>8.12</td>
<td>10.82</td>
<td>+2.70*</td>
<td>33.25</td>
<td>3.4</td>
</tr>
<tr>
<td>TOTAL</td>
<td>16</td>
<td>21.10</td>
<td>24.68</td>
<td>+3.58*</td>
<td>16.97</td>
<td>6.3</td>
</tr>
</tbody>
</table>

*significant at p<0.05(two-tailed)
As can be seen from Table 54, there is a significant increase in the comprehensive score of 16.97% (3.58) over one term of study. The only sub-component that seems to develop over one term and which accounts for the largest part of the increase in DEPTH is SYN which increases by 33.25% (2.70). The sample was split according to whether they obtained a below average score. Those with a below average total score made a significant increase of 28% and those with a below average on the synonyms sub-test made a significant increase of 70%. For the two other sub-tests no significant increase was observed among the below-average groups.

Accordingly, the more demanding dimensions of vocabulary knowledge such as productive size of vocabulary knowledge and depth of vocabulary knowledge seem to develop already after only one term of study.

Table 55 shows the mean increase of DEPTH scores between beginning of term one and end of term two.

Table 55. *Change in DEPTH between the beginning of term one and end of term two (N = 16).*

<table>
<thead>
<tr>
<th>DEPTH v.1+v. 2</th>
<th>Beginning-of-term-one mean</th>
<th>End-of-term-two mean</th>
<th>Mean-change between beg. of term one and end of term two</th>
<th>Mean-change between beg. of term 1 and end of term 2 in %</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH TOTAL</td>
<td>23.68</td>
<td>28.28</td>
<td>4.59**</td>
<td>19.38</td>
<td>7.53</td>
</tr>
<tr>
<td>COLL</td>
<td>9.19</td>
<td>10.97</td>
<td>+1.78**</td>
<td>19.36</td>
<td>4.33</td>
</tr>
<tr>
<td>WD</td>
<td>6.03</td>
<td>5.90</td>
<td>-0.13</td>
<td>-2.16</td>
<td>1.40</td>
</tr>
<tr>
<td>SYN</td>
<td>8.47</td>
<td>11.41</td>
<td>+2.94*</td>
<td>34.71</td>
<td>3.58</td>
</tr>
</tbody>
</table>

*significant at p<0.05(two-tailed); **significant at p<0.01(two-tailed)

As shown in Table 55 there are significant changes in both COLL and SYN scores among the sub-sample who proceeded to second term of study over two terms.

In order to determine whether the changes in the scores reflect an increase among both test-version groups, a separate paired-samples t-test was conducted for the two groups. A similar pattern was observed for both subsamples, i.e. an increase in COLL and SYN scores, but no increase in the WD scores. Accordingly a similar developmental pattern was observed among each of the test-version sub-samples as for the whole sample.

In summary, the largest increase in percentage figures seems to be in the informants’ depth of vocabulary knowledge rather than in their vocabulary size, both receptive and productive. If we look at the development of the individual components comprising depth of vocabulary knowledge, there is a clear pattern of an increase in COLL and SYN over two terms of study. The
only component that develops both after one and two terms is SYN. Moreover, there is no significant increase in WD between the three points of time examined.

Having looked at the development of size and depth of vocabulary knowledge we will examine whether there is a difference in lexical richness between essays produced in the students’ first and second term.

5.5.5 Is there a difference in the mean-B2000 score of all essays between term one and two?

As has been reported in the previous section, there is a development of the learners’ vocabulary knowledge over one and two terms of study. Accordingly, the question is if the use of advanced vocabulary in student essays exhibits a similar pattern. Table 56 compares the B2000 scores of first and second term essays. An average score was calculated, which includes all the essays produced in the respective terms.

Table 56. Change in B2000ALL after two terms (N = 13).

<table>
<thead>
<tr>
<th>Time</th>
<th>Means</th>
<th>Change B2000ALL over two terms</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term 1</td>
<td>17.84%</td>
<td></td>
<td>3.97</td>
</tr>
<tr>
<td>Term 2</td>
<td>21.03%</td>
<td>3.19%*</td>
<td>4.71</td>
</tr>
</tbody>
</table>

*significant at p<0.05 (two-tailed)

As can be seen from Table 56 there is a significant difference in the proportion of B2000 words between first and second term essays. Whether this is attributable to an actual development of their vocabulary knowledge or whether the higher requirements in the second term have forced them to produce more advanced vocabulary that were already known is difficult to ascertain. However, since we have seen an increase in the informants’ size and depth of vocabulary knowledge, it would not be too far-fetched to conclude from the data that some of that knowledge has become incorporated into their essay writing.

As was shown for the significant increase on the other measures, the below-average students produced the highest growth rate also in regard to the lexical richness score. The below-average informants (N = 6) increased their lexical richness scores by 3.85 percentage points, while no significant increase was found among the above-average students. However, this should be interpreted with some caution due to the small sample size.

In the next chapter the empirical findings of the study will be discussed.
6. Summary and discussion of results

The primary purpose of this chapter is to evaluate and interpret the results in relation to previous research, and to address theoretical and pedagogical implications. The chapter is divided into two main parts. The first part provides a short summary of the findings.

The second part discusses the findings of the current study. It is further divided into three sections each of which deals with a specific area. The first section begins by discussing the development of the individual facets of vocabulary knowledge and use. The next addresses the relationship between the variables. The final section of the chapter evaluates and interprets results pertaining to the development of the relationship between the variables examined in the present study. Each section begins by relating the findings to previous research and goes on to discuss possible underlying causes of the findings. The final part of each section addresses the theoretical and pedagogical implications.

6.1 Summary

Five main research questions were addressed in the present study (cf. section 1.2). Questions 1-4 concern the relationship between the variables investigated. The interrelationship between facets of learners’ vocabulary knowledge and use and academic achievement was examined both from a static and a longitudinal perspective. Question 5 addressed the development of learners’ vocabulary knowledge and use longitudinally. Each research question will form a separate heading under which relevant findings will be summarized.

6.1.1. To what extent are Swedish students’ receptive and productive vocabulary size and depth of vocabulary knowledge related?

A close relationship was found between learners’ receptive and productive size. The relationship between receptive and productive size increased in strength towards the end of the students’ first term of study. The degree to which the students’ scores on the receptive and productive size test were
found to correlate remained stable between the end of their first term and the end of their second term.

The relationship between receptive size and productive depth (see below) was found to be moderate over the three data points (i.e. beginning of term one, end of term one and end of term two). The relationship between productive size and productive depth was found initially to be moderate but increased in strength to high at the end of term two. In regard to the relationship between the students’ size and depth overall the pattern discerned was that of a stronger relationship between the same kinds of vocabulary knowledge, i.e. productive size and productive depth than between different kinds of vocabulary knowledge, i.e. receptive size and productive depth.

The relationship between size and depth was also examined from the point of view of the extent to which receptive and productive size were related to three components conceptualized as elements of depth. Overall among the three components of depth, knowledge of collocations was found to exhibit the strongest relationship with receptive and productive size across the three data points. Accordingly, there seems to be a stronger link between syntagmatic associational knowledge and size than between size and either paradigmatic associational word knowledge or morphological word knowledge. There was a strong relationship productive size to knowledge of collocations and across the three data points. The relationship of knowledge of collocations to receptive size was modest in the first term of study and high at the end of the second term of study. This may be explained by register and meta-knowledge increasing with time at a language department, where students practice translation and linguistics and become familiarized with different text types. They come into university as linguistic novices and already after one term have a good overview of semantics and morphology.

Of the two other knowledge components, i.e. knowledge of derivations and synonyms, only the former was found to exhibit a relationship with size, namely productive size at beginning of term one. At the end of term both knowledge of word derivations and synonyms were moderately related to vocabulary size, both receptive and productive. At the end of the second term of study the students’ knowledge of word derivations and synonyms displayed a strong relationship with productive size. Moreover, there was a moderate relationship between knowledge of word derivations and receptive size. Overall, the findings indicate that the different components are not equally related with size. The increase in the strength to which size was found to be related to depth from the beginning of term to the end of term two is to a large extent accounted for by an increase in the strength to which size was related to knowledge of word derivations and synonyms.

In addition to the examination of the relationship between size and depth, the relationship between the three vocabulary-depth knowledge components was investigated. At the beginning of term the only significant relationship
between the components found was between knowledge of collocations and word derivations. At the end of terms one and two there was a modest and a strong relationship among the three components, respectively. Overall, the components coalesced over time towards a stronger interrelationship. In regard to the extent to which the individual components were able to predict the performance on the other two, it corresponded to the developmental pattern of the interrelationship between the components. At the beginning and end of term one, knowledge of collocations and word derivations each exhibited some predictive power of the performance on the other two knowledge components. At the end of term two all three components could each explain a large portion of the variance in the other two components.

6.1.2 To what extent is Swedish students’ vocabulary knowledge related to lexical richness in academic essays?

At the end of term one a modest relationship was found between depth of vocabulary knowledge and lexical richness of all essays (i.e. not divided by discipline, i.e. literature or linguistics). This overall relationship is primarily a reflection of a modest relationship between knowledge of word derivations and the lexical richness of literature essays. In addition knowledge of collocations and synonyms were found to be weakly related with the lexical richness of literature essays and linguistics essays, respectively. At the end of term two a strong relationship was found between knowledge of synonyms and the lexical richness of all essays. The largest contributor to this relationship was the modest relationship between knowledge of synonyms and the lexical richness of literature essays. This may be explained by literature-essay writers being more encouraged to vary their language. This is not an issue in linguistics, since the task demands place more emphasis on content than style. Overall, the findings indicate a development towards a stronger link between depth of vocabulary knowledge and at the same time a variation among the depth components in the extent to which they are related to lexical richness. Mainly two components of depth, morphological and paradigmatic knowledge i.e. word-derivation and synonymy knowledge, exhibit a relationship with lexical richness. Overall, the pattern discerned is a strengthening of the relationship between depth and lexical richness\(^\text{21}\), and variation in the extent to which depth was related to lexical richness according to type of component and discipline. No relationship of receptive and productive size to lexical richness was found. The tendency to use a wider vocabulary is associated with the quality rather than the quantity of one’s vocabulary knowledge.

\(^{21}\text{This seems to be the case at least between the beginning and end of term one. The relatively small sample size at the end of term might be a reason why no relationship was found between the aggregate score of depth and lexical richness.}\)
6.1.3 To what extent is vocabulary knowledge related to academic achievement?

All discrete vocabulary test scores correlated at a modest level with first-term course grade.

Table 57 illustrates how the different levels of the RVLT and PVLT are related with course grade in the first and second term of study.

Table 57. Levels of the RVLT and PVLT exhibiting a significant correlation with course grade at different points in time.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Relationship of the scores on the Receptive Levels to first- and second-term course grade</th>
<th>Relationship of the scores on the Productive Levels to first- and second-term course grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First term</td>
<td>Second term</td>
</tr>
<tr>
<td>2K</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>3K</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AWL</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5K</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td>10K</td>
<td>++</td>
<td>-</td>
</tr>
</tbody>
</table>

- = Non-significant; + = Weak correlation; ++ = Modest correlation; +++ = High correlation

As illustrated in Table 57, two main patterns can be identified. First both the 2K and 5K level of the PVLT seem to form a recurrent relationship to course grade over the two data points. Secondly, there is a noteworthy variation in which the levels of the RVLT are related to course grade at the first and second term of study, which makes it difficult to single out any level or levels that might be good predictors of academic achievement over the course of the learners’ studies of English.

Moreover, a strong relationship of productive knowledge of academic words to course grades was found in the second term of study. Overall the scores of the individual levels of the productive size test exhibit a more consistent pattern in terms of how they are related with course grade than the scores of the levels of the receptive test. In addition to a strong relationship between productive knowledge of academic words and course grade, one can conclude that the PVLT is probably a better predictor of academic achievement than the RVLT among learners in this context.

6.1.4 To what extent is lexical richness of student essays related to essay grade and overall academic achievement?

A weak relationship was found between lexical richness of all essays, i.e. essays written within both disciplines, and first-term course grades. However, no relationship was found between lexical richness and essay grade. Survey data revealed that faculty teachers do not place much
emphasis on lexical richness features when assessing essay quality, which means that other features of essay quality are weighed in the grading.

6.1.5 How do Swedish students’ vocabulary knowledge and student-essay lexical richness develop over time?

The informants had a large onset receptive vocabulary of about 7800 word families. An increase was found only between the end of term one and two. The increase was limited to the 10K level. At the beginning of term a majority of the students mastered the 5K level. Although their size of low-frequency vocabulary increases, a majority do not reach the cut-off point for mastery of the 10K level by the end of the second term.

A relatively large onset productive vocabulary was also found among the informants, of 4900 word families. Unlike their receptive size, which was only found to increase after two terms, their productive vocabulary grew already after one term as well as after two terms. The increase was at the 5K and 10K levels.

In contrast to the students’ receptive level-mastery profile, which remains stable across the two terms, their productive level-mastery profile improves with a majority of the students reaching the cut-off point for mastering the 3K level by the end of term two. On the basis of the levels that a majority of the students were found to master at the three data points, three vocabulary learning milestones can be identified. In regard to receptive size, it seems as we should expect Swedish learners to be able to have receptive mastery of the 5000 most frequent word families in English after completing upper-secondary school. The other two milestones are identified for productive size. For productive size two vocabulary learning milestones can be identified. The first one is the 2K level which should be mastered by Swedish learners at the end of upper-secondary school. The second one is the 3K level which a majority of the students investigated here mastered by the end of the second term of study. However, since the sample was relatively small in the second term, the results obtained for that term should be interpreted with caution.

The largest increase in percentage figures is in the informants’ depth of vocabulary knowledge. There is an observed increase in their knowledge of synonyms over both one and two terms of study, whereas knowledge of collocations increases only over two terms of study.

A relatively high increase in the lexical richness of student essays could be observed between the first and second term of study.
6.2 The development of vocabulary knowledge and use among Swedish university students of English

An important question informing our understanding of the developing vocabulary ability of advanced learners as well as having a crucial relevance to pedagogy is how advanced learners’ vocabulary knowledge develops in the context of a language-oriented university course with no particular focus on vocabulary. Such insight can provide teachers and administrators with a reference point for the evaluation of the effectiveness of vocabulary learning components in English language courses at universities in Sweden and other comparable situations.

