



Faculty of Arts and Education
English

Pia Sundqvist

Extramural English Matters

Out-of-School English and Its Impact on
Swedish Ninth Graders' Oral Proficiency and Vocabulary

DISSERTATION
Karlstad University Studies
2009:55

Pia Sundqvist

Extramural English Matters

Out-of-School English and Its Impact on
Swedish Ninth Graders' Oral Proficiency and Vocabulary

Pia Sundqvist. *Extramural English Matters - Out-of-School English and Its Impact on Swedish Ninth Graders' Oral Proficiency and Vocabulary.*

DISSERTATION

Karlstad University Studies 2009:55

ISSN 1403-8099

ISBN 978-91-7063-278-5

© The Author

Distribution:

Faculty of Arts and Education

English

SE-651 88 Karlstad

+46 54 700 10 00

www.kau.se

Printed at: Universitetstryckeriet, Karlstad 2009

To my mother and my father

Abstract

The present study examines possible effects of extramural English (EE) on oral proficiency (OP) and vocabulary (VOC). The study is based on data collected from Swedish learners of ESL in grade 9 (aged 15-16; N=80; 36 boys, 44 girls) over a period of one year. EE was defined as linguistic activities that learners engage in outside the classroom in their spare time. EE was measured with the help of a questionnaire and two language diaries, each covering one week. In the diaries, the learners recorded how much time they had spent on seven given EE activities (reading books, reading newspapers/magazines, watching TV, watching films, surfing the Internet, playing video games, listening to music). There was also an open category. Speech data were collected with the help of five interactional speaking tests; learners were in random dyads on each occasion. Each student performance was assessed by three raters with the help of a profile scheme, resulting in an overall grade. Based on these grades from the tests, a mean grade for OP (the OP grade) was calculated for each student. OP was defined as the learner's ability to speak and use the target language in actual communication with an interlocutor. Learners' VOC was measured with an index variable based on the scores on two written vocabulary tests. For a selection of ten learners, additional analyses were made of oral fluency and the use of advanced vocabulary in speech. A mixed methods research design was used, but the lion's share of data was analyzed using inferential statistics.

Results showed that the total amount of time spent on EE correlated positively and significantly ($p < .01$) both with learners' level of OP and size of VOC, but that the correlation between EE and VOC was stronger and more straightforward than the one between EE and OP. The conclusion drawn was that although EE impacts both OP and VOC, the causal relationship is more salient in the case of VOC. Results also showed that some activities were more important than others for OP and VOC respectively; i.e., the type of EE activity mattered. EE activities that required learners to be more productive and rely on their language skills (video games, the Internet, reading) had a greater impact on OP and VOC than activities where learners could remain fairly passive (music, TV, films). An important gender difference was identified. Boys spent significantly more time on productive EE activities than girls; therefore, EE had a greater impact on OP and VOC for boys than for girls. Four background variables were also studied. The conclusion was that EE is an independent variable and a possible path to progress in English for any learner, regardless of his or her socioeconomic background.

Keywords: Extramural English, out-of-school learning, implicit learning, learner language, oral proficiency, oral fluency, second language acquisition, second language learning, ESL, EFL, vocabulary acquisition, assessment, speaking tests, language diary, motivation, self-efficacy, anxiety, speech, oral communication, core vocabulary, peripheral vocabulary.

Acknowledgments

I have learned that completing a thesis is very much like running a marathon or skiing from Sälen to Mora – a very long process with a definite end. Now that I see the finish line, I would like to thank y'all who cheered me on.

I would like to begin by thanking Elisabeth Wennö. If you had not called, I never would have applied; it is as simple as that. My career suddenly took a new turn. Thanks ever so much for believing I could do this.

Next I would like to express my gratitude to my supervisors, who have all helped me, but in different ways. Thanks to Moira Linnarud, who was my main supervisor the first semester, for your open and witty mind, for your great support when I chose my research topic, and, in particular, for your kind words at the CSL conference, 2008. Thanks to my assistant supervisor, June Miliander, for sharing my interest in applied linguistics – and for giving me Benson's book. Thanks to my other assistant supervisor, Hugo Wikström, who joined me about one-third into the race. I admit that your statistics course got me totally hooked. Since then you have taught me so much, things I never knew I would (or could) learn. Thanks for asking me the right questions, for patiently letting me formulate my answers, and for being such a good listener. It was a pleasure all the way. Finally, most of all I would like to thank Solveig Granath, who became my main supervisor when Moira retired. It is very difficult to imagine what would have become of my thesis if you had not been there. So, a million thanks for always giving me valuable feedback (on major as well as on minor matters), for showing me how to approach academic writing, and, in particular, for outstanding support over the past six months.

I also owe many thanks to other colleagues, at the English department and elsewhere: Michael Wherrity, for wise comments which helped improve my text; Maria Holmgren Troy, for insightful feedback; Erica Sandlund, for sharing my interest in spoken English and for giving me encouraging comments (especially “the boring paper revenge”); Marika Kjellén Simes, for maintaining your part of our pact and for surprising me with ice-cold champagne; Julie Skogs, for being my one and only PhD colleague in linguistics and for teaching me hilarious idioms from the land down under; Pilar Álvarez, for moral support (on several occasions); Gunilla Ericsson, for encouragement and guidance; Berith Nordenberg, for booking trips and providing administrative assistance with a smile on your face; and finally, all the members of the faculty seminar for doctoral students, for your comments. To those of you I have not mentioned, be sure that I also heard and appreciated your cheers.

Two colleagues and friends proved to be extra supportive: Marie Tåqvist and Andreas Nyström. To both of you, my deepest thanks, especially for picking me up that one time when I hit rock bottom. In addition, Marie, thanks for happily volunteering to proofread my manuscript, and for outstanding linguistic discussions all the way. And Andreas, you have helped me immensely with all sorts of things: IT-support, photography, the design of my dust jacket, comments on my work, and energy bars. One word: thanks.