6.2.1 The state and development of receptive size

As noted in section 3.4.1, most like the current study which have documented growth in receptive vocabulary among adolescent and adult learners were carried out in formal learning contexts where vocabulary is not explicitly taught but unlike the present study, all were characterized by a relatively low onset size. Before dealing with the extent to which the students’ vocabulary size develops over the course of their studies, it is of relevance to first discuss its onset size.

6.2.1.1 The students’ state of receptive size

As expected, the total onset receptive size among the students was above 7000 word families (cf. Gyllstad 2007). At the beginning of their university studies they had an estimated overall vocabulary size of about 7800 word families. The relatively large onset vocabulary size among the students was expected also in light of English being taught at school from a relatively early age. Moreover, in the Swedish learning context there is a high receptive exposure to English (cf. sections 4.3.1; 3.3.1.1).

In regard to the expected lexical profile of the students, they were not expected to master the 10K level at any of three data points examined here (cf. Gyllstad 2007). However, it should be pointed out that they probably have an overall vocabulary size well over 8000 word families. A test which had included lower frequency bands such as the Vocabulary Size Test (cf. section 2.4.3.1.3 would probably have captured this. On the basis of their scores on the RVLT it would not be far-fetched to infer that they have met the minimal vocabulary size of 10000 words needed for university studies in a second language context according to Hazenberg and Hulstijn (1996). Moreover, a majority of the students (24/34) mastered the 5K level already at the onset of their university studies. Since the 5K level has been shown to be a crucial lower threshold for being able to function appropriately in an L2 (cf. Henriksen et al. 2008; Hazenberg and Hulstijn 1996; Staehr Jensen 2005) and since a majority of the students examined here already master this
level at the onset, it could be used as a diagnostic threshold for detecting which students run the risk of failing the English language course. Among the ten informants who did not master the 5K level at the onset of their studies, five subsequently failed the course, among a total of eight students who did not pass the course.

It should be pointed out that Hazenberg and Hulstijn’s (1996) estimate is applicable to university studies in which the medium of instruction is the second language. A slightly lower estimate might be in place for university studies through the medium of a foreign language, since such courses are typically designed with the foreign language learner in mind, while courses which are taught in the second language typically target the native speaker. It is important to point out that Hazenberg and Hulstijn’s (1996) conclusion that “the minimal vocabulary size needed for university studies is 10,000 base words” can be somewhat misleading since it only covers the amount of vocabulary needed for reading not the amount needed for productive tasks.

If we look at what kind of tasks Swedish students should be able to carry out on the basis of the onset receptive size observed here, we see that they would have no problems with listening comprehension of academic spoken texts (cf. Schmitt 2008; Staehr Jensen 2005). However, as noted in section 3.3.1.1, the amount of vocabulary needed for receptive tasks such as reading and listening comprehension is expressed in terms of thresholds. One problem in the research literature is that no distinction is made between mastery of certain levels in order to be able to comprehend texts and the overall vocabulary size needed to perform such tasks.

For instance mastery criteria adopted here (cf. section 4.2.1) might indicate that a learner had not mastered any of the frequency bands up to the 5K level and yet this learner might have an overall vocabulary size of 5000 words or word families. Although an overall vocabulary size of 5000 words or word families entails knowledge of the most frequent vocabulary - since words are typically acquired according to their frequency of occurrence – you might very well encounter learners with an overall vocabulary size of 5000 who have mainly acquired lower-frequency words and whose vocabulary provides much less text coverage than the vocabulary of learners who have acquired words strictly according to their frequency of occurrence. In other words, an overall vocabulary size of 5000 word families does not necessarily imply that all frequency bands up to the 5K level are known. For instance Nation (2006) states that:

If we take 98% as the ideal coverage, a 8,000–9,000 word-family vocabulary is needed for dealing with written text, and 6,000–7,000 families for dealing with spoken text.

If we conceive this statement as expressing the overall size needed for those tasks, then our students would have little difficulty comprehending spoken
texts and come close to the threshold needed for comprehending written
texts. However, since Nation 2006 bases his estimate on coverage figures,
the recommendation should be understood as referring to knowledge of a
specific frequency range, i.e. knowledge of all or most word families up to
the 8000 or 9000 most frequent word families for understanding written texts
and up to the 6000-7000 most frequent word families for being able to
understand spoken texts. This would entail that our informants should have
had a much larger overall vocabulary size to be able to reach these
thresholds. This highlights the importance of clarifying what it is one is
referring to when it comes to how much vocabulary is needed for performing
specific tasks.

In addition, it is important to define such recommendations in relation to
authentic language use. As pointed out by Nation (2006) a 95% coverage
would suffice if dictionaries are used when reading. On the basis of such a
proportion of coverage, the minimum vocabulary size needed for obtaining
an appropriate word coverage would be considerably less. In the university
context one could expect a high reliance on dictionaries as well as on
explanations from lecturers, since in most cases the course literature is
discussed and explained in lectures and seminars. Another aspect that is not
reflected by estimates of how much vocabulary is needed for comprehension
and which is relevant to the language use in higher education is that these
learners, presumably, have well developed higher-order skills which can be
called upon when reading in an L2 (e.g; Mcmillion and Shaw 2009;
Schoonen, Gelderen, de Glopper, Hulstijn, Simis, Snellings and Stevenson
2003). Laufer (1997) suggests that a receptive vocabulary size of 3000 word
families is required for a learner to be able to transfer L1 higher-level skills
(e.g. inference making and comprehension monitoring) to L2 tasks.

When estimating learners’ vocabulary size and relating it to thresholds
there are thus a few aspects one needs to take into account. First, how much
coverage is provided by the test items from which we draw inferences about
vocabulary size? Second, are we talking about overall size or mastery of
frequency bands? Third, in relation to a specific threshold, to what extent can
learners compensate for not reaching that threshold? Vocabulary size
thresholds for reading do not take learners’ compensatory abilities into
account. In other words, the context of language use, i.e. the third aspect, is
important in how we understand whether learners reach a certain threshold.
This line of argumentation coincides with Chapelle’s (1994) stress on the
context of language use as an important component of vocabulary ability.

Above the students’ onset receptive size in light of thresholds for reading
and listening was discussed. In addition ambiguities in the way such
thresholds have been formulated and some of the difficulties involved in
using them as benchmarks for shedding light on whether learners are
equipped with an appropriate vocabulary size have been highlighted. Next,
we will look more closely at how our students’ receptive vocabulary size compares to that of other learners.

The students’ onset receptive vocabulary is relatively large in comparison to previous estimates obtained for university students in other learning macro-contexts (e.g. Barrow et al 1999, Cobb and Horst 1999, Nurweni and Read 1999 and Tschirner 2004). Their onset receptive size can also be described as relatively large in comparison to the overall receptive size estimates for second language learners in Cameron (2002). These adolescent second language learners of English in Cameron (2002) had an estimated receptive vocabulary size of almost 6700 word families. In light of this estimate one could argue that with regard to receptive vocabulary knowledge, the Swedish learning context fits the categorization of a High-middle exposure learning macro-context.

This might mean that Swedish students in general are well equipped to use English in their studies. Accordingly, the findings of the current study do not lend support to Schmitt’s (2008; forthcoming) contention that the vocabulary sizes of learners typically fall well short of Nation’s (2006) recommendations, especially if we take compensatory strategies and the kind of learning macro-context into account.

As can be expected the Swedish students come to English studies equipped with a relatively large receptive size, and with regard to degree of knowledge followed by productive size and productive deep knowledge of academic words. Within depth the following degree-of-knowledge hierarchy can be established: knowledge of collocations, parts of speech and synonyms. Most studies that have examined learners’ vocabulary ability from the point of view adopted in the present study have included learners with a relatively less developed vocabulary ability such as Nurweni and Read (1999) or looked at a wider range of proficiency levels, making it difficult to pinpoint what proficiency level the findings can be generalized to, such as Mochizuki and Aizawa (2000) and Qian (1999; 2002). Therefore the findings about the Swedish students’ vocabulary knowledge in the present study are quite significant because they are probably typical of an increasing population of students in the on-line and globalized world.

6.2.1.2 The development of the students’ receptive size

Having discussed the onset receptive vocabulary profile of the sample, we will now consider how their receptive vocabulary size changes over one and two terms of university studies. One of the main aims of the current study was to establish what improvements take place in academic language courses with no particular focus on vocabulary learning, because such information can be used as a yardstick for assessing the relative benefits of laying more emphasis on vocabulary in English language courses. Since the testing described here, the faculty members at the English department at Stockholm University responsible for developing the course curricula have incorporated
more focus on vocabulary learning as well as acknowledging the importance of more qualitative word knowledge aspects, such as knowledge of collocations. In light of the state of the students’ onset vocabulary knowledge established in the current study, test developers at the department should devise or employ receptive vocabulary size tests that have a relatively high ceiling, as is corroborated by Shaw and McMillion (in preparation). For instance, in the sample examined here four of the respondents reached the cut-off point for mastering the 10K level and almost half of the informants obtained a score of 20 or above out of 30 on this level, already at the beginning of the first term.

Studies have shown that vocabulary acquisition is a continuous process among adult EFL learners and that when English is mainly encountered in school a large increase of between 300 to about 1000 word families can take place over one to two terms of studies among students with a below 5000 onset overall receptive size (e.g. Bennett 2006; Laufer 1995; Milton and Meara 1995; Schmitt and Meara 1997).

As noted in section 3.4.1 it was expected that no significant overall increase would be observed among the students in a learning context without focus on vocabulary, because of a supposed high onset receptive size. It was also anticipated that growth should only take place among below-average students, since the task demands would be relatively higher on these learners. No development was found after one term even among the 16 two-termers, and though there was a significant average increase of 700 word families after two terms, this was predominantly among the lower-scoring students. Thus there is an increase in the receptive vocabulary even among highly proficient learners who have a large onset receptive vocabulary, but as expected this growth seems to be limited to learners with a below-average onset receptive size. In the light of this observation, a receptive vocabulary size test with a higher ceiling might have captured a significant potential increase of receptive size also among the above-average students. However, generally the lower-scorers improved more than higher-scorers on the other tests as well.

If we consider the students’ average receptive size attainment level at the end of the second term of study, they seem to have an adequate coverage of words that frequently occur in spoken English (Nation 2006). Although their receptive vocabulary improves over one year of study, a majority of the students do not fully master the 10K level at the end of term two (cf. Shaw and McMillion in preparation).

Schmitt and Meara (1997) state that word frequency is not by itself a reliable determinant of the probability that a word is known. In the present study no substantial gaps were found in the informants’ receptive repertoire and as opposed to the informants in Laufer (1998) and Schmitt and Meara (1997) the increase was limited to the 10K level. One reason why the learners in the present study have a different learning profile is that they...
have enough basic vocabulary to carry out necessary language tasks, while the increase in low-frequency vocabulary is necessitated by the language demands of university studies. These levels represent the type of vocabulary needed for academic language tasks. In other words depending on the context of learning and the demands placed on learners’ vocabulary knowledge, if enough basic vocabulary is known to be able to carry out a wide range of language tasks, the increase will likely occur in the lower-frequency levels in order to meet the requirements of a more demanding context.

There is interplay between vocabulary acquisition and on the one hand the vocabulary already known and on the other hand the degree of sophistication of the language task subjects are expected to carry out. It might also be the case that the acquisition of lower-frequency vocabulary has to do with what words they are exposed to. For instance, despite gaps in a learner’s knowledge of frequent vocabulary, in an academic learning context a learner might primarily acquire words that are less frequent in general but nevertheless frequently occurring or perceived as important in the academic setting. In light of this and based on the results of the present study we should not talk about vocabulary size increases without reference to onset level, learning context and task demands. As noted above in this section the reason why the below-average students here increased more may be that they are being exposed to relatively more demanding tasks.

We can conclude from this that the analysis of the Swedish students’ receptive size has yielded three main findings that might benefit pedagogy and future research. First an important milestone was identified namely mastery of the 5000 most frequent words in English at the onset of university studies. This might be set as a vocabulary learning goal in the curricula for the upper-secondary school. It can also be included in the diagnostic assessment of beginner students at English Language and Literature courses at Swedish universities for identifying which learners risk failing their university studies. Since the private-school reform, an increasing number of Swedish university students come from different private schools where the quality of English teaching can be questioned and there is a growing number of students from other countries where English is not taught at school or taught considerably less than in the Swedish contexts. An easily administrable diagnostic test, such as the 5K level, can prove to be a valuable asset for identifying learners with too low receptive sizes to be able to manage university studies in English at an early stage. As is discussed in the sections pertaining to the development of the relationship between the variables, the varied formal learning experience among today’s students seems to be a prominent factor behind how different types of knowledge are interrelated at the beginning in comparison to further along in the studies.

In terms of the kind of growth of receptive size that can be expected, two main findings have emerged. First, growth seems to be limited to the 10K
level and, second, mainly below-average students were found to increase their vocabulary knowledge in a context without specific focus on vocabulary knowledge.

It might be the case that more emphasis on vocabulary learning could push the above-average students to further develop their receptive size. If we interpret Nation’s (2006) goals restrictively, then, in order for the students to reach these goals, more focus on vocabulary is needed. Schmitt (2008) suggests that an effective vocabulary learning programme needs to be long-term and based on providing a maximum amount of engagement with vocabulary, through both incidental and intentional learning. He acknowledges that there is no “‘best’ teaching methodology” (354), but that all learning partners should work according to the general principle of maximizing the engagement with lexis by integrating intentional and incidental components within an overall teaching methodology. Accordingly, a more principled approach to vocabulary teaching might work a long way towards pushing even relatively well developed lexicons further along the vocabulary learning continuum.

6.2.2 The state and development of productive size
6.2.2.1 The students’ state of productive size

In view of Tschirner’s (2004) findings of 22% percent of the German students mastering the 2K level productively, in a middle exposure learning macro-context, it should come as no surprise that a majority (26/34) of the Swedish master the 2K level productively. In line with the categorization of the Swedish learning context as a High-middle exposure learning macro-context, this result was expected.

On the basis of the students’ onset mastery level observed here, one could argue that productive mastery of the 2K level can be identified as a milestone in the development of Swedish learners’ productive size. It can be considered for inclusion as a vocabulary learning goal in the upper-secondary curriculum of English. Mastery of this level at the onset of English university studies might be considered as a threshold for passing the course, since out of the eight students who did not master the 2K level, four (among eight who failed the course) subsequently failed the course. It might thus be a useful diagnostic tool for identifying learners who run the risk of failing the course. However, with a small sample size, caution must be exercised, as the findings might not be transferable to the wider population of students.

It should be stressed again that when discussing the students’ productive size a more restrictive understanding of the number of members in the word family unit of counting should be employed than when talking about receptive size. In other words productive knowledge of a given lexical item
does not extend to as many of its related forms (i.e. word family members) as receptive knowledge would entail. Moreover, productive size as it is measured by the PVLT should not be perceived as the number of items a learner can use freely in authentic language use. The estimate should be perceived as the number of words that the informants can access when dealing with off-line language tasks, such as take-home-essay writing when they work with a well defined task and in which the context of the topic at hand aids lexical retrieval. The kind of vocabulary knowledge measured here can be characterized as lexical items that can be accessed for production with some aid, such as a context, and without time pressure.

With regard to the students’ productive size, Swedish students have had substantially less experience of putting their vocabulary knowledge into productive than into receptive use. The typical Swedish EFL learner interacts in English only in the classroom setting. Accordingly, one should probably expect a lower P/R ratio than among native speakers or learners in an L2 context.

Most studies estimating learners’ vocabulary size have looked at it receptively. One of the few studies that have performed an extensive investigation of learners’ productive size is Zimmerman (2004), who found that foreign exchange students at the highest proficiency level in a university English-language center had a productive overall vocabulary size of about 3600 word families. However, on average the highest-level students did not reach the cut-off point for mastery of any of the levels. In comparison to Zimmerman’s (2004) highest-level students, the Swedish students seem to have a large onset productive size. An estimate of the overall productive vocabulary size of the university students in Laufer and Nation (1995) is about 4350 word families. The onset total productive found among the Swedish students was 4893 word families.

In light of Chapelle’s (1994) emphasis on the context of vocabulary use, one might argue that tests of this kind, which measure productive mastery that can be categorized as form recall tests, are useful for the type of vocabulary use discussed here. Moreover, in the light of the move towards more authentic language tasks in school and the increased access to aids, such as computer thesauruses and dictionaries, it might not be entirely relevant to assess truly productive ability, i.e. unrestricted vocabulary use, unless one is concerned with online production. It might be the case that the call for truly productive context-independent tests has to some extent been made with tasks such as timed compositions and online-conversation situations in mind (e.g. Meara and Fitzpatrick 2000; Webb 2008).

6.2.2.2 The development of the students’ productive size

Having discussed the students’ onset productive size and on the basis of that proposed a tentative milestone in Swedish learners’ vocabulary, we will turn our attention to how it developed over the course of university studies. A
different developmental profile was expected for the students’ productive size than for their receptive size. In line with the expectations a significant increase was found over one and two terms of study. It was further anticipated that the increase would primarily be found among the students with a below-average onset productive size. This seems to be the case, since only these students made a significant improvement over one term of study. The low-scorers on the PVLT increased their productive size with about 10% over one term. Among the two-term students in general a significant increase was found over both one and two terms. A similar level of growth was observed between the beginning and end of term one as between end of term one and term two. This indicates that the development of productive size is incremental in nature.

In light of the students’ well developed onset receptive size, it makes sense that the increase in productive vocabulary size is more rapid. Schmitt (forthcoming) suggests that vocabulary learning is incremental in the sense that less advanced knowledge aspects, such as receptive size, form the basis for development of more advanced knowledge aspects, such as productive size and productive depth. This view seems to be corroborated by the findings. Unlike the students’ relatively well developed onset receptive vocabulary, which increased only over two terms, their productive size increased more rapidly. Another possible explanation for the more rapid increase of productive size might be that the students’ receptive vocabulary knowledge is not subjected to the same task demands, since they presumably have enough receptive knowledge to manage the receptive tasks at this level. In other words, although the difference in the extent to which the two vocabularies develop probably should be attributed to the different ceilings of the measures, there might be other explanations.

With regard to whether we should expect a development of learners’ vocabulary size in educational settings in which no particular emphasis is placed on vocabulary, Schmitt (forthcoming: 14) states that:

The scope of the learning task, and the fact that many learners fail to achieve even moderate vocabulary learning goals, indicates that it can no longer be assumed that an adequate lexis will simply be ‘picked up’ from exposure to language tasks focusing either on other linguistic aspects (e.g. grammatical constructions) or on communication alone (e.g. communicative language teaching). Rather, a more proactive, principled approach needs to be taken in promoting vocabulary learning, which includes both explicit teaching and exposure to large amounts of language input, especially through extensive reading.

Although on average there is an increase in the students’ productive size both over one and two terms of study, 13 out of the 34 students do not exhibit any growth over one term of study. Almost half of these 13, six students, have an onset productive size below the group average at the
beginning of the first term. Accordingly, if we disaggregate the average growth we can identify students who do not display any growth. As suggested by Schmitt (forthcoming), these students would probably benefit from a more direct vocabulary instruction. Moreover, Schmitt (forthcoming) highlights the importance of extensive reading for vocabulary development. We can speculate that in the absence of a focus on vocabulary, the growth that was observed over a range of vocabulary knowledge dimensions is a consequence of the kind of language tasks, such as the reading of course literature, essay writing, and listening to lectures, they engage in. From these they pick up new lexical items as well as entrench the knowledge of already acquired vocabulary.

By introducing more focus on vocabulary learning we can enhance the chances of growth among a larger portion of the student population. I would argue that an effective way would be to employ diagnostic tests through which we can identify the range of vocabulary that should be targeted for each individual learner at the onset of the students’ university English studies. Such an approach would enable us to make the above-average students aware of what vocabulary range lies immediately beyond the range they already know and thus enable the students to go beyond their current level of vocabulary knowledge. This approach might also benefit the below-average students in that the course designers and faculty teachers avoid setting the vocabulary learning goals too high in relation to their onset level of knowledge.

Having discussed the onset state and development of the students’ receptive and productive size, in the next section we will discuss their depth of vocabulary knowledge from the perspective of onset state and development.

6.2.3 The state and development of productive depth

6.2.3.1 The students’ state of productive depth

In the previous two sections it was established that we did find growth in learners’ receptive and productive size in a learning context without focus on vocabulary. Before discussing the development of depth, the students’ onset level will be briefly touched upon again. It will be discussed in terms of the degree of difficulty of the sub-tests and will form the basis for determining whether the development is related to degree of difficulty. This is important since depth is perceived as a knowledge dimension which develops slowly and it is of interest to see whether degree of difficulty determines the development of components of depth. Depth will be discussed in relation to criteria for choosing knowledge components for a more focused vocabulary instruction in that those likely to develop more slowly should be specifically targeted. In light of the cross-sectional results of earlier research a growth in
learners depth was expected (e.g Greidanus and Nienhuis 2001; Greidanus, Beks and Wakely 2005).

Judgments as to the relative difficulty of the sub-tests and inferences as to the relative degree to which the various components have been acquired at the onset of study are highly dependent on the test instrument used. In other words, it is difficult to compare learners’ degree of mastery of sub-skills across studies, since studies often adopt different measures. If we look at the students’ depth of productive knowledge of academic words, on average they can supply the correct collocation for eight of the academic prompt words out of a total of twenty. Roughly speaking, this would mean that the informants were able to produce an appropriate collocation for 40% of the overall population of academic words. In the case of the informants’ knowledge of the parts of speech of academic words, on average the informants knew 25% of the parts of speech of the test items. This depth of vocabulary knowledge component seems to be relatively less developed than their knowledge of collocations of academic words. Furthermore, they knew on average 20% of the synonyms of the test items. One reason why these two knowledge components are relatively less developed might be that they tapped into the students’ metalinguistic knowledge. Accordingly, the relatively low average scores on these sub-tests might be explained by a low onset metalinguistic knowledge and awareness among the students. The test rubrics in these two sub-tests entail knowledge of the meaning of terms such as synonym, noun, etc. At any rate, it is difficult to ascertain whether the students have a well developed deep knowledge of academic words, since we do not have any established frame of reference. Therefore, it must be pointed out that the test measured productive knowledge of academic words and that their onset depth of vocabulary knowledge must be interpreted in light of that.

The research field would benefit from the development of standard depth tests that measure both receptive and productive knowledge so that a higher degree of comparability can be obtained across studies. Developing such tests could be a daunting task, as was the case for the present study, since if we set out to measure depth according to the components approach, we must ascertain that the various sub-tests measuring the knowledge of a specific component have the same degree of difficulty in order to allow comparisons across components as well as across studies. This is almost impossible, since there are no criteria for determining that tests of different depth components have a similar degree of difficulty. The only way of possibly obtaining such criteria would be by using native-speaker score norms. Moreover, in order to obtain an appropriate level of construct validity we must work out the best way to measure learners’ knowledge of components of depth and thus avoid ways of testing this which tap into knowledge of other constructs, which is a problem that might have arisen in the present study. This might work a long way towards enhancing the validity of measures of depth and also our
knowledge of different facets of learners’ depth of vocabulary knowledge. However, as pointed out by Milton (2009: 17):

It often proves hard to find a method that can unambiguously measure just a single element of vocabulary knowledge. Because language knowledge has to be assessed indirectly, a variety of other types of knowledge and skill may be required by the learners, which may colour the vocabulary produced.

This statement is indeed applicable to the depth test developed here, and as noted above in this section for two of the sub-tests (i.e. the word derivation and synonym sub-tests) other knowledge types than those targeted were probably tapped. More generally Milton’s (2009) observation raises an important question whether our conceptualization of depth as a unitary construct encapsulating highly interrelated components is a viable one. A number of researchers have questioned whether depth should be seen as a separate independent dimension (e.g. Vermeer 2001, Gyllstad 2007, Meara and Wolter 2004 and Read 2004). Read (2004: 224) states that:

It may be time, then, to dispense with the term depth and to recognize that any substitute that one might propose – precision, richness, elaboration, quality – is equally problematic as a cover term for the state of a learner’s vocabulary knowledge that goes beyond a rough estimate of how many words are known.

In light of these reservations and Milton’s (2009) observation, when we measure what we term depth we are in reality tapping into multiple different knowledge types that may or may not be connected. It must be said that depth and other constructs within the humanities are seldom easily definable (e.g. justice, power and democracy). Nevertheless, they are useful as a means of interpreting complex phenomena in the specific fields where they have been used, such as vocabulary knowledge.

The way forward is probably as stated by Milton (2009: 169) “that depth may have to be modified if it is to remain useful as a dimension at all, since nothing keeps the elements comprising it together terribly persuasively …”.

Going back to the findings of the present study in terms of how they compare to previous findings, we see that the degree of difficulty is to some extent corroborated by other studies. Staehr Jensen (2005), who also operationalized depth according to the components approach, and who examined two similar depth of vocabulary knowledge components (viz. knowledge of synonyms and collocations) found that the synonym sub-test was more difficult than the collocation sub-test for his informants. Moreover, Schmitt and Meara’s (1997) results indicate that a test measuring suffix knowledge had a higher degree of difficulty than a test measuring word association knowledge. These findings together with the results of this study seem to suggest that a rough hierarchy in terms of the extent to which
different components of depth of vocabulary knowledge are mastered can be established, notably, syntagmatic before paradigmatic. However, such a hierarchy is contradicted by the findings of Greidanus et al. (2005) who found that their informants mastered paradigmatic associations to a higher extent than syntagmatic associations. It is likely that the test employed here measures a different construct from the one employed in Greidanus et al. (2005). This lack of comparability between studies as noted above is perhaps one of the reasons why the validity of the construct has been questioned in the research literature.

6.2.3.2 The development of the students’ productive depth

Having dealt with the students’ onset depth of vocabulary knowledge and having discussed some of the inherent problems with the concept of depth we will turn to the extent to which our students’ depth conceptualized according to the components approach developed over the course of their university studies. A significant overall increase in depth was predicted on the basis of previous research that has shown that learners’ depth develops in learning contexts where vocabulary is not taught explicitly (e.g. Greidanus, Beks and Wakely 2005; Greidanus and Nienhuis 2001; Schmitt 1998; Schmitt and Meara 1997). The significant increase in the students’ depth of vocabulary knowledge in this study corroborate these earlier findings.

Moreover this study confirms that knowledge of the components develops in a nonparallel manner (cf. Schmitt 1998; Schmitt and Meara 1997). On the question of how the three components of depth developed, this study found that only one of the components, knowledge of synonyms, developed over the course of one term. The increase was quite large, i.e. 33%. Over the course of one year there was a progression in both knowledge of synonyms and collocations of 35% and 19%, respectively. In view of the relatively short time period of one and two terms these results are quite encouraging, especially in view of the relatively high increase found in the present study. As noted in relation to the growth found for the students’ productive size, the relatively high course demands in combination with a well developed onset sight vocabulary are probable factors behind an increase of more sophisticated facets of vocabulary knowledge. However, it is important to note that the differences observed in the development of the variables examined can be due to different degrees of sensitivity of the tests at capturing a potential increase. Moreover, as noted above, comparing the development of depth of vocabulary knowledge is very difficult because of different conceptualizations, and even if studies work with the same conceptualization, the lack of standard tests gives rise to different operationalizations of the construct.

The observed growth found in the current study might be a result of the formal learning context where there is a requirement for depth of vocabulary knowledge, i.e. the ability to produce appropriate collocates of academic
words as a way of expressing different nuances of an idea and where the language exposure is of a nature that leads to development of such facets of vocabulary knowledge.

These results suggest that although one can expect a development of advanced learners’ receptive vocabulary size (mainly among the below average students), a relatively larger and more rapid increase seems also to take place in productive size and productive depth (i.e. knowledge of synonyms), presumably, because these are more important above a certain receptive size threshold to academic success and therefore it is in these areas students are motivated to improve. Consequently, vocabulary development must be understood in relation to learners’ onset vocabulary knowledge and the communicative demands of the learning context.