Thanks also to Liss Kerstin Sylvén, Gothenburg University. Not only did you provide invaluable help regarding the measuring tools I ended up using in my study, but also, you have been a constant support since the first time we met. I really look forward to our joint research projects in the future. Moreover, I am greatly indebted to Britt Marie Apelgren, Gothenburg University, who was the opponent at my mock viva. I appreciate your thorough reading of my text, and your constructive and very valuable comments.

In preparing my study, several people assisted me on technical matters. I would like to thank Marie Pettersson, who kindly solved computer-related problems that arose, and John Ivan, Christer Mellberg, and Fredrik Tigér, who suggested what camcorder to buy and, then, patiently taught me how to use it. Thanks also to Marjolein Deunk, PhD colleague at the University of Groningen, who introduced me to Transana (as well as to some tricks in Excel), and to Annelie Ekberg Andersson, librarian, who got me started using EndNote, and who kindly helped me whenever I had questions.

Many warmest thanks to you, Lars Lerin, for being so generous, allowing me to use your art on the cover of my thesis. I really, really, really love my cover. Love it.

And, of course, this thesis would never have materialized if it were not for three outstanding teachers, who opened up their classrooms, and eighty equally outstanding students, who willingly let me study their English. Thank you so much for taking part in this project; I owe you greatly.

It is now time to wrap things up. As you might have seen, I dedicate this thesis to my parents. Thank you, *mamma och pappa*, for your love and support, and for being there, always. Finally, I have three wonderful children, Surya, Sixten, and Nirmala; life would be so awfully boring and meaningless without you. Thanks to each one of you, for being exactly who you are. And from the bottom of my heart, thank you Martin, husband and main cheerleader. Can you see the finish line? I think you can, and so can I, and I can't wait to cross it, because then, I will fall into your arms. The best reward there is.

Table of contents

1	Introduction.....	1
1.1	Background.....	2
1.2	Aims and research questions.....	5
1.3	Outline of thesis.....	8
2	Second language acquisition.....	10
2.1	Key concepts and theories of L2 acquisition.....	10
2.2	The role of awareness in language learning.....	14
2.3	Individual difference variables.....	16
2.4	Gender and language learning.....	20
2.5	Cultural capital.....	22
3	Extramural English.....	24
3.1	The out-of-class context and language learning.....	24
3.1.1	Towards a definition of extramural English.....	25
3.1.2	Early out-of-class language learning studies.....	26
3.1.3	The role of English in Europe and Sweden.....	27
3.1.4	Learner self-assessment and retrospection studies on EE.....	30
3.1.5	Other studies on out-of-class learning.....	32
3.2	Empirical studies combining data on EE with learning outcomes.....	33
4	Oral proficiency in learner English.....	39
4.1	On language proficiency and oral proficiency.....	39
4.1.1	Basic types of learner speech.....	41
4.1.2	Characteristics of learner speech and communication.....	41
4.1.3	L2 oral proficiency studies.....	43
4.1.4	L2 oral fluency studies.....	45
4.2	Accommodation theory and the role of the interlocutor.....	49
4.3	Assessment of oral proficiency and oral fluency.....	51
5	Vocabulary in learner English.....	57
5.1	Learner vocabulary in speech and writing.....	57
5.2	Receptive and productive vocabulary.....	59
5.3	Measuring vocabulary.....	62
5.3.1	What is a word?.....	67
5.3.2	Previous empirical studies measuring vocabulary.....	68
5.4	Core and peripheral vocabulary.....	69
5.4.1	The syllable structure of English words.....	71
5.4.2	The etymology of English words.....	72
6	Material and methods.....	75
6.1	Design and procedure.....	75
6.1.1	Pilot studies.....	75
6.1.2	Outline of main study.....	79
6.1.3	Criteria for participation and selection process.....	81
6.1.4	Three schools, four classes – and generalizability of results.....	82
6.1.5	Participants and participant attrition.....	83
6.1.6	Parental contacts.....	85
6.1.7	Ethical considerations.....	85
6.1.8	Mixed methods research design.....	86
6.2	Material.....	88

6.2.1	Questionnaire	88
6.2.2	Language diary.....	89
6.2.3	Speaking tests.....	91
6.2.4	Assessment data	95
6.2.5	Vocabulary tests	97
6.2.6	The Swedish national test of English 2007.....	97
6.2.7	Leaving certificate	98
6.2.8	Interviews	99
6.3	Methods of analysis	99
6.3.1	Group design.....	100
6.3.2	Extramural English.....	102
6.3.3	Oral proficiency.....	103
6.3.4	Vocabulary	106
6.3.5	Background variables, motivation, and views on English.....	109
6.3.6	Statistical methods	110
6.3.6.1	Independent samples t test.....	112
6.3.6.2	Paired samples t test	112
6.3.6.3	ANOVA	112
6.3.6.4	Chi-square	113
6.3.6.5	Correlation analysis.....	113
6.3.6.6	Backward linear regression analysis.....	115
6.3.6.7	Cronbach's alpha and index variables.....	115
7	Extramural English and oral proficiency: Analysis and results.....	116
7.1	Extramural English based on language diary data	116
7.1.1	Results for the four participating classes and gender.....	119
7.1.1.1	A comparison with Swedish and other languages.....	121
7.1.2	Results for subsets based on the EE variable.....	122
7.2	Extramural English based on questionnaire data	124
7.2.1	Media.....	124
7.2.1.1	Media and gender.....	126
7.2.1.2	Media and the EE subsets	127
7.2.2	Reading.....	128
7.2.2.1	Reading and gender	129
7.2.2.2	Reading in English and the EE subsets.....	129
7.2.3	Online role-playing games	130
7.2.3.1	Online role-playing games and gender	131
7.2.3.2	Online role-playing games and the EE subsets.....	131
7.3	Oral proficiency based on assessment data from the speaking tests	132
7.3.1	Results for the four participating classes and gender.....	132
7.3.2	Progression over time for the four classes.....	135
7.3.3	Results for the EE subsets	136
7.3.4	The factorial grade for fluency.....	137
7.3.5	Pauses and fluency.....	137
7.3.6	Understanding the data from a different perspective.....	139
7.4	A comparison of students' extramural English and oral proficiency	140
7.4.1	Correlations between extramural English and oral proficiency.....	141
7.4.2	Oral proficiency viewed through the EE subsets	142
7.4.3	EE viewed through the subsets based on oral proficiency	143
7.4.4	Rank of importance: EE activities in relation to oral proficiency.....	143
7.5	Extramural English and oral proficiency: Summary and conclusion.....	144