However, these results must be interpreted with some caution. Samples were relatively small. Moreover, the increase might in part reflect a development in other skills such as test-taking skills and metalinguistic knowledge, especially in the synonym knowledge sub-test where the test-takers had to be familiar with the meaning of the word *synonym* for them to be able to provide correct responses. As mentioned above, the construct of depth has been questioned on the basis of difficulties in delineating the construct. More research is needed before we can obtain a clearer picture of how advanced learners’ depth of vocabulary knowledge develops over time. Furthermore, before we can say anything about the relative degree of development of different components of depth we need to develop better test instruments which measure the target components and which control for different degrees of learning burden of a word. For some words it may be easier for a learner to know the collocational profile than the part-of-speech or synonymy profile. One factor for this might be that whereas collocations by their very nature occur together in the language input, the synonym or part-of-speech form of a specific lexical item occurs separately from the item itself, which to a learner obscures the word-form and synonymy links between words.

Test results are inevitably dependent on the particular lexical items selected and their possibilities. Gyllstad (2007: 215) reports a personal communication from Meara who questions whether depth should be conceived as a dimension in the same way as size, because whereas size is used to describe the whole lexicon, depth is frequently used to describe the degree of knowledge of individual words. This use of depth is probably a crucial factor behind the alleged ‘fuzziness’ of the construct, because it is used to describe knowledge of a property that varies according to which lexical item is targeted.

One pedagogical implication of the lack of development of word derivation knowledge found here is that the introduction of a vocabulary component should include focus on word form. Schmitt (2008) argues for the importance of placing focus on word form in vocabulary instruction. He
also stresses the need to highlight formulaic sequences for learners in vocabulary instruction.

6.2.4 Vocabulary use: the lexical frequency profile of students’ take-home essays

Turning our attention to the lexical frequency profile of the students’ first-term essays, we see that it differed according to discipline. They produced more B-2000 vocabulary in their linguistics essays than in their literature essays. This might be due to a higher reliance on technical vocabulary within the linguistics discipline than in the literature discipline. The literature texts were close readings of different kinds of literary works and probably call for more high-frequency vocabulary. Consequently, the measure seems to be able to discriminate between different text types and the students seem to have a certain degree of register awareness in terms of vocabulary use.

If we look at the average lexical frequency profile of the essays of both disciplines lumped together, we see a relatively high proportion of advanced vocabulary of 17%. Obviously, as is the case with different assessments of depth of vocabulary knowledge, it is difficult to make meaningful comparisons across studies since differences in lexical frequency profiles can be attributed to other variables, such as task and genre type. However, in order to obtain an idea of whether this is a high or low figure, for exploratory reasons, we can compare with the lexical frequency profiles obtained for learners in other studies. For instance, Muncie (2002) found that his subjects produced approximately 12% B-2000 vocabulary in multiple-drafts essays. Intermediate to advanced ESL learners at a Canadian university have been found to produce on average 11.6% B-2000 vocabulary in timed compositions (Morris and Cobb 2004). A slightly lower profile was found among Israeli EFL learners at the beginning of their first term of English Language and Literature studies (Laufer 1994). It might have been a cause for concern if the frequency profiles found in the current study had been lower than these. Naturally, more data is needed before we can draw firmer conclusions about how the lexical frequency profiles obtained for the Swedish learners compare to other learners and writing tasks. It is also important to keep in mind that there are many factors in addition to language proficiency and text type that might account for these differences, thus making it difficult to ascertain whether this is a somewhat typical lexical frequency profile for advanced learner-take-home essays. The learners in the present study, however, seem to roughly produce the amount of vocabulary one might expect given the type of writing task and proficiency level. Overall, their frequency profile seems to be within an expected level i.e. higher than in timed compositions and slightly higher than in intermediate-learner multiple-drafts essays in Muncie (2002). Moreover, as noted above,
the frequency profile captures discipline differences. The next question is how much it develops between term one and two.

6.2.4.1 The development of lexical richness
As regards the development of lexical richness, on the basis of the findings of Laufer (1991; 1995) an increase in the Swedish students’ use of advanced vocabulary was anticipated. It was noted in section 3.4.3 that Laufer and Paribakht (1998) found no growth in lexical richness among Israeli upper-secondary learners. One possible reason for these conflicting results is that there are many intervening variables between a learner’s productive knowledge and vocabulary use as reflected in lexical richness which give rise to considerable variation in lexical richness from one time to another and/or across one writing task to the other. Figure 11 illustrates some possible intervening variables which can be assumed to cause variation in lexical richness.

Figure 11. Intervening variables between a learners’ vocabulary knowledge and the lexical richness of the output.

When comparing the lexical richness of learner output either across learners or between the same learners at different points in time, we should acknowledge that the construct is dependent on many intervening factors, such as the ones illustrated in Figure 11 as well as on task type (cf. Bulté, Housen, Pierrard and Van Daele 2008). With this said, in the present study it was shown that the proportion of advanced vocabulary in the students’ essays increased by 3 percentage points. If we assume that the development is in tandem with a development of language proficiency, it supports Laufer and Nation’s (1995) claim that there is a relationship between lexical richness and language proficiency. However, it must be pointed out that the
writing conditions in the second term are not identical to those of the first term, which means that the difference could be due to different writing conditions. It might be the case that no real development has taken place in the kind of vocabulary knowledge which learners can access for unrestricted lexical use. One can only speculate that the increase found for the kind of productive size examined here and productive depth of vocabulary knowledge percolated into our students’ vocabulary use. In light of the increase in their productive size and depth, it is not too far fetched to infer that some of this knowledge becomes incorporated into the learner’s essay writing. Another explanation is that their higher-order skills have developed, which decreases the load on their short-term memory enabling them to access more advanced vocabulary when engaging in academic writing tasks. Thirdly, one may speculate that as their writing skills improve they become more prone to use, and better at using, different aids at their disposal, such as the language tools in word processors.

6.2.5 Final remarks on the development of the students’ vocabulary knowledge and use

In view of the relatively large onset receptive size and the development of both productive size and depth over one term, one might be inclined towards concluding that there is no immediate need for a greater emphasis on vocabulary learning in the English Language and Literature course. However, this large onset receptive size and the observed increase in productive size and depth might suggest that the students perceive a gap in their productive knowledge which leads them to work on vocabulary. Moreover, we see an increase in learners’ knowledge of low-frequency vocabulary – presumably, words encountered in the course of their studies. The vocabulary is probably acquired mainly by incidental learning. These learners might benefit from a more focused vocabulary teaching that complements and guides the natural intake of new vocabulary as well as the deep learning of already known lexical items that was found among the students (cf. Nation 2001; Schmitt forthcoming). Moreover, although there is an average overall development of the students’ vocabulary knowledge and use as reflected by the increase in receptive size, productive size, productive depth and their use of advanced vocabulary in their academic writing, one must take note of the individual variation underlying the average increase. For instance, six out of the 16 students exhibited a small or no increase in receptive size over two terms of study. Pedagogical improvements must target these learners. The vocabulary learning goals should be set in relation to both below- and above-average students by setting appropriate vocabulary learning goals which force them to look upward.
The main theoretical implication of the results pertaining to the development of the variables examined here is that we should not examine learners’ vocabulary acquisition divorced from (i) their initial level of vocabulary knowledge, since the onset level affects the course of development and (ii) the learning context, because learning contexts provide different kinds of exposure to lexis, which might affect what kind of lexical knowledge develops. Accordingly, the slow growth of receptive vocabulary among the students examined here is not particularly alarming in view of these two factors, i.e. onset level and context.

6.3 The relationship between dimensions of vocabulary knowledge

When looking at the static relationship of the variables examined here, the second data point will be used as a point of departure, since all informants were administered the same depth test, which means that this data point provides more reliable results. In regard to the discussion of the relationship of vocabulary knowledge and lexical richness to grades, only the end of term data points will be treated, since lexical richness, essay grade and course grade are mainly end of term measures.

6.3.1 The relationship between receptive and productive size

The overall close relationship found between receptive and productive size was expected and confirms earlier findings (e.g. Laufer 1998, Laufer and Paribakht 1998 and Meara and Fitzpatrick 2000).

The strong relationship indicates that an appeal is made to the same kind of knowledge. This fits well with the notion of viewing the distinction between receptive and productive knowledge in terms of a continuum rather than a binary distinction. However, although it is the same kind of knowledge it is also somewhat different because receptive size and productive size are differently related to productive depth. The difference can probably be characterized as a difference in the degree of productive knowledge the two tests tap into.

6.3.2 The relationship between size and depth

On the basis of the review of the literature a strong relationship of size to depth was anticipated. The relationship of size to depth has mainly been investigated receptively and within the framework of the network knowledge approach. In regard to the relative relationship that receptive and productive size displays with productive depth, it was expected that there would be a
stronger relationship between the same kind of vocabulary knowledge i.e. productive size and productive depth. Overall the relationship between size and depth supports earlier findings of a moderate to strong relationship (e.g. Greidanus, Boogards, van der Linden, Nienhuis and de Wolf 2004, Qian 1999, 2002 and Qian and Schedl 2004 Albrechtsen et al 2008 and Staehr-Jensen 2005). One can conclude that we should expect to find a strong relationship between size and depth among learners of different proficiency levels. No evidence was found that the relationship is higher or lower in this sample than what has typically been found in other samples. However, because students at the tertiary level come equipped with varied vocabulary knowledge profiles when starting their university studies of English, it might be rather difficult to set goals appropriate for all students. As suggested by Milton 2007, foreign language learners might display an irregular frequency profile with knowledge gaps in high-frequency lexis due to the kind of language exposure typically found in foreign language contexts. In such contexts words are not acquired according to frequency of occurrence but are rather learned thematically from the text books which contain a high proportion of infrequent lexis. It might be the case that foreign language education also produces gaps across different dimensions of vocabulary knowledge, such as size and depth, due to schools placing different emphasis on such dimensions.

In order to shed light on the extent to which an appeal is made to the same kind of knowledge in regard to the relationship of receptive size to productive depth a further analysis was conducted. On the basis of a regression analysis with depth as the dependent variable in which PVLT was entered first and RVLT second it was shown that RVLT did not add any significant predictive value on top of that already afforded by PVLT. This indicates that for assessment purposes, one should probably only employ a size test as a predictor of the corresponding kind of depth of vocabulary knowledge. Moreover, when talking about the nature of the relationship of size to depth, one should arguably need to further specify what type of vocabulary knowledge (i.e. receptive or productive).

On the basis of the results of the present study one can only speculate that the question of whether the two constructs, i.e. size and depth, are identical should, arguably, be approached from the point of view of the receptive/productive distinction and also the extent to which they contribute to other skills. Those (Qian 1999, 2002 and Staehr-Jensen 2005) who have attempted to answer this question from the point of view of whether size and depth can make unique contributions to reading and listening performance maintain that they should be treated as two separate constructs. In light of the results of the present study it might be the case that the relationship varies as a function of the receptive-productive continuum, even to the extent that at certain points of this continuum the relationship might be so close that they should be treated as the same construct. For instance, test-takers’
performance on a size and a depth test that tap into the same kind of productive knowledge would probably be more strongly related than test-takers’ performance on a size and a depth test that reflect different kinds of receptive/productive knowledge.

Talking about the relationship in an either-or manner forces the empirical findings into a narrow framework that makes us miss the complexity of the matter at hand. Maybe we should perceive the relationship as a scale in which certain aspects of depth cannot be distinguished from size and that the degree to which they produce a separate contribution is a function of the task at hand. For instance, it was found that depth was related to lexical richness, while receptive size did not exhibit a relationship. Moreover, the current findings suggest that we should expect a closer relationship according to the same kind of knowledge, i.e. productive size and productive depth rather than perceiving the relationship as uniform across the receptive/productive continuum. This idea finds some support in Qian and Schedl (2004: 30), who state referring to the dimensions comprising vocabulary knowledge:

“These dimensions are not only intrinsically connected but also interact closely with one another in fundamental processes of vocabulary use and growth. The importance of the various factors in these dimensions will vary according to the specific purpose of language use. For instance, a receptive process may involve a different set of factors than those which may be involved in a productive process.”

According to Qian and Schedl (2004) the relative importance of size and depth differs according to what language task is performed. The degree of interrelationship from this point of view can be regarded as a function of the task at hand. In view of this one might argue that depth is not a unified construct but presumably a cluster of characteristics with varying relations to size.

The results of the present study indicate that there is a higher likelihood of finding a learner with a relatively large receptive vocabulary size but a less developed productive depth in comparison to a mean average in the sample than finding a learner with a relatively large productive vocabulary knowledge but a relatively less developed productive depth of vocabulary knowledge. The question is whether we would typically find a closer relationship between size and depth at the receptive end of the control continuum. It might be the case that since in general learners in Sweden attain a large receptive vocabulary it is not as good a predictor of depth as productive size since a large productive vocabulary size indicates that the learners have engaged in production, which has implications for productive depth. However, we cannot rule out that we have a ceiling effect which depresses the results. I would argue that it is not the type of word knowledge, (receptive or productive size) per se but the relative experience with
encoding and decoding language that produce the variable relationship receptive and productive size exhibit with productive depth. In other words, the reason why a stronger relationship is found between the same kind of knowledge, i.e. productive size and productive depth, is simply because both knowledge types derive from the same kind of language exposure, i.e. productive use.

An important implication of the finding of a variable relationship of receptive and productive size to productive depth among advanced learners is that if we want to obtain a picture of the overall state of the advanced learners’ mental lexicon we cannot solely rely on receptive size measures, because such measures might only provide an accurate picture of overall receptive knowledge. From an assessment perspective, by using a test battery which comprises both a receptive and a productive size test (not necessarily the size tests adopted here), we might obtain a better assessment of the test-takers’ receptive and productive vocabulary knowledge.

Another implication concerns the validity of the PVLT as a measure of productive knowledge. Webb’s (2008) claim that the PVLT might actually be a measure of receptive knowledge does not seem to be corroborated by the findings of the present study, since they were differently related to productive depth. If the two tests actually both measure receptive knowledge one should arguably have found a similarly strong relationship with depth of vocabulary knowledge (cf. section 6.3.2). Moreover, Webb (2008) seems to be working with a binary R/P distinction not a continuum, which might explain his way of categorizing measures in an either-or manner. However, since the PVLT is widely used in empirical studies as a measure of productive vocabulary size (e.g. Fan 2000, Waring 1997 and Tschirner 2004) further studies are needed into the validity of the test, for more certainty in this area.