8	Extramural English and vocabulary: Analysis and results	147
8.1	Productive Levels Test.....	147
8.1.1	Results for the four participating classes and gender	147
8.1.2	Results for the five EE subsets.....	148
8.2	Vocabulary Levels Test.....	149
8.2.1	Results for the four participating classes and gender	149
8.2.2	Results for the five EE subsets.....	150
8.2.3	EE habits among students with extreme scores	151
8.3	Cloze test: “Working for Change”	153
8.4	The factorial grade for vocabulary	154
8.5	Correlations between extramural English and vocabulary	154
8.5.1	An index variable for vocabulary.....	155
8.6	Rank of importance: EE activities in relation to vocabulary	156
8.7	Polysyllabic vocabulary in the High and Low OP groups in Test 5	157
8.8	Summary and conclusion.....	162
9	Background variables, motivation, and views on English: Analysis and results	165
9.1	Background variables	165
9.1.1	Travels abroad	166
9.1.2	Parents’ educational background.....	168
9.1.3	Number of books in the home	170
9.1.4	Rural versus urban residency.....	172
9.2	Motivational factors.....	175
9.2.1	Self-efficacy	175
9.2.2	Anxiety related to speaking	177
9.2.2.1	Students who speak English regularly in their spare time	179
9.3	Students’ views on English.....	180
9.3.1	Beliefs about where English is learned	180
9.3.2	Self-assessment of work in the school subject English.....	182
9.3.3	Opinions about the school subject English	183
9.4	Correlations between EE/OP and overall grades	184
9.5	Summary and conclusion.....	186
10	Summary and concluding remarks	190
10.1	Brief description of the study.....	190
10.2	Discussion of main findings.....	191
10.2.1	Extramural English among Swedish ninth graders.....	191
10.2.2	Extramural English and oral proficiency	192
10.2.3	Extramural English and vocabulary	195
10.2.4	Extramural English and background variables.....	198
10.2.5	Oral proficiency and background variables	199
10.2.6	Extramural English and motivation.....	200
10.2.7	Oral proficiency and motivation.....	200
10.2.8	Students’ views on English.....	201
10.2.9	Summary of main findings	202
10.3	Implications for classroom practice.....	205
10.4	Concluding remarks and suggestions for further research.....	206
	Summary in Swedish – Sammanfattning på svenska	208
	References	213

Appendices

Appendix 1: Letter to the parents	227
Appendix 2: Form of consent.....	228
Appendix 3: Western Svealand	230
Appendix 4: Questionnaire	231
Appendix 5: Language diary (sample page)	239
Appendix 6: Technical details	240
Appendix 7: Expository speech (Test 4)	241
Appendix 8: Test-instructor's manuscript for Test 4.....	242
Appendix 9: Instructions for the assessment of speaking tests.....	243
Appendix 10: Performance profile scheme	246
Appendix 11: Interrater reliability	248
Appendix 12: Productive Levels Test (page 1).....	250
Appendix 13: Teacher's instructions (VLT)	251
Appendix 14: Vocabulary Levels Test (page 1).....	252
Appendix 15: Figures	253
Appendix 16: Polysyllabic types and their etymology	267
Index	272

List of tables

Table 6.1. Outline of the main study.....	80
Table 6.2. The schools in the study.....	82
Table 6.3. External and internal attrition rates for the four classes and in total.....	84
Table 6.4. An overview of the five speaking tests according to Brown's (2004) taxonomy.....	93
Table 6.5. Transcription notations used in the present study.....	95
Table 6.6. Assessment totals for the four raters.....	95
Table 6.7. Overview of the mixed methods research design.....	100
Table 6.8. Frequency levels of the test items in the Productive Levels Test.....	106
Table 6.9. Frequency levels of the test items in the Vocabulary Levels Test.....	106
Table 6.10. Schematic description of the analysis of students' background, motivation, and views on English.....	110
Table 7.1. Amount of time (hrs/w) spent on the eight EE activities.....	117
Table 7.2. Percentage and number of students with no reported time per extramural English activity.....	118
Table 7.3. EE and the four participating classes; significance values for gender differences.....	120
Table 7.4. Descriptive statistics for the five EE subsets.....	123
Table 7.5. Descriptive statistics for the EE subsets and the eight EE activities.....	124
Table 7.6. OP grades for the four classes on Tests 1-5.....	133
Table 7.7. Improvement over time (paired samples t tests).....	136
Table 7.8. EE subsets and OP grades.....	137
Table 7.9. Pauses and words for the Low and High OP group.....	138
Table 7.10. Time spent on EE for the Low and High OP group.....	139
Table 7.11. Descriptive statistics for the five OP subsets.....	140
Table 7.12. Rank of importance for EE activities in relation to the OP grade.....	144
Table 8.1. Mean scores for the Productive Levels Test for the four participating classes; significance values for gender differences.....	148
Table 8.2. Mean scores for the EE subsets on the Productive Levels Test.....	149
Table 8.3. Mean scores for the Vocabulary Levels Test for the four participating classes; significance values for gender differences.....	150
Table 8.4. Mean scores for the EE subsets on the Vocabulary Levels Test.....	151
Table 8.5. Mean scores for the EE subsets on the Working for Change Test.....	153
Table 8.6. EE subsets and the factorial grade for vocabulary.....	154
Table 8.7. Rank of importance for EE activities in relation to the Voc. Index Variable.....	156
Table 8.8. Tokens, types, and ratios for the Low OP group in Test 5.....	158
Table 8.9. Tokens, types, and ratios for the High OP group in Test 5.....	159
Table 8.10. The students' own polysyllabic types.....	161
Table 9.1. Students' experiences of traveling abroad: EE and OP grade.....	166
Table 9.2. Results for students on five points of measure with cut off points between the sample at 50 and 200 books (case = student).....	171
Table 9.3. Results for six points of measure and residency: rural versus urban.....	173
Table 9.4. EE (hours/week) for boys and girls with regard to residency.....	174
Table 9.5. Mean OP grades for boys and girls with regard to residency.....	174
Table 9.6. Self-efficacy and the five EE subsets.....	176
Table 9.7. Self-efficacy and the five OP subsets.....	177
Table 9.8. Anxiety and the five EE subsets.....	178
Table 9.9. Anxiety and the five OP subsets.....	179
Table 9.10. Mean overall grade for the four classes and in total.....	185
Table 9.11. The five EE subsets and the final grade in English.....	186

List of figures

In text

Figure 5.1. The core-peripheral distribution of the lexicon (based on Minkova & Stockwell, 2006:464).....	70
Figure 7.1. Students' total amount of EE (the EE variable).....	117
Figure 7.12. The five EE subsets and gender distribution.....	123
Figure 8.1. The relationship between polysyllabic types and oral proficiency for the High and Low OP group.....	162

In Appendix 15

Figure 7.2. EE "listening to music" for the four classes and gender.....	253
Figure 7.3. EE "playing video games" for the four classes and gender.....	253
Figure 7.4. EE "watching TV" for the four classes and gender.....	254
Figure 7.5. EE "watching films" for the four classes and gender.....	254
Figure 7.6. EE "surfing the Internet" for the four classes and gender.....	255
Figure 7.7. EE "other activity" for the four classes and gender.....	255
Figure 7.8. EE "reading books" for the four classes and gender.....	256
Figure 7.9. EE "reading newspapers and/or magazines" for the four classes and gender.....	256
Figure 7.10. EE activities for boys as percentages of boys' time spent on EE per week (20.8 hours).....	257
Figure 7.11. EE activities for girls as percentages of girls' time spent on EE per week (16.4 hours).....	257
Figure 7.13. The distribution of the four classes in the five EE subsets.....	258
Figure 7.14. Responses to question 7 (item b: 5), "I read books in English" – gender distribution.....	258
Figure 7.15. Responses to question 15 b, item 6, "I use the computer to play computer games" – the four classes.....	259
Figure 7.16. Responses to question 15 b, item 6, "I use the computer to play computer games" – gender distribution.....	259
Figure 7.17. Responses to question 15 b, item 6, "I use the computer to play computer games" – the five EE subsets.....	260
Figure 7.18. Histogram of the OP grade, with normal curve.....	260
Figure 7.19. OP grade means for the four classes on Tests 1, 3, and 5.....	261
Figure 7.20. Mean OP grade (5 tests) versus mean pause length beyond one second (Test 5) for the Low and High OP group.....	261
Figure 7.21. EE versus mean intra-utterance pause length beyond 1 second, with fit lines for the Low and High OP group.....	262
Figure 7.22. Five OP subsets and gender distribution.....	262
Figure 7.23. A correlation of students' EE and their OP grades, with markers for all students, fit lines for the four classes, and a fit line for the total sample.....	263
Figure 7.24. Mean OP grades and total EE (hours/week) for boys and girls, with added fit lines.....	263
Figure 7.25. The five OP subsets dispersed in the five EE subsets.....	264
Figure 7.26. The correlation between EE and the OP grade for EE subset 1.....	264
Figure 7.27. The five EE subsets dispersed in the five OP subsets.....	265
Figure 9.1. The dispersion of students depending on residency (EE subsets).....	265
Figure 9.2. The dispersion of students depending on residency (OP subsets).....	266

1 Introduction

The present study focuses on Swedish ninth graders' extramural English, oral proficiency, and vocabulary. *Extramural English* (henceforth *EE*) refers to the English learners come in contact with or are involved in outside the walls of the classroom (see section 3.1.1). During my time as an English teacher, I encountered many students whose English was influenced by EE. I will briefly relate three specific cases here. The first is of a student who spoke English with remarkable fluency and a Scottish accent. This was quite surprising because he did not have any Scottish relatives or friends. The reason for his accent turned out to be Sean Connery, the famous actor. The student loved to watch his films and, since Connery is Scottish, the student consciously strove to speak with a Scottish accent. The second example is of a student who read English aloud exceptionally well, with excellent pronunciation and native-like prosody. The student himself thought it could be due to his trips to cousins in New Zealand. It later turned out that he had also acquired a large vocabulary, evidenced by his use of such rare words as *squid* and *hoi polloi*.¹ According to the learner himself, the explanation for his use of such infrequent words could perhaps be that he had an electronic subscription to "Word of the Day" from www.dictionnaire.com. The third and final example is of a student whose written production was grammatically poor but good in terms of vocabulary size and idiomaticity. His speaking skills were also surprisingly good in comparison to his writing skills. This time the explanation turned out to be that the student played online video games with native speakers of English daily; hence he was involved in spoken interaction in English on a regular basis.²

These three non-native speakers of English exemplify learners who engage in activities in English outside school, i.e. EE activities. Their activities included watching films, traveling abroad, using the Internet, and playing video games.³ At least two of them had conscious strategies for improving their English: one by imitating a Scottish accent and one by subscribing to "Word of the Day". Other sources of EE than the ones mentioned above are, for example, music, literature, and TV. Several studies show that Swedish adolescents as well as adolescents from other countries have extensive contacts

¹ The frequency in the BNC corpus for *squid* is 1.53 times per million words and for *hoi polloi* 0.17 (accessed via <http://www.byu.edu>, Nov. 29, 2005).

² All three examples are from boys. This is due to the fact that I mainly taught boys during the latter part of my high school teaching career before I became a doctoral candidate.