6.3.3 The interrelationship of components of depth

A modest to high interrelationship among the components examined here was anticipated on the basis of earlier findings (e.g. Schmitt and Meara 1997; Staehr Jensen 2005).

The focus of the network knowledge approach in the definition of depth of vocabulary knowledge is on syntagmatic and paradigmatic sense relations since these are viewed as indicators of the organization of the mental lexicon. By contrast, in the present study, knowledge of word class derivations was operationalized as one of three components of depth of vocabulary knowledge.

The results indicate that knowledge of word class derivations is a valid component of depth of vocabulary knowledge since it correlates both with the other components and with size of vocabulary knowledge, thus lending the construct some concurrent validity.
6.3.4 The relationship of vocabulary knowledge to lexical richness

In light of previous findings of a modest relationship between vocabulary size and lexical richness in timed compositions, a significant weak relationship of vocabulary size to lexical richness was anticipated. A weaker relationship than found in previous studies was predicted, since the writing task examined here places different demands on the writer. Contrary to expectations no significant relationship of receptive and productive size to lexical richness was found.

In reference to the relationship of depth to lexical richness an unspecified significant positive relationship was expected on the basis of Bulté et al.’s (2008) model of lexical competence, in which it is postulated that lexical sophistication, which is the type of lexical richness investigated here, should mainly be seen as a correlate of qualitative word knowledge. Accordingly, the results seem to corroborate their description of the nature of the relationship of lexical sophistication to the underlying construct of lexical competence.

Although lexical richness should primarily be perceived as a text feature rather than a skill, it can be perceived as an indirect index of writing ability or as an instantiation of learners’ ability to use advanced vocabulary in their writing. The findings of the present study showed a modest relationship between depth of vocabulary knowledge and lexical richness of students’ take-home essays. In contrast to Laufer and Nation (1995) who found a close relationship between productive size and the LFP of learners’ timed compositions, no such relationship was found in the present context. One possible reason for there being a difference between the degree to which depth in comparison to size of vocabulary knowledge is related to lexical richness is that lexical richness, albeit to a limited extent, is a reflection of academic writing ability which draws on sophisticated knowledge of academic words. As stated by Albrechtsen, Haastrup and Henriksen (2008), learners progressively acquire more and more language abilities which in turn build on each other as well as on additional knowledge and skills accumulated. Moreover, they suggest that these abilities are relatively contained within specific task domains. In view of this, depending on the task at hand different abilities and knowledge types are called upon. The relative reliance on a specific ability or knowledge type should arguably vary according to the specific language task performed.

It might also be a function of language proficiency. Among less proficient learners who are unable to transfer the higher-order skills they have acquired in their L1 one might find a stronger link between vocabulary knowledge and lexical richness in their writing. Accordingly, Laufer and Nation’s (1995) finding that there is a strong link between learners’ vocabulary size and the lexical richness of their written output, can only be applicable to
similar learners and context. However, the applicability of their findings is rather restricted since the study was conducted in an uncontextualized fashion, i.e. we lack information about aspects such as the learning context of the subjects, which makes it difficult to know which contexts and learners the results can be generalized to.

Earlier findings show that depth and vocabulary size are equally good predictors of reading performance (e.g. Qian 1999 and 2002), but in the present study neither receptive nor productive size showed a significant correlation with lexical richness. The pedagogical implication of this lack of relationship is, in contrast to Qian’s (2002) recommendation that both measures should be used in reading assessment, that it would only make sense to use productive depth measures as potential predictors of the extent to which learners use advanced vocabulary in writing.

However, since assessing lexical richness, as opposed to assessing reading comprehension, in most cases is not a pedagogical end in itself, the empirical results of the present study suggest that it would make sense to further examine the relative predictive value of vocabulary size and depth in relation to other indices of language use. If we look at previous findings with reference to the relationship between vocabulary knowledge, both size and depth, and writing ability, Albrechtsen et al (2008) found that neither depth nor size correlate with language quality in the written texts of high-proficiency learners. This result seems to further support the idea that among highly proficient learners who can rely on a wider range of sub-skills in both their L1 and L2 a straightforward relationship between vocabulary knowledge and indices of productive skills might be difficult to identify.

Depth and the components: knowledge of collocations and word derivations showed a modest correlation with lexical richness in all the essays. Depth of vocabulary knowledge seems to facilitate the use of more advanced vocabulary, whereas mere familiarity with the meaning of words is not enough to be able to use them in academic writing. It might very well be the case that depth of vocabulary knowledge implies more connections between words in the mental lexicon, hence faster retrieval. Learners with a less developed depth of knowledge of academic words, might retrieve these words more slowly and opt for higher-frequency words which are more readily accessible. Even if depth is found to be more closely related than size (receptive or productive) to lexical richness, it is rather tenuous to claim that this relationship means that depth is more closely related to writing proficiency than size. There are simply too many mediating factors between depth of vocabulary knowledge and writing proficiency to be able to draw such a conclusion. However, what we can conclude from the findings is that there is a relationship between depth and a specific instance of vocabulary use in a writing task for which the writers are not pressed for time and have access to aids that can be used to compensate for a lack for words. Accordingly, although all students have access to various aids that they can
use to enhance the lexical richness in their essays, learners with a deep knowledge of academic vocabulary tend to use more advanced vocabulary than learners whose knowledge is shallower in this area. This seems to suggest that even though learners might have access to various aids for a given writing task, such as dictionaries to help them enhance the lexical richness of their writing, the vocabulary knowledge in the form of depth of vocabulary knowledge that they bring to the writing task affects the lexical richness of the text. Another possible explanation is that learners’ depth of vocabulary knowledge contributes to a more effective use of different aids or that depth of vocabulary knowledge plays a role in decreasing the cognitive load enabling the learners to produce more infrequent vocabulary in their academic writing.

One important reservation should be made regarding the relationship of depth to lexical richness. The relationship has been discussed here mainly in terms of the cover term depth. This is not an entirely accurate way of seeing the relationship. Although all three components were found to be related to lexical richness, the correlational profile varied both over time and in terms of the strength of relationship of the components to lexical richness. Whenever we use measures of depth which are a composite score of various skills and find a relationship with some other skill, we are in reality superimposing a framework on the empirical data. Meera and Wolter (2004) have stated that testing word knowledge in terms of isolated elements is a classic example of not seeing the wood for all the trees. However, a way forward as suggested by Milton 2009 and Read (2004) would be to look at individual elements of vocabulary knowledge, such as collocational knowledge without invoking the notion of depth, because there is a risk that we will not see the trees for the wood. By working with clearly defined constructs without attempting to lump disparate elements of vocabulary knowledge into one single construct and thus losing sight of what it is we are examining we can better apply our findings to vocabulary teaching.

The term might very well be appropriate pedagogically to direct teachers’ and learners’ attention to the fact that there is more to vocabulary acquisition than merely learning the meaning of words. However, for research purposes the term is probably too fuzzy for being used to study learners’ vocabulary knowledge (cf. Milton 2009).

6.3.5 The relationship of vocabulary knowledge to course grade
A number of studies have examined vocabulary size among EFL learners studying English or through the medium of English at the university level in the light of the extent to which they reach curriculum vocabulary learning goals. Generally, EFL learners’ vocabulary size has been found to fall short of vocabulary learning goals perceived as crucial for being able to succeed academically (e.g. Nurweni and Read 1999, Barrow, Nakanishi and Ishino
However, none of these studies have investigated the relationship between vocabulary size and academic achievement in the specific learning context examined in order to determine whether falling short of vocabulary learning goals is detrimental to academic success. Most of the subjects in the studies above had relatively small vocabularies (i.e. between 1300-3000 basewords) but it is difficult to believe that practically all of them would actually fail their university studies. Although this claim is not put forth, these studies tend to perceive the state of affairs as very problematic. However, it begs the question of how vocabulary size is related to academic achievement.

Although there are some contradictory findings regarding the relationship between vocabulary size and school and academic performance, the vast indirect evidence of a relationship between vocabulary size and overall grade suggests that we should expect to find a positive relationship between size and academic achievement. If, as suggested by David (2008), grades at higher levels of education incorporate to a higher extent productive knowledge of vocabulary, we should also anticipate a stronger association of productive size and productive depth with course grade.

In regard to the relationship between vocabulary knowledge and grade, both dimensions as well as components of depth show a similar level of correlation with course grade in the first term of study. As expected a positive relationship of vocabulary size to academic achievement was found. In addition as noted in section 3.3.2 a positive relationship of depth to academic achievement was expected. The present findings seem to be consistent with other research, which have mainly examined the relationship indirectly by looking at how depth is related to various language tasks (e.g. Mecartty 2000; Qian 1999; 2002)

However, contrary to expectation, this study did not find a stronger relationship of productive size than of receptive size to course grade at least on the basis of first-term data.

The positive relationship of size to course grade accords with our earlier suggestion, that students who did not master the 2K level productively and the 5K receptively at the onset of their studies were overrepresented among those students who subsequently failed the course. This finding supports the recommendation put forth in sections 6.3.1 and 6.3.2 to incorporate these level tests as diagnostic measures for identifying which students do not have enough receptive and productive vocabulary for managing English studies at the university.

In regard to the relationship of the components of depth to course grade, if we see the grade as a very indirect reflection of proficiency, then a rather wide range of vocabulary knowledge components are equally related to proficiency. This together with the findings of a patchy interrelationship between the components seems to suggest, interestingly enough, that the components are not equally related to each other but equally related to the
broader construct of language ability or perhaps better yet academic ability if we regard course grades to some extent as an indicator of these. One possible reason might be that grades tap into a multiple of different skills which are more or less related to different facets of vocabulary ability.

A practical implication of the relationship found between vocabulary knowledge and course grade, although we cannot be sure of there being a causal relationship between the former and the latter, is that course developers should formulate the course aims or requirements in the form of vocabulary learning goals. This would make the link between academic performance and vocabulary knowledge more salient among the students who after all are studying in a foreign language. One way of achieving this would be to evaluate the English Language and Literature course in terms of what range of vocabulary including phraseology a learner should master receptively and productively to be able to cope with the kind of language tasks they are exposed to at a foreign language discipline course. The question is to what extent we could rely on available word lists such as the AWL or estimates of the amount of vocabulary needed to cope with university studies done in a different learning context than the Swedish one. In regard to the AWL and the extent to which it could be used as a basis for formulating vocabulary learning goals in this context, Hyland and Tse (2007) found that the words in the AWL were considerably overrepresented in certain disciplines, such as computer engineering, which raises doubts as to whether the list is an appropriate basis for vocabulary learning goals in the discipline of English Language and Literature.

We cannot stress the importance of vocabulary knowledge for general language proficiency without also offering the means by which learners can improve their vocabulary. At the beginner stages a more hands-on approach is probably a cost-effective method, but at the advanced stages a more indirect method in which we supply learners with well-defined vocabulary lists for specific language needs is probably more appropriate.

One good strategy, which is probably appropriate for our Swedish university students, is to supply them with a productively-oriented academic word list. A promising attempt at developing such a list is documented by Paquot (2005) who presented a new methodology for the identification of lexical items in existing word lists, such as the GSL and the AWL that should be part of a productively-oriented academic word list. She argues for the need to establish productively-oriented word lists as opposed to solely relying on receptively-oriented word lists which do not supply the learners with usage information and lack inclusion of the phraseological and contextual properties of words.
6.3.6 The relationship of lexical richness of student essays to essay and course grades

On the basis of previous findings of a weak to modest relationship of lexical richness of learners’ timed compositions and overall text quality, a similar relationship was anticipated here. In addition evidence indicates that lexical richness of learners’ timed compositions is related to school and academic grades. Accordingly, a weak to modest relationship between the lexical richness of students’ take-home essays and course grade was expected.

With regard to the relationship of lexical richness to essay grade, contrary to expectations no significant positive relationship was found. This finding seems to contradict the claim by Laufer and Nation (1995) that lexical richness at least as reflected by the LFP (the Lexical Frequency Profile) is a reliable index of overall text quality. Of course the discrepancy between their claim and the findings of the present study must also be seen in light of the relationship being investigated in different contexts, such as the proficiency profile of the samples, the different writing tasks examined and the different assessment criteria. However, in reference to the relationship of lexical richness to course grade, this study does confirm that lexical richness is associated with overall grade (Morris and Cobb 2004; Richards, Malvern and Graham 2008).

One major determinant of whether one should expect to find a relationship between lexical richness and any global rating of a text, such as essay grades or any other quality rating is how much emphasis on vocabulary is included in the assessment of essay quality. Consequently the lack of relationship found here can be explained by the tendency among the teachers to focus relatively more on extra-lexical features when grading the students’ essays. The LFP measure does not seem to be a good measure of overall text quality when it comes to take-home essays, regardless of discipline, simply because the assessment of take-home essays seems to be more content and grammar oriented. Consequently, one plausible reason for earlier research having found such a relationship might be that the holistic rating to which the lexical richness scores were correlated emphasizes lexical features, such as precisely lexical richness and quality. In light of this, when examining the relationship between linguistic features and the overall quality of a written text, it is important to delineate the criteria in the holistic rating, to which the linguistic features examined are related. However, although the LFP does not discriminate between take-home essays in terms of overall text quality as reflected by the grade awarded, it seems to be a valid predictor of academic achievement. Presumably, this is because both reflect underlying academic ability.

From a pedagogical point of view two main conclusions can be drawn from the findings. First in an academic context in which students’ written texts are assessed on the basis of other features than lexical richness the LFP
might not be a valid indicator of text quality as reflected by the grade awarded to the specific essay and should probably not be used for assessing student essays. However, the LFP measure can be used as a diagnostic tool to identify students who predominantly rely in their writing on high-frequency vocabulary. It can, thus, be used as a pedagogical tool to spot learners early on who run the risk of failing the course due an over-reliance on high-frequency words in their essay writing.

6.4 The development of the relationship between dimensions of vocabulary knowledge and use

In this section the development of relationship between the variables examined will be discussed in light of previous findings. Some theoretical and practical implications will be proposed.