³ As suggested by Gee (2003:1-3), the term *video games* is adopted to cover both games played on game platforms (e.g. PlayStation, Nintendo, and Xbox) and those played on computers.

with EE in their spare time (see e.g. Berns, de Bot, & Hasebrink, 2007a; Forsman, 2004; Oscarson & Apelgren, 2005; Skolverket, 2004a; Sylvén, 2004) and it can be assumed that their English is influenced by these activities in one way or another (cf. e.g. Crystal, 2001:237). In fact, at least in Sweden in the 2000s, it is not unusual to hear comments such as “children learn more English outside school than they do in school these days” or “all kids speak English”. The scientific basis for making such claims has, however, yet to be established. Empirical studies which combine data on EE with learning outcomes are very rare (cf. Higgins, 2009:401-402). Likewise, research on oral proficiency among adolescent learners is also limited, at least in comparison with the number of available studies which focus on, for example, writing. This situation is probably due to the fact that it is more arduous to collect, study, and assess oral rather than written productions of language. In comparison with EE and oral proficiency, vocabulary – the third topic of interest in my thesis – is well-researched, both regarding the acquisition of L2 vocabulary (see e.g. Nation, 2001b) and the assessment of L2 vocabulary (see e.g. Daller, Milton, & Treffers-Daller, 2007; Read, 2000).

The current status of research in the field combined with my teaching background led me to the aims and design of the present study. The results of my study should be of interest to teachers of English as a second language and to anyone who is professionally involved in second language acquisition research.

1.1 Background

Students’ access to EE increased dramatically from the mid 1990s with the widespread use of information technology, IT (Modiano, 2005:34). Unfortunately, not all learners had access to EE activities offered by IT due to socioeconomic constraints on their homes (Lamb, 2004:232; Skolverket, 2004b:49-50). In Sweden, one effort by the government to compensate for the problem and to ensure equal opportunities for learners regarding access to IT was the initiative for “IT in Schools”, “ITiS” (Skolverket, 2004b:19). It has been claimed that the emergence of IT and ITiS has had a great impact on schools regarding teaching methodology and learning strategies (Karlsson, 2004:224-225). However, the report *Engelska i åtta europeiska länder – en undersökning av ungdomars kunskaper och uppfattningar* (Skolverket, 2004a) did not find that there had been any major changes in teaching methodology, at least

not in the ESL classroom.⁴ On the contrary, the picture we get is a traditional one, albeit with a stronger focus on communicative skills as compared with earlier curricula. Surprisingly, despite all these new opportunities, students say that they are not actively encouraged by their teachers to speak English outside the classroom (Skolverket, 2004a:50). On the other hand, the syllabus explicitly states that the subject English should make use of the rich and varied exposure to English that children and teenagers encounter outside school (Skolverket, 2000:61).⁵

In spring 2003, there was a national evaluation of the ninth grade in the compulsory school in Sweden.⁶ The evaluation was initiated by the government and the Swedish National Agency for Education was in charge of it. The main purpose was to provide an overall picture of goal attainment (goal attainment by subject and general goal attainment). According to Skolverket (2004b:8), a second purpose was to see what changes had taken place since the previous national evaluation in 1992, and a third was to point out the need for measures.⁷ Sixteen subjects were covered, including English. The results of the English evaluation were presented in a so-called subject report: *Nationella utvärderingen av grundskolan 2003: Engelska* (henceforth *NU-03*) (Oscarson & Apelgren, 2005). In total, approximately 7,000 pupils, 378 teachers, and 120 schools participated in the English subject study (Oscarson & Apelgren, 2005:13). The statistical population was all ninth graders in Sweden (Oscarson & Apelgren, 2005:28). The pupils filled out questionnaires (response rate: 90%) and were also tested in English (participation rate: c. 80%) (Oscarson & Apelgren, 2005:32). There were three questionnaires, two of which were general and one subject specific. The English test was divided into three parts: listening comprehension, reading comprehension, and written production (Oscarson & Apelgren, 2005:36-38). The two comprehension tests contained questions with open answers and were identical to the ones used in the 1992 national evaluation (see Skolverket, 1993). The listening test took 20 minutes and included 13 questions; the time allotted to the reading test was 30 minutes and it contained 11 questions (Oscarson & Apelgren, 2005:37). Both these tests examined students' general understanding as well as comprehension of

⁴ ESL, English as a Second Language (see chapter 2); report title in English: *English in Eight European Countries*.

⁵ In the student questionnaire used in Skolverket's (2004a:110) study, a majority of the students responded negatively to the statement "Läraren uppmuntrar eleverna att prata engelska utanför klassrummet" ("The teacher encourages students to speak English outside the classroom", my translation). However, the conclusion drawn in the report can be questioned because "outside the classroom" is vague. Note that the statement refers to speaking, i.e., it does not refer to other uses of English outside the classroom.

⁶ Fifth grade was also evaluated.

⁷ It is possible to criticize the third purpose of the evaluation. That is, should a need for measures be stated in advance?

specifics. In order to evaluate their writing skills, the written part of the national test was used (Test time 80 minutes; two topics to choose from; no dictionaries allowed) (Oscarson & Apelgren, 2005:38). There were no tests of students' oral proficiency, but the grades from the speaking part of the national test were collected (Oscarson & Apelgren, 2005:30).

From *NU-03* (Oscarson & Apelgren, 2005:82) we learn that students who had above average final grades claimed that they learn as much English outside school as they do in school. In contrast, average and below average students claimed they learn most of their English in school. Moreover, the above average group engaged in more spare time English activities than students in the other groups. The results also showed that a majority of students (more than 80%) considered English to be one of the most important, useful, popular, and interesting subjects in school (Oscarson & Apelgren, 2005:42-43). The importance and popularity of the English subject was incontestable also in *Engelska i åtta europeiska länder* (Skolverket, 2004a:60), where 99% of the girls and 97% of the boys said that it is important to know English. Somewhat surprisingly, half of the students in *NU-03* (Oscarson & Apelgren, 2005:44) said that they find English "rather difficult" or "very difficult", which was a higher percentage than the one found in the national evaluation of 1992. Among those who found English difficult were, not surprisingly, the students who did not meet the grading criteria for pass and, therefore, failed to receive a final grade in English. In comparison with students who actually did receive a final grade in English, these students were considerably less involved in spare time activities in English (Oscarson & Apelgren, 2005:80-83).

Interestingly, the results of the previously mentioned report *Engelska i åtta europeiska länder* (Skolverket, 2004a:62-64) do not quite agree with those of *NU-03* (Oscarson & Apelgren, 2005) regarding the relationship between grades and EE. We have to bear in mind, though, that *NU-03* is a much larger study that based its results on 7,000 students as compared to 1,500 in *Engelska i åtta europeiska länder*. In the latter report, it is pointed out that the group of Swedish students who had yet to receive a grade in English began learning English in school later than their peers. Furthermore, in that particular group, which was rather small (80 individuals; 6% of the informants), a majority of the students were not born in Sweden and/or spoke another language than Swedish or English at home (Skolverket, 2004a:62). Moreover, they were rarely enrolled in another language course (such as French, German, or Spanish). Instead, they were enrolled in "Swedish-English", a class which often is considered an easy

option.⁸ Enrollment in Swedish-English means that the students had in fact more hours of English in school than their peers (Skolverket, 2004a:62). Apparently, these hours did not suffice in order to pass the English course. In addition, the report shows that these students together with the students who had the grade “pass” spent more time on a number of EE activities than their peers with higher grades, for example on watching TV/video, using the computer (video games, the Internet), and listening to music (Skolverket, 2004a:62). This is a result which does not at all agree with those of *NU-03* (Oscarson & Apelgren, 2005). The contradictory results in these two studies indicate that there is no simple explanation for the relationship between EE and grades. That is, students who report a great amount of EE do not automatically also have high grades in English.

It should be mentioned that oral proficiency in English was not tested in *Engelska i åtta europeiska länder* (Skolverket, 2004a). That report covers four other areas of English proficiency: listening comprehension, linguistic accuracy, reading comprehension, and written production. An interesting finding was that the countries which reported the highest amount of EE for their students (Norway and Sweden) also had the highest total scores, whereas Spain and France, where the amount of EE was less, had the lowest scores (Skolverket, 2004a:53-55). In short, the contradictory results in *NU-03* and *Engelska i åtta europeiska länder* present a bewildering picture of the relationship between learners’ EE and learning outcomes that motivates further investigation.

1.2 Aims and research questions

The main aim of the present study is to see whether extramural English has an impact on students’ oral proficiency and vocabulary. Additional aims are to identify and explicate any potential correlations there might be between the variables, i.e., EE and oral proficiency on the one hand, and EE and vocabulary on the other. The study also aims to map out students’ EE. My main research question is as follows: *Does extramural English have an impact on students’ oral proficiency and vocabulary?* I designed an empirical study to answer the question (see section 6.1). Studies are needed in which EE data and data on learning outcomes are correlated for the same group of learners. My sample consists of eighty students in ninth grade (aged 15-16) and it is representative of the part of Sweden referred to as Western Svealand (see App. 3).

⁸ Swe. “*svensk-engelska är ‘slapp’*” (Enkvist, 2005:40).

Regardless of what method is used, it is difficult – if not impossible – to claim that X (extramural English) is the cause of Y (oral proficiency/vocabulary).⁹ However, with empirical evidence, it is possible to evaluate the potential impact of EE on oral proficiency and vocabulary. Furthermore, by adopting a mixed methods research design in which I investigate EE, oral proficiency, and vocabulary in several ways, both quantitatively and qualitatively, and view the results from various perspectives (see e.g. sections 7.4.2–7.4.3), the reliability of my interpretations increases. Thus, my subsequent conclusions about the potential impact of EE on oral proficiency and vocabulary should be valid. The null-hypothesis is, nevertheless, the starting point: EE has no effect on the variables to be examined.

The additional aims of my study, i.e. to identify and explicate any potential correlations there might be between the variables (EE and oral proficiency/vocabulary) and to map out students' EE, follow from the main aim. To achieve those additional aims, a number of other research questions are used. The three questions below are addressed and answered first (chapter 7), in order to pave the way for finally answering the main research question:

- Which EE activities are most common among Swedish ninth graders?
- Do the participating school classes differ in any respect regarding their EE habits? If so, how?
- Do the results indicate any gender differences with regard to EE? If so, what are they?

In addition, before the main research question can be addressed, it is necessary to examine the correlations between (a) EE and oral proficiency and (b) EE and vocabulary. Thus, other research questions in the present study are the following three:

- Is there a correlation between students' EE and their oral proficiency/vocabulary?
- If there is a correlation between students' EE and their oral proficiency/vocabulary, how can it be described?
- Does the amount of time of EE seem to be important? Or is it rather the type of EE that seems to matter? Or both?

⁹ Y might as well be the cause of X, or a third unknown factor might be the cause of both X and Y (see section 6.3.6.5).

Answers regarding EE and oral proficiency are presented in chapter 7, while questions regarding EE and vocabulary are addressed and answered in chapter 8. Chapter 9 is mainly about the results for background variables and motivational factors and their relationship to EE and oral proficiency, but also about the results regarding students' views on English. The research questions for this part are the following:

- What is the relationship between typical background variables (e.g. the educational level of the parents) and EE/oral proficiency?
- What is the relationship between students' motivation and EE/oral proficiency?
- What are the students' views on English? For example, what are their beliefs about where English is mainly learned: in school or outside school?

On the basis of the results presented in chapters 7-9, the main research question is then addressed and answered in the final chapter, i.e. chapter 10.