6.4.1 The development of the relationship between receptive and productive size

If we consider the relative development of the students’ receptive and productive size, learners with a relatively larger productive vocabulary size do not develop either in their productive or receptive size, while learners with a relatively larger receptive vocabulary develop their productive size. This pattern can be seen in Figure 12. In order for a stronger correlation to be observed, there has to be a relatively larger increase along one axis. The circles represent the beginning-of-term scores of the RVLT and PVLT and the stars represent the end-of-term scores.

The RVLT/PVLT scores at the bottom right quadrant seem to move upwards along the Y-axis by the end of term. The upward move is represented by the stars in the bottom right quadrant which are higher up along the Y-axis.

This developmental profile might be explained by a number of different factors. One key factor pertains to the learning context of Swedish EFL learners who tend to develop a large receptive vocabulary due to the high input of English. This might explain the relatively slower development of the informants’ receptive size.
Figure 12. Overlay scatterplot of the relationship between receptive and productive size at the beginning and end of term one.

Subsequently, the learners become exposed to increasing demands on their productive vocabulary as they attempt to meet the course requirements at the English Language and Literature programme. Many researchers have highlighted the importance of learners’ noticing processes in the development of the L2. The notion of ‘noticing’ is intrinsically linked to the input/interaction approach to understanding L2 acquisition. This intrinsic relation of noticing to the input/interaction approach as defined by Gass and Selinker (2003: 224):

takes as its starting point the assumption that language learning is stimulated by communicative pressure and examines the relationship between communication and acquisition and the mechanisms (e.g., noticing, attention) that mediate between them.

Although this view stresses the local language needs, i.e. a break down of communication in a specific instance of interaction, it might be extended to more global language needs, namely the students perceiving a gap as a result of the language demands of a global context, i.e. the university setting, rather than a more local context such as specific interactions. In general, there seems to be an increase in the relationship between two variables as a function of a relatively larger increase in one.
6.4.2 The development of the relationship between size and depth

Previous research has shown that the relationship of size to depth varies according to sub-level proficiency\(^{22}\) (Nurweni and Read 1999; Staehr Jensen 2005). This has mainly been examined cross-sectionally by comparing the correlations between size and depth scores obtained for sub-samples split according to some measure of sub-level proficiency (e.g. Nurweni and Read 1999; Staehr Jensen 2005). Both these studies have found that the relationship is stronger among higher-sub-level proficiency learners than among lower-sub-level proficiency learners. The results of the present study suggest that it is possible to detect a corresponding stronger relationship between the two dimensions within the same learner over time.

However, it must be pointed out that we are dealing with different sets of tests correlated at the different data points. It might be the case that the stronger correlation between the depth scores and the size scores at the second data point in comparison to the first is an artifact of the test administration procedure employed here (cf. section 4.4.1). In other words the differences might be due to differences in the measures rather than differences in time. This needs to be controlled for in future studies in order to obtain a clearer picture of this area. However, if this were genuinely the case then we would expect a similar relationship between the variables at data point one and three since the same set of measures were used (cf. section for details about the test administration). This does not seem to be the case, which means that the differences between the data points can be attributed to time and the variable development of the variables in relation to each other.

The relationship between different dimensions of vocabulary knowledge has been approached from a static point of view in many studies (Greidanus, Boogards, van der Linden, Nienhuis and de Wolf 2004, Qian 1999, 2002 and Qian and Schedl 2004, Staehr Jensen 2005). We need to take into account factors such as the learning context and developmental factors which arguably are important factors in how dimensions of vocabulary ability are related – especially in view of the continuous development of learners’ vocabulary ability as was shown in the present study as well as previous studies (e.g. Schmitt and Meara (1997), Nurweni and Read (1999), Lewis 2009, Milton 2006). It might be problematic to infer from studies showing a strong relationship of size to depth that it somehow reflects a close relationship between the two dimensions in learners’ mental lexicon. If indeed, as indicated here, the relationship between the two becomes stronger over time, it suggests that contextual factors are at play. In connection to a

\(^{22}\) Sub-level proficiency was defined in section 3.1.1 as proficiency differences within broader categories such as “beginner”, “intermediate” and “advanced”.

182
discussion of lexical profiles Milton (2007) suggests that if language instruction is thematically structured learners’ lexical profile might appear irregular. In other words if learners’ acquire language by thematically based instruction, vocabulary learning might not follow a neat word frequency trajectory. If we relate Milton’s (2007) line of reasoning to the findings of a weaker relationship between size and depth at the onset of the students university studies a possible explanation for this pattern emerges. This might be that since the university learning context has learners from different schools with diversity in the quality of English teaching and diversity in which language skills are emphasized, at the onset of their university studies there are more students with an asymmetrical vocabulary knowledge profile than further along in their studies, due to a similar-exposure effect. This accords with our previous observations, which showed that the increase in vocabulary is typically found among the students with a below-average onset level. Two processes are operating: they are getting better, and their profiles are getting more similar. Rather than simply improving uniformly, they are filling gaps.

However, a few reservations should be made regarding what conclusions can be drawn from these results. First, we are dealing with group scores, and the distance from these to associations between constructs in individual learners’ minds is too great to allow reliable conclusions. Second, we cannot rule out that the discerned pattern of a development of the relationship between size and depth scores is a reflection of a number of other factors, test formulation being one, rather than evidence of a real insight into how different knowledge dimensions are related both statically and longitudinally. To this should be added that the depth dimension is a hotly debated construct (cf. section 6.2.3).

From an assessment perspective, it would make sense to employ a wider range of measures for learners from different formal language learning contexts in order to obtain a fuller picture of the overall state of their vocabulary knowledge.

6.4.3 The development of the interrelationship of the components of depth

The development of the interrelationship of the components of depth was expected to be similar to that of size and depth. The results show that learners’ knowledge of collocations, word derivations and synonyms does indeed become more interrelated over time. This finding accords with previous studies such as Schmitt and Meara (1997), who found tentative evidence of an increase in the strength of correlation between components over time. However, since Schmitt and Meara (1997) found a modest relationship at the receptive end in the first term between knowledge of
derivative suffixes and word associations, it might be the case that the 
interrelationship between components of receptive depth increases in 
strength earlier in the course of learners’ L2 language development. This 
might suggest that the development of the interrelationship between 
components of depth at the receptive end differs from the development of the 
interrelationship at the productive end.

For instance, low-proficiency learners might exhibit a relatively strong 
relationship between different receptive abilities such as receptive 
knowledge of word derivations and receptive knowledge of synonyms of 
words. When the same learners are tested on their productive knowledge of 
the same aspects, they perform less consistently on the two measures which 
generates a low correlation between the two scores. As we move higher in 
terms of level of proficiency the relationship between productive abilities lag 
behind that of receptive abilities.

In this context that would mean that if we tested the same learners on two 
receptive measures of depth each targeting a different component, one can 
speculate that a strong relationship would have been found already at the 
onset of their studies as opposed to the modest relationship found between 
productive knowledge of the components examined here. This is just 
speculation but it highlights the importance of making the context of 
learning and the proficiency profile of the subjects examined explicit to 
allow for valid comparisons and in extension an insight into the different 
developmental stages a typical learner progresses through.

6.4.4 The development of the relationship of vocabulary 
knowledge to the lexical richness of student essays

Since a significant relationship of only one dimension of vocabulary 
knowledge, i.e. depth, to lexical richness was found, we will only discuss the 
development of vocabulary knowledge to lexical richness from the point of 
view of depth of vocabulary knowledge.

In the first term of study a modest relationship of depth to lexical richness 
was found. This relationship is attributable to knowledge of word 
derivations. Accordingly, since not all components were similarly related to 
lexical richness, we cannot in a strict sense talk about a positive relationship 
between depth and lexical richness.

In the second term a different component exhibited a modest relationship 
to lexical richness, namely knowledge of synonyms. Indeed, the variability 
to which the components of depth were related to lexical richness both 
synchronously and longitudinally, suggest that we cannot off-hand say that 
there is a general relationship between depth as a cover term and vocabulary 
use.
Another issue arising is that the variability to which components are related to lexical richness indicates that it is difficult to neatly categorize different sub-skills under one single cover term.

However, the main reasons why we found a variable relationship should be sought after in lexical richness and in the small sample size in term two. Vocabulary use in authentic tasks is subjected to a complex set of variables which might explain the variable relationship longitudinally found here (cf. fig. 11 in section 6.2.4.1). Moreover, the small sample size in the second term would seem to have generated irregularities such as these.

6.4.5 The development of the relationship of vocabulary knowledge and lexical richness to course grade

Despite clear data about how the relationship develops in other contexts, a stronger relationship of lexical richness to essay and course grade was expected in the second term than in the first term.

However, in regard to the development of the relationship between lexical richness scores and essay and course grade, the results of the two data points do not enable a comparison due to the insignificant results obtained for the analyses. Further research is needed to explore whether there is a relationship between the proportion of advanced vocabulary, here operationalized through the B-2000 measure, and authentic essay quality and academic performance assessment.

If we look at the development of the relationship of vocabulary knowledge to overall grade, a similar relationship was observed for the two terms. This suggests that vocabulary knowledge is an important predictor of academic success both at the beginning of university studies and further along.

From an assessment perspective if prediction of academic success is the focus then it would be more practical for English teachers at the university to opt for a more easily administered size test, since the two size measures employed here were found to be relatively stable predictors of academic success over time. The efficacy of using size tests as measures of learners’ language development at various levels is supported by Milton (2009) who argues for the use of vocabulary size tests on the basis that they provide the same information as more comprehensive and less practical language proficiency measures that are used to test learners’ progress.
7. Conclusion

The following conclusions can be drawn from the current study: On the basis of the Swedish students’ onset size, two crucial frequency levels of receptive and productive size of vocabulary knowledge were identified, the 2K and 5K levels, respectively. Those students who did not master these levels when they entered university were more likely to fail the course. These two levels can therefore be considered for inclusion in vocabulary learning goals for students entering university studies of English. It was also proposed that a practical way of diagnosing first-term students’ chances of managing university studies of English in the Swedish context, is to test them on these levels.

The study also showed that a rough estimate of the average student’s receptive and productive size is 7800 and 4900 word families, respectively. These kinds of data are useful for informing the teaching of English, generally, and could be used for evaluating the effectiveness of English language teaching in the final years of upper-secondary school.

In regard to the development of the vocabulary knowledge and use of Swedish university students of English, two main conclusions can be drawn. Even in a learning context where no particular focus is placed on learning vocabulary one can expect growth in learners’ vocabulary knowledge and use. However, the increase were mostly found among those learners with a below-average onset vocabulary knowledge and use of advanced vocabulary. The strongest relative growth was found in the students’ productive size and depth of vocabulary knowledge which accords with Milton’s (2009: 77) proposition that “… very advanced learners, who already have large vocabularies, might develop depth and fluency of use at the expense of size, in order to make gains and become more proficient”. Moreover, although obtaining a relatively high overall receptive size, only a minority of the students mastered the 10K level receptively at the end of the second term of study. These findings have implications for pedagogy in that they highlight the need for a more focused vocabulary instruction in order for the above-average students and the minority of weaker students who do not improve to develop their vocabulary and in order for the students to reach the 8000-9000 threshold for being able to comprehend a wide range of English written texts identified by Nation (2006).

Another implication is that the extent of increase found here could be used as a point of reference for evaluating the efficiency of vocabulary
instruction treatment at the tertiary-level. A theoretical implication of the findings is that the context of learning is an important determiner of vocabulary growth. This should be included in a theory of vocabulary acquisition. Vocabulary development seems to be a function of onset level as well as context. If the growth found among the below-average students can be attributed to the learning context in the sense that the task demands are set at a level just beyond their onset level, but not high enough for the above-average students, it might explain the relatively smaller growth among the students with relatively large onset vocabularies.

It was shown in the current study that there is a closer relationship between types of knowledge at the same point on the receptive-productive continuum, i.e. productive size and productive depth. An implication of this is that teachers of English at the university should not only rely on assessing receptive size but also productive size in order to obtain a fuller picture of students’ productive mastery of words. By employing a test battery which includes a productive size test, which is typically easier to administer than productive depth tests, teachers at the university are able to obtain a practical predictor of the overall state of vocabulary knowledge.

However, the findings also indicate that a measure of productive depth should be considered for inclusion in a test battery of vocabulary knowledge, since it was the only dimension that displayed a significant correlation with the lexical richness of student essays at a variety of levels. Among relatively advanced learners, size tests are not necessarily valid predictors of the sophistication of vocabulary use. One issue emerging from this finding relates to the question of what an estimate of advanced learners’ vocabulary size really can tell us about their vocabulary use in authentic high-stakes writing tasks. If teachers are interested in obtaining an indication of whether students will be able to make use of an appropriate register of words in the writing tasks they are required to carry out at the university level, assessing depth of vocabulary rather than solely relying on size measures might be more appropriate. However, it must be stressed that no significant relationship was found here between lexical richness and essay grade. It might be the case that lexical richness at least as it was measured here is not a valid index of text quality as reflected by essay grade. In order to be able to formulate a theory of second language vocabulary competence, we must in addition to establishing how dimensions of vocabulary interact also establish the extent to which dimensions of size and depth, which are regarded as central dimensions of learners’ lexical competence, play a role in different contexts of vocabulary use.

In regard to what knowledge aspects should be part of a definition of depth, on the basis of the findings which showed a strong interrelationship between learners’ knowledge of collocations, word derivation forms and synonyms, a definition of depth should include knowledge of word derivation since this was found to be strongly related to knowledge of both
collocations and synonyms, which are identified in the network knowledge framework as indicators of the quality of learners’ mental lexicon.

This study has shown that there is a development in the degree to which different dimensions of vocabulary knowledge are related over time within the same learners. These findings are consistent with those of Nurweni and Read (1999) and Staehr Jensen (2005) who found that the relationship between the dimensions size and depth varies in strength according to sub-level proficiency. An implication of this is the possibility that learners’ knowledge of different dimensions and components becomes more integrated due to similar learning experience in the university learning context. If this is the case then a wider range of vocabulary tests is preferable for diagnosing the overall state of learners’ vocabulary ability at the onset of their university studies of English when students come equipped with a more varied vocabulary ability profile due to different learning experiences at pre-tertiary education. Moreover, employing a wider range of vocabulary tests that tap into different dimensions of vocabulary knowledge might provide an indication of which dimensions of vocabulary knowledge learners need to further develop in order for them to meet the departmental requirements.