As was mentioned above, speech data were collected with the help of speaking tests and these tests form the basis for the investigation of *oral proficiency* (henceforth *OP*).¹⁰ In Swedish schools, English is a mandatory subject up until grade 9, which is also the last year of compulsory school. There are two reasons why the level of spoken English produced in ninth grade is particularly interesting to study. First, the level of English produced here is a measure of what students acquired in compulsory school. Second, it is also the starting point for senior high school. For the same reasons, it is interesting to study vocabulary among ninth graders. Both OP and vocabulary are emphasized in Swedish national guidelines and curricula for English (Skolverket, 2000:64-65). Also, the research community seems to agree that vocabulary is at the core of second language acquisition (see e.g. Adolphs & Schmitt, 2004:40; Bergström, 2001:51; Meara, 1992:2-4).

For a selection of ten students in the sample, additional investigations were made of two aspects of oral proficiency: fluency and vocabulary. The question then was whether a measure of oral fluency (pause length) and use of

¹⁰ Four external raters were engaged to make qualitative assessments of the OP of all participants. They took part in a project on assessment of oral proficiency in English financed by the Centre for Research on the Teaching and Learning of Languages and Literature, Karlstad University, and I was given permission to use their assessment work for my thesis.

advanced vocabulary (polysyllabic words) discriminate between learners at different levels of OP.

The lack of studies which combine data on EE with learning outcomes has already been mentioned. The need for further research in this area is stressed in Sylvén (2004:234). One finding in her study was that EE played a crucial role in incidental vocabulary acquisition: “Students who receive English input elsewhere than school are those who score the best, regardless of group. This is especially true of students who read English texts on their own, outside of school” (Sylvén, 2004:224). In her concluding discussion, Sylvén (2004:234) suggests that further research be done on

reading habits as well as TV, movie and Internet habits. Research into how much reading and what types of text are most beneficial for the L2 learner is called for. /.../ This also goes for TV, movie and Internet habits. More information about these types of language input would provide a firm basis for recommendations regarding CLIL as well as traditional EFL teaching.¹¹

My study is an attempt to answer Sylvén’s (2004) – and others’ – call for further research.

1.3 Outline of thesis

This introductory chapter is followed by four background chapters. Chapter 2 discusses key concepts and some background variables in second language acquisition. Chapter 3 deals with extramural English. In the chapter, I define extramural English and discuss the out-of-school context in relation to language learning. Furthermore, previous empirical studies on the topic are accounted for. Chapter 4 discusses oral proficiency in learner English and definitions of language proficiency. In addition, chapter 4 discusses assessment of oral proficiency and presents previous studies on the topic. Characteristics of learner speech are also reported on. Chapter 5 gives a background to how vocabulary in a second language can be investigated and discusses topics such as receptive and productive vocabulary, how to measure vocabulary, and what is meant by core and peripheral vocabulary. Previous studies are also accounted for.

Material and methods are presented in detail in chapter 6. After chapter 6 there are three chapters which present the results of my study. Chapter 7 is the longest of the three. It presents the results with regard to EE and OP. In chapter 8, results are presented regarding EE and vocabulary. Chapter 9 presents the results on background variables, motivation, and students’ views

¹¹ CLIL = Content and Language Integrated Learning. EFL=English as a Foreign Language (see chapter 2).

on English. Background variables and motivational factors are compared with EE and OP, but not with vocabulary as that was beyond the scope of the present study. In the results regarding students' views on English, it was possible to make comparisons with *NU-03* (Oscarson & Apelgren, 2005).

Chapter 10 is the final chapter. In that chapter, I summarize the results and interpret them. Some results from student interviews are included here. Chapter 10 also suggests topics for further research and implications for language teaching practice. A summary in Swedish is provided after chapter 10.

Finally, I would like to make some general comments on practical matters regarding the format of my thesis. I use 31 figures in total; all but four are available in Appendix 15, where they are presented in the order in which they appear in various chapters. Four figures are placed in the text, namely Figures 5.1, 7.1, 7.12, and 8.1. Footnotes are numbered starting with "1" in each chapter.

2 Second language acquisition

The present chapter starts out with a discussion of terminology used in second language acquisition (SLA) of relevance for the present investigation, followed by a brief survey of SLA research. The role of awareness in language learning is discussed next (section 2.2). In section 2.3 there is a discussion on individual difference variables, for example how personality and motivation are related to language learning. Research regarding gender and learning in general is discussed in section 2.4. The closing section is a discussion of the concept “cultural capital”. In addition to a presentation of the concept, section 2.5 includes some findings from previous research related to cultural capital.

2.1 Key concepts and theories of L2 acquisition

Second language acquisition, or *SLA*, is a branch of applied linguistics. According to Ellis and Barkhuizen (2005:3), SLA frequently refers to “the learning of another language (second, third, foreign) after the acquisition of one’s mother tongue is complete. That is, it labels the *object* of inquiry” (italics in original). It is important to realize that “second” may refer to any other language than the learner’s first language. The same authors also say that the term *second language acquisition* is sometimes used to refer to the study of how people learn a second language. In this latter sense, SLA labels the field of enquiry itself. This dual use of terminology might cause confusion. To avoid such confusion, I will use *L2 acquisition* in reference to ‘the learning of a second (third or foreign) language’ and *SLA* in reference to ‘the general field of enquiry within which language learning is studied’ (cf. R. Ellis, 1994a:6; R. Ellis & Barkhuizen, 2005:3; Mitchell & Myles, 2004:5-6).