The main theoretical implication of this finding is that we should avoid talking about a relationship between size and depth in a static absolute sense. It is important to specify the context of the relationship and also take into account that it is subject to change even over a relatively short period of time.

The relatively strong and stable relationship of vocabulary knowledge to course grade could lead developers of English language syllabuses of the importance to place more emphasis on vocabulary learning. Moreover, the relationship also highlights the need for an ongoing assessment of learners’ vocabulary even at the university level for identifying learners who have gaps in their vocabulary knowledge and therefore run the risk of not reaching the English language requirements of the specific educational setting. This way awareness of potential gaps in learners’ vocabulary knowledge can be raised and appropriate measures can be taken proactively within the course of studies. In light of the stronger relationship found between aspects of the student’s vocabulary knowledge and use as they progress through the study of English, a smaller and thus more practical set of vocabulary tests can be used at the later stages of study which would considerably decrease the resources needed for test administration.

7.1 Limitations of the study

In the present study there are two weaknesses that need to be acknowledged and addressed. First in regard to the research design, both a correlational and a longitudinal approach were employed. These two designs are fairly
difficult to combine on practical accounts. When conducting correlation analyses the significance of a correlation coefficient of a particular magnitude will change depending on the size of the sample from which it was computed. The minimum sample size needed to detect a significant correlation varies as a function of the strength of correlation. The second issue that might affect the reliability of the results of a correlation analysis is the degree of homogeneity in the variables in the data set. Since parallel tests were used in a counter-balanced design, there is a risk that the scores comprising the variables are not entirely compatible.

These two weaknesses are the result of limitations generated by the longitudinal design. In a longitudinal study it is difficult to obtain a large number of informants who are prepared to take tests on more than one occasion. Moreover, there is a higher drop-out rate. In the present study it was deemed necessary to use parallel tests in order to reduce the risk of a practice effect. As mentioned in the method section, since the study elicited participants on voluntary basis, there is a possibility that these learners take a particular interest in their language learning. This together with the fact that the population of students is relatively small leads to a situation in which participants know each other and might discuss the content of the tests. Accordingly, a parallel test design was deemed necessary to ascertain that a real increase has taken place rather than an increase due to the participants having retained the test items in their memories through discussions among themselves. Moreover, since it is very difficult to obtain and also develop two exactly parallel tests a counter-balanced test administration design was used.

Nevertheless, the advantages of conducting a longitudinal study are that we obtain an overall picture of how different aspects of vocabulary knowledge develop over time within the same learners. In contrast, in a cross-sectional study it is difficult to determine whether differences between learners at different course levels can be assigned to an actual development of aspects pertaining to their vocabulary knowledge, rather than some other factor related to differences in learning contexts, since course curricula are subjected to continuous revisions which affect the learning context.

Criticism can also be directed to the depth of vocabulary test employed. Due to the time constraint caused by the longitudinal design, the test was not sufficiently validated which among other things resulted in the word derivation sub-sections having a variable number of correct answers. This in turn had a negative effect on the validity of the results since it reduced the sample for which the development of depth of vocabulary knowledge could be tracked. Accordingly, the main strength of the study is also to some extent its main weakness.
7.2 Directions for further studies

In view of the lack of vocabulary growth among the students with a relatively well developed onset level, it might be a potential future research area to examine what kind of vocabulary students of English at Swedish universities are exposed to and whether it is a factor in vocabulary development at this level.

Another potential venue for further research is to examine the extent to which the components within the components framework are indeed related to depth of vocabulary knowledge or if they instead should be conceptualized as belonging to other language modules. For instance, correlation analyses in which additional components identified by Nation (1990; 2001) as comprising depth are related to other measures of vocabulary knowledge are needed in order to evaluate and refine the components framework. Until we obtain a clearer picture of which components comprise depth of vocabulary knowledge, we run the risk of examining knowledge components that are not relevant for understanding learners’ depth of vocabulary knowledge. The only way we could hope to reach some level of conceptual clarity regarding the nature of depth of vocabulary knowledge is if we can identify which elements are strongly related and thus constitute valid candidates for being included in a definition of it.
References


Shaw, P. and A. McMillion in preparation. “Vocabulary size and reading comprehension in first and second languages”,


Thorndike, E. L. and I. Lorge. 1944. The Teacher's Word Book of 30,000 Words. Teachers College, Columbia University, New York


Appendices

Appendix 1: Productive Vocabulary Levels Test, Version 1
Name: ______________________

Levels Test of Productive Vocabulary
Version 1
Instructions:
Complete the underlined words. The example has been done for you.

He was riding a bicycle

The 2000-word level
1. They had to cl__________a steep mountain to reach the cabin.
2. The doctor ex__________the patient thoroughly.
3. The house was su__________by a big garden.
4. The pirates buried the trea__________on a desert island.
5. Her beauty and ch__________had a powerful effect on men.
6. La__________of rain led to a shortage of water in the city.
7. He takes cr__________and sugar in his coffee.
8. The rich man died and left all his we__________to his son.
9. Pup__________must hand in their papers by the end of the week.
10. The railway con__________London with its suburbs.
11. She wan__________aimlessly in the streets.
12. The organisers li__________the number of participants to fifty.
13. If you blow up that balloon any more it will bu__________.
14. In order to be accepted into the university, he had to impr__________his grades.
15. The telegram was deli__________two hours after it had been sent.
16. The differences were so sl__________that they went unnoticed.
17. They sat down to eat even though they were not hu__________.
18. You must have been very br__________to participate in such a dangerous operation.
The 3000-word level
1. He has a successful car__________ as a lawyer.
2. The thieves threw ac__________ in his face and made him blind.
3. Before writing the final version, the student wrote several dra__________.
4. It was a cold day. There was a ch__________ in the air.
5. The cart is pulled by an o__________.
6. Anthropologists study the struc__________ of ancient societies.
7. After two years in the Army, he received the rank of lieu__________.
8. The statue is made of mar__________.
9. Some aristocrats believed that blue blood flowed through their ve__________.
10. Sudden noises at night sca__________ me a lot.
11. France was proc__________ a republic in the 18th century.
12. Many people are inj__________ in road accidents every year.
13. The secretary assi__________ the boss in organizing the course.
14. His beard was too long. He decided to tr__________ it.
15. People were whir__________ around on the dance floor.
16. He was on his knees, ple__________ for mercy.
17. I live in a small apa__________ on the second floor.
18. I won’t tell anybody. My lips are sea__________.
The 5000-word level
1. Some people find it hard to be independent. They prefer to be tied to their mother’s ap_________ strings.
2. After finishing his degree, he entered a new ph________ in his career.
3. They keep their valuables in a vau________ at the bank.
4. The old lady is too ill and fra______ to live alone.
5. The kitten is playing with a ball of ya__________.
6. The workmen cleaned up the me_________ before they left.
7. The building is heated by a modern heating appa__________
8. He received many com__________ on his dancing skill.
9. At the bottom of the blackboard there is a le________ for chalk.
10. The child was holding a doll in her arms and hu_________ it.
11. We’ll have to be inventive and de______ a scheme for earning more money.
12. The picture looks nice; the colours bl_________ really well.
13. The management held a secret meeting. The issues discussed were not disc_________ to the workers.
14. We could hear the sergeant bel________ commands to the troops.
15. The boss got angry with the secretary and it took a lot of tact to soo_________ him.
16. Nuts and vegetables are considered who_________ food.
17. Many gardens are full of fra_________ flowers.
18. Many people feel depressed and gl_________ about the future of mankind.
The Academic Wordlist level
1. There has been a recent tr__________ among prosperous families toward a smaller number of children.
2. Farmers are introducing innov__________ that increase the productivity per worker.
3. Phil__________examines the meaning of life.
4. The Far East is one of the most populated reg______ of the world.
5. You usually pay less if you buy in bu______.
6. He usually read the sports sec__________of the newspaper first.
7. He backed up his assertions by quoting the latest research and stat______.
8. There are several misprints on each page of this te__________.
9. His new book will be pub_______ at the end of this month by a famous University Press.
10. They insp__________ all products before sending them out to stores.
11. Despite his physical condition, his int_______ was unaffected.
12. The students were inhi_______ by the presence of the Dean of the Faculty.
13. He is irresponsible. You cannot re__________ on him for help.
14. It’s impossible to eva__________ these results without knowing about the research methods that were used.
15. He finally att__________ a position of power in the company.
16. The story tells about a crime and subs_________ punishment.
17. She didn’t openly attack the plan, but her opposition was imp_______ in her attitude.
18. The urge to survive is inh__________ in all creatures
The 10 000-word level
1. The baby is wet. Her dia________ needs changing.
2. For many people, wealth is a prospect of unimaginable felic________.
3. Second year university students in the US are called soph________.
4. Her favourite flowers were or________.
5. The insect causes damage to plants by its toxic sec________.
6. The evacu________ of the building saved many lives.
7. Much to his chag________, he was not offered the job.
8. The crowd soon disp________ when the police arrived.
9. The approaching storm stam________ the cattle into running wildly.
10. The hurricane whi________ along the coast.
11. Some coal was still smoul________ among the ashes.
12. The dead bodies were mutil________ beyond recognition.
13. She was sitting on a balcony and bas________ in the sun.
14. For years waves of invaders pill________ towns along the coast.
15. She found herself in a pred________ without any hope for a solution.
16. The deac________ helped with the care of the poor of the parish.
17. Computers have made typewriters old-fashioned and obs________.
18. Watch out for his wil________ tricks.
Appendix 2: Productive Vocabulary Levels Test, Version 2
Name: ______________________

Levels Test of Productive Vocabulary
Version 2
Instructions:
Complete the underlined words. The example has been done for you.
He was riding a bicycle

The 2000-word level
1. It is the de__________that counts, not the thought.
2. Plants receive water from the soil through their ro__________.
3. The nu__________was helping the doctor in the operating room.
4. Since he is unskilled, he earns low wa__________.
5. This year long sk__________are fashionable again.
6. Laws are based on the principle of jus__________.
7. He is walking on the ti__________of his toes.
8. The mechanic had to replace the mo__________of the car.
9. There is a co__________of the original report in the file.
10. I’m glad we had this opp__________to talk.
11. There are a doz__________eggs in the basket.
12. Every working person must pay income t__________.
13. This sweater is too tight. It needs to be stret__________.
14. Ann intro__________her boyfriend to her mother.
15. Teenagers often adm__________ and worship pop singers.
16. This work is not up to your usu__________standard.
17. The dress you’re wearing is lov__________.
18. He wasn’t very pop__________when he was a teenager, but he has
many friends now.
1. She has been changing partners often because she cannot have a sta__________ relationship with one person.
2. The pro__________ of failing the test scared him.
3. To improve the country’s economy, the government decided on economic ref__________
4. She wore a beautiful green go__________ to the ball.
5. The government tried to protect the country’s industry by reducing the imp__________ of cheap goods.
6. The children’s games were amusing at first, but finally got on the parents’ ner__________.
7. The lawyer gave some wise coun__________ to his client.
8. Many people in England mow the la__________ of their houses on Sunday morning.
9. The farmer sells the eggs that his he__________ lays.
10. Suddenly he was thru__________ into the dark room.
11. He perc__________ a light at the end of the tunnel.
12. Children are not independent. They are att__________ to their parents.
13. She showed off her sle__________ figure in a long narrow dress.
14. You’ll sn__________ that branch if you bend it too far.
15. You must wear a bathing suit on a public beach. You’re not allowed to bath na__________.
16. Crying is a nor__________ response to pain.
17. The Emperor of China was the supr__________ruler of his country.
18. You must be awa__________ that very few jobs are available.
The 5000-word level
1. Soldiers usually swear an o__________ of loyalty to their country.
2. The voter placed the ball__________ in the box.
3. The thieves have forced an ent__________ into the building.
4. On Sunday, in his last se__________ in church, the priest spoke against child abuse.
5. I saw them sitting on st__________ at the bar drinking beer.
6. His favourite musical instrument was a tru__________.
7. The small hill was really a burial mou__________.
8. We decided to celebrate New Year’s E__________ together.
9. People manage to buy houses by raising a mor__________ from a bank.
10. The soldier was asked to choose between infantry and cav__________.
11. After falling off his bicycle, the boy was covered with bru__________.
12. This is a complex problem that is difficult to compr__________.
13. The angry crowd sho__________ the prisoner as he was leaving the court.
14. Don’t pay attention to this rude remark. Just ig__________ it.
15. We do not have adeq__________ information to make a decision.
16. She is not a child, but a mat__________ woman. She can make her own decisions.
17. The prisoner was put in soli__________ confinement.
18. He is so depressed that he is cont__________ suicide.
The Academic Wordlist level

1. The ar__________ of his office is 25 square meters.
2. I’ve had my eyes tested and the optician says my vi__________ is good.
3. In their geography class, the children are doing a special pro__________ on North America.
4. In a free country, people are not discriminated against on the basis of colour, age, or s__________.
5. The money from fruit-picking was a suppl__________ to their regular income.
6. The drug was introduced after medical res__________ indisputably proved its effectiveness.
7. These courses should be taken in seq__________, not simultaneously.
8. A considerable amount of evidence was accum__________ during the investigation.
9. People have proposed all kinds of hypot__________ about what these things are.
10. You’ll need a mini__________ deposit of $20,000.
11. Results from the study ind__________ that men find it easier to give up smoking than women.
12. In a lecture, a lecturer does most of the talking. In a seminar, students are expected to part__________ in the discussion.
13. The airport is far away. If you want to ens__________ that you catch your plane, you’ll have to leave early.
14. It’s difficult to ass__________ a person’s true knowledge by one or two tests.
15. The new manager’s job was to res__________ the company to its former profitability.
16. Even though the student didn’t do well on the midterm exam, he got the highest mark on the fi__________.
17. His decision to leave home was not well thought out. It was not based on rat__________ considerations.
18. The challenging job required a strong, successful, and dyn__________ candidate.
The 10 000-word level
1. The new vic__________was appointed by the bishop.
2. If your lips are sore, try lip sal__________, not medicine.
3. The prisoner was released on par__________.
4. The actors exchanged ban__________with the reporters.
5. She wanted to marry nobility: a duke, a baron, or at least a vis__________.
6. The floor in the ballroom was a mos__________ of pastel colours.
7. She has contributed a lot of money to various charities. She is known for her generosity and bene__________.
8. This is an unusual singer with a range of three oct__________.
9. A thro__________ controls the flow of gas into an engine.
10. Anyone found loo__________ bombed houses and shops will be severely punished.
11. The rescue attempt could not proceed quickly. It was imp__________ by bad weather.
12. The wounded man squi__________on the floor in agony.
13. The dog crin__________ when it saw the snake.
14. He imme__________himself in a hot bubbly bath forgetting all his troubles for a moment.
15. I wouldn’t hire him. He is unmotivated and indo__________.
16. The problem is beginning to assume mam__________proportions.
17. His vind__________behaviour toward the thief was understandable.
18. He was arrested for illi__________trading in drugs.
Appendix 3: Depth of Vocabulary Knowledge Test, Version 1
Name: _________________________

Depth of Vocabulary Knowledge Test
of
Academic words
Version 1
Instructions:
Read the sentence and write one word in each gap. The word should make sense in the sentence/text and be grammatically correct. Try to choose a word that most native speakers would choose.

1 a) Many soldiers maintain they have never recovered from what they experienced in the Vietnam war and suffer from post-traumatic stress.

b) What word class forms can stress be realized as in addition to the one in the sentence (write the correct word form)?
Noun_________  Verb stress_____
Adjective stressed_____  Adverb_________

N.B. Not all word form slots can be filled for each and every word.

c) Provide two synonyms or near-synonyms of stress
1. _______strain____________
2. _______pressure_________

1 a) What exactly the author meant by that statement is __________ to interpretation. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What word class forms can interpretation be realized as in addition to the one in the sentence? (Write one correct word form per slot)
Noun_________  Verb_________
Adjective_________  Adverb_________

c) Provide two synonyms or near-synonyms of interpretation.
1. ____________________ 2. ____________________

2 a) Use the five steps and chart below to work out a reasonable estimate of the __________ of rebuilding your home. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What word class forms can estimate be realized as in addition to the one in the sentence? (Write one correct word form per slot)
Noun_________  Verb_________
Adjective_________  Adverb_________

c) Provide two synonyms or near-synonyms of estimate.
1. ____________________ 2. ____________________
3 a) This is a restricted __________. You are not authorized to enter. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can restricted be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of restricted.
1. ____________________   2. ____________________

4 a) More than two years after New York City cut 35% from its funding for the arts, The Brooklyn Museum continues to ________ the consequences. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can consequences be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of consequences.
1. ____________________   2. ____________________

5 a) People from various __________ minorities often face prejudice and discrimination. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can minorities be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of minorities.
1. ____________________   2. ____________________
6 a) I believed the drill instructor was __________ core, nose to the grindstone, always screaming and shouting. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can core be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of core.
1. ____________________   2. ____________________

7 a) In the last few years, attitudes have __________ and society now expects smokers to wipe out 70 years of brainwashing overnight. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can attitudes be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of attitudes.
1. ____________________   2. ____________________

8 a) In research conducted in the 1960s and 1970s substantial __________ emerged which showed that schools made little or no difference to pupils’ outcomes. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can emerged be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of emerged.
1. ____________________   2. ____________________
9 a) There is a growing conflict of _______________ between her position as a politician and her business activities. *(No grammatical words, e.g. pronouns, prepositions etc.)*

b) What other word class forms can *conflict* be realized as in addition to the one in the sentence? *(Write one correct word form per slot)*

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of *conflict*.

1. ____________________   2. ____________________

10 a) Mr Kinnock said it was wrong to propose a *fundamental__________* in the voting system for a short-term tactical vote to get rid of the Thatcher Government. *(No grammatical words, e.g. pronouns, prepositions etc.)*

b) What other word class forms can *fundamental* be realized as in addition to the one in the sentence? *(Write one correct word form per slot)*

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of *fundamental*.

1. ____________________   2. ____________________

11 a) You won’t succeed unless you _______________ the *instructions* rigidly. *(No grammatical words, e.g. pronouns, prepositions etc.)*

b) What other word class forms can *instructions* be realized as in addition to the one in the sentence? *(Write one correct word form per slot)*

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of *instructions*.

1. ____________________   2. ____________________
12 a) We accept as a __________ principle that this firm has a special obligation to participate in public service activities without expectation of compensation. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can principle be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of principle.

1. ____________________   2. ____________________

13 a) His fear is that he will be tortured to extract ___________ about his brother. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can extract be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of extract.

1. ____________________   2. ____________________

14 a) As I passed the hall in a large comprehensive ____________, where notices proclaimed “Quiet please: examination in progress”. I saw… (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can comprehensive be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of comprehensive.

1. ____________________   2. ____________________
15 a) In this sentence there are many ambiguous ____________ to which a range of possible meanings might be attached. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can ambiguous be realized as in addition to the one in the sentence? (Write **one** correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of ambiguous.
1. _________________  2. _________________

16 a) Terrorism has depressed the tourist industry, Morocco’s biggest single source of ____________currency, and ended remittances from Moroccan workers in Kuwait. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can **currency** be realized as in addition to the one in the sentence? (Write **one** correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of **currency**
1. _________________  2. _________________

17 a) Daniel, Panayiotis and Pavlos Xidis, three Jehovah’s Witnesses, are still serving four-year sentences in Greece for refusing to perform military ____________. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can **military** be realized as in addition to the one in the sentence? (Write **one** correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of **military**
1. _________________  2. _________________
18 a) We are establishing an ethics ______________ to look at the effects of advanced techniques in animal breeding. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can ethics be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of ethics.
1. ____________________   2. ____________________

19 a) Some people are easily influenced by the opinion of others, whereas others are able to maintain their ______________ integrity. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can integrity be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of integrity.
1. ____________________   2. ____________________

20 a) While the Poles appreciated German fears about the ______________ posed to Central Europe by a rise of Soviet power they did not see that they would be serving Poland’s interests by agreeing to Hitler’s demands. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can posed be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of posed.
1. ____________________   2. ____________________
Appendix 4: Depth of Vocabulary Knowledge Test, Version 2
Name: _________________________

Depth of Vocabulary Knowledge Test
of
Academic words
Version 2
Instructions:
Read the sentence and write one word in each gap. The word should make sense in the sentence/text and be grammatically correct. Try to choose a word that most native speakers would choose.

1 a) Many soldiers maintain they have never recovered from what they experienced in the Vietnam war and suffer from post-traumatic stress.

b) What word class forms can stress be realized as in addition to the one in the sentence (write the correct word form)?

Noun__________ Verb__stress______
Adjective_stressed______ Adverb__________

N.B. Not all word form slots can be filled for each and every word.

c) Provide two synonyms or near-synonyms of stress

1. _____strain____________   2. _______pressure____ ______

1a) In shifting responsibility from government to the people: “The Thatcher government has___________ a novel approach. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What word class forms can approach be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________ Verb__________
Adjective__________ Adverb__________

c) Provide two synonyms or near-synonyms of approach.

1. ____________________   2. ____________________

2a) Alcoholic beverages have __________a role in many societies for over 7000 years - before the existence of written history. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What word class forms can role be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________ Verb__________
Adjective__________ Adverb__________

c) Provide two synonyms or near-synonyms of role.

1. ____________________   2. ____________________
3a) Further, large-scale studies are needed to confirm the encouraging
obtained so far. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can obtained be realized as in addition to the
one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of obtained.
1. ____________________  2. ____________________

4a) Aware of his political inexperience, he sought ___________ from
his cabinet in domestic matters and the National Security Council in foreign
affairs.
(No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can sought be realized as in addition to the
one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of sought.
1. ____________________  2. ____________________

5a) Some of the men asked questions and ___________ comments, also
in sign language, and I asked Kalchu what they were saying. (No
grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can comments be realized as in addition to
the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of comments.
1. ____________________  2. ____________________
6a) British egg production had decreased because of constraints on it by the ministry. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can constraints be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of constraints.

1. ____________________   2. ____________________

7a) For obvious, the people of Hong Kong have long been reluctant to speak their minds. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can obvious be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of obvious.

1. ____________________   2. ____________________

8a) The Conservatives claim to have cut tax since 1979 while statistics show that total taxation as a proportion of GNP has increased. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can statistics be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of statistics.

1. ____________________   2. ____________________
9a) Death rates were expressed as standardised __________ ratios, with the national average as 100. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can ratios be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of ratios.
1. ____________________   2. ____________________

10 a) This produced complete deadlock in the conference, which now went to the highest governmental level, showing the __________ attached to the matter by both the USA and the UK. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can attached be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of attached.
1. ____________________   2. ____________________

11a) What all these tropical forests have in common, however, is their astonishing __________ diversity. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can diversity be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of diversity.
1. ____________________   2. ____________________
12a) It is often considered that there is much greater sexual tolerance than used to be the case and the public response to homosexuality and lesbianism is sometimes cited as __________ for this. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can cited be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of cited.
1. ____________________   2. ____________________

13a) Senior members of both firms refused to __________ or deny the story last night. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can deny be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of deny.
1. ____________________   2. ____________________

14a) The National Security Council (NSC) submitted a __________ to President Truman on 2 April 1948 reviewing the situation in Korea and recommending the principles that should govern future policy. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can submitted be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________  Verb__________
Adjective__________  Adverb__________

c) Provide two synonyms or near-synonyms of submitted.
1. ____________________   2. ____________________
15a) In cases of race and ________ bias, such decrees often have produced quotas and preferential treatment for the aggrieved party.

(No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can bias be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun_________ Verb_________
Adjective_______ Adverb_______

C) Provide two synonyms or near-synonyms of bias.
1. ____________________   2. ____________________

16a) John Wakeham, the new Secretary of State for Energy, had risen to public prominence in his previous Cabinet job by refusing to have nuclear ________ dumped in his constituency. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can nuclear be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun_________ Verb_________
Adjective_______ Adverb_______

C) Provide two synonyms or near-synonyms of nuclear.
1. ____________________   2. ____________________

17a) If employees had to move into temporary ________ while trying to find new apartments, they received a tax-free allowance towards the cost of the rent for a maximum period of 13 weeks. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can temporary be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun_________ Verb_________
Adjective_______ Adverb_______

C) Provide two synonyms or near-synonyms of temporary.
1. ____________________   2. ____________________
18a) In a”__________ scenario”, the UK Atomic Energy Authority has suggested that up to 10,000 Soviet citizens could be expected to die of radiation-induced cancer as a result of the Chernobyl accident, with a further 30,000 fatalities possible worldwide. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can scenario be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________ Verb__________
Adjective__________ Adverb__________

c) Provide two synonyms or near-synonyms of scenario.
1. ____________________  2. ____________________

19a) Despite persistent __________ in Washington of further infidelities, Mr Clinton says there will be no more surprises about him.(No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can persistent be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________ Verb__________
Adjective__________ Adverb__________

c) Provide two synonyms or near-synonyms of persistent.
1. ____________________  2. ____________________

20a) Furthermore, he had been summoned to appear in court in Grenoble on Sept. 26 in connection with an ongoing __________ into corruption charges in his capacity as president of the Alpes-Maritimes Council. (No grammatical words, e.g. pronouns, prepositions etc.)

b) What other word class forms can ongoing be realized as in addition to the one in the sentence? (Write one correct word form per slot)

Noun__________ Verb__________
Adjective__________ Adverb__________

c) Provide two synonyms or near-synonyms of ongoing.
1. ____________________  2. ____________________
Appendix 5: Teacher Assessment Procedures
Questionnaire
Questionnaire on Student Essay Assessment
Name:_____________________________

N.B. The questions pertain to the time before the implementation of the Bologna process, i.e. up to last term.

Instructions: Please read and answer the questions in the order they appear in the questionnaire. I want to know what spontaneously comes to mind when answering each question.

1. What kind of essays have you marked the last two years (Tick all relevant boxes)?
   a) Linguistics □ Literature □
   b) A-level □ B-level □
   C-level □ D-level □

N.B. Please note that none of the following questions pertain to C- and D-essays (C- eller D- uppsatser).

2. Do you follow any set of written criteria when rating essays?
   Yes □ No □

   If yes to question 2, what kind of criteria? (Please attach or provide a short description below)

3. If no to question 2, when grading an essay how do you decide what grade to give it (Even if you do not have anything written, what criteria do you use)?

4. How much weight do you put on content vs. language features? (Please give a rough estimate as a percentage, e.g. 50% content/50% language features)
5. What are the strong language features of a good essay (top 25%)? (Please rank the language features from 1-5)

Rhetoric □ Structure & Org. □ Grammar □
Vocabulary □ Spelling □

Comments: __________________________________________
_________________________________________________
________________________________________

6. What according to your experience are the poor language features of a weak essay (bottom 25%)? (Please rank the language features from 1-5)

Rhetoric □ Structure & Org. □ Grammar □
Vocabulary □ Spelling errors □

Comments: __________________________________________
_________________________________________________
________________________________________

7. Within language how much weight do you put on lexical features/vocabulary? (Please give a rough estimate as a percentage, e.g. 50% grammar/50% vocabulary)

8. In your experience what are the two major vocabulary features of a good essay (top 25%)?

1) __________________________________________
2) __________________________________________

9. In your experience what are the two major vocabulary features of a poor essay (bottom 25%)?

1) __________________________________________
2) __________________________________________
10. Can you have a good essay with poor grammar?
Yes ☐ No ☐
Comments:__________________________________________
___________________________________________________
___________________________________________________

11. Can you have a good essay with poor vocabulary, e.g. low degree of lexical variation and high dependence on high frequency words?
Yes ☐ No ☐
Comments:__________________________________________
___________________________________________________
___________________________________________________

12. Can you have good vocabulary but a weak essay, e.g. a high degree of lexical variation and a low dependence of high frequency words?
Yes ☐ No ☐
Comments:__________________________________________
___________________________________________________
___________________________________________________

13. Is it possible to distinguish the content from vocabulary?
Yes ☐ No ☐
Comments:__________________________________________
___________________________________________________
___________________________________________________

14. To what extent do you think that take-home essays mirror students’ language proficiency considering the risk of plagiarism and the possibility of obtaining extensive aid from friends and relatives?
Thank you for your time!