The language of interest here is English taught in a Swedish context. This type of English is sometimes labeled *English as a foreign language (EFL)* in SLA, which indicates that English is mainly learned in school through instruction and not in out-of-class contacts with native speakers (Viberg, 2000:28). At times a distinction is made between *EFL* and *ESL*, i.e. *English as a second language*, where *ESL* would be English taught to students with another mother tongue than English and in an English context, such as in countries where English is the dominant language of institutions and in the community. However, the distinction is difficult to maintain in the 2000s because in many countries, including Sweden, English is easily accessed outside the classroom even though English is not an official language (Viberg, 2000:28; see also chapter 3). Thus, in line with Ellis and Barkhuizen (2005:3) and others (see e.g. Gass, 1997:ix;

Mitchell & Myles, 2004:5-6), I use “second language” in reference to any language other than the learner’s mother tongue and, as a consequence, I talk about “L2 acquisition/learning” and “ESL” rather than “foreign language acquisition/learning” and “EFL”. According to Mitchell and Myles (2004:6) it is sensible to include “foreign languages” under the more general term “second languages” because the underlying learning processes of “foreign” and “second” languages are essentially the same, despite differing learning purposes and circumstances. Even so, it should be pointed out that from the perspective of learners in Sweden, English might be perceived as a “foreign” language by some, and as a “second” language by others. Finally, the language one aims to learn is generally called the *target language (TL)* and one’s mother tongue the *first language (L1)*.

The terms *learning* and *acquisition* are used interchangeably in my discussion above and, to some, the terms are indeed interchangeable, but to others they are not. There is an ongoing debate in SLA about the distinction between the terms. Krashen (1981:1-3) states that we have two independent ways of developing an L2, either through acquisition (subconsciously, similar to the way a child acquires an L1, picking up the L2 through exposure) or through learning (consciously, through explicit, formal instruction) (see also Krashen, 1987:35-36). Krashen himself admits that it is difficult to test the acquisition/learning hypothesis directly and this is one of the reasons why many researchers choose to use the terms interchangeably (see e.g. R. Ellis, 1994a; Mitchell & Myles, 2004:6). This will also be done in the present thesis. What is more, there is an additional controversy regarding acquisition and learning, namely whether learning can turn into acquisition or not. Krashen asserts that this is impossible (the so-called *non-interface position*) while others disagree with him, claiming, for instance, that older learners might explicitly ask for grammar rules and eventually internalize those rules which, indeed, would be equivalent to Krashen’s *acquisition* (i.e. the interface position, see e.g. N. Ellis, 1994:3-4). Although Krashen’s ideas have been criticized, both regarding the distinction between the two terms and the non-interface position, his contributions were very influential during the last two decades of the 20th century, especially in L2 teaching (Kachru, 2006:249). With the acquisition/learning hypothesis, teachers found that the distinction helps to explain the lack of correspondence between error correction and direct teaching on the one hand, and learners’ accuracy of performance on the other, exemplified by Mitchell and Myles (2004:45) with the teaching of third person singular *-s* in English (*he likes*). They say that learners may have conscious knowledge of the rule but at the same time lack

the ability to apply it in, for instance, spontaneous conversation. The last example, then, would show that the learners have learned, but not acquired, the rule.

As has been noted, the present study focuses on Swedish learners of English and their English. In SLA, such language is commonly referred to as *learner language*, which is simply defined as “the oral or written language produced by learners” (R. Ellis & Barkhuizen, 2005:4). Learner language is, thus, learner output. Often the term *interlanguage* appears in discussions on learner language. The term was coined by Larry Selinker in 1972 (R. Ellis, 1994a:114) and it is used to describe the intermediate states of a continuum that learners move along from having zero knowledge of an L2 to (ideally) native-like proficiency. Interlanguage is said to be dynamic and to have its own internal structure (Mitchell & Myles, 2004:137). Bardovi-Harlig (2006:69) stresses that the study of interlanguage development is crucial because “without an understanding of language development we are unable to specify what the influence factors [for L2 acquisition] may be except in the broadest strokes” (cf. Færch & Kasper, 1983:21).

Not only *interlanguage* but also the terms *competence* and *performance* are commonly discussed concurrently with learner language (see e.g. R. Ellis, 1994a:155-156; Hedge, 2000:45; Lightbown & Spada, 2006:36; Mitchell & Myles, 2004:10-12; Tornberg, 2005:40). The terms *competence* and *performance* were first introduced by Noam Chomsky (see e.g. Chomsky, 1965:4). In generative theory, *competence* refers to a speaker’s internalized knowledge of his or her grammar, whereas *performance* refers to the way in which speakers put that knowledge to use in speaking and understanding a language (Poole, 2005:55). It should perhaps be pointed out that in post-Chomskyan literature, some maintain that there is great confusion regarding the meaning and use of his terminology, in particular regarding competence. Freed (1995:124), for instance, mentions how the three terms *language ability*, *communicative competence*, and *language proficiency* are often used interchangeably and thus subject to multiple and overlapping definitions. In addition, Taylor (1988:149) argues that Chomsky never included *ability* in competence; this was rather an aspect which Dell Hymes added when he suggested the concept of communicative competence in the early 1970s: “...he [i.e. Hymes] has extended the notion of competence in subtle and not always recognized ways, so that it conveys something very different from what was originally intended by Chomsky” (Taylor, 1988:156). Since Hymes’ concept of communicative competence was introduced, Taylor continues, “different conceptions of competence and the distinction between

competence and performance reflect different aims and interests and it is impossible to make comparisons without taking this into account” (Taylor, 1988:157). I have now briefly presented some of the criticism of the (mis)interpretations of Chomsky’s terms; another “misinterpretation” might be that his terminology is often used in SLA research, but Chomsky himself has never written about SLA (Noam Chomsky, email communication, Nov. 21, 2007). I believe that the potential confusion of terms that Taylor fears, supposedly caused by an avoidance of explicitly addressing, for instance, the concept of competence in scientific discourse, is well-founded. Nevertheless, as Taylor (1988:161) himself says, in the domain of current applied linguistics and SLA research, competence has come to be firmly associated with *proficiency*. Thus, when we speak of *competence* in the year 2009, it does include the notion of *ability*.

Considering the fact that learners differ greatly in their ability to learn a second language (for a vivid description, see Reinders, 2006:1-2), how can L2 acquisition be explained? Several suggestions have been put forth over the years. In the 1980s, Stephen Krashen presented five hypotheses as explanations (Krashen, 1981, 1982, 1985, 1987). First, there is the acquisition-learning hypothesis (discussed above). Second, there is the monitor hypothesis, which suggests that knowledge of L2 rules only helps the learner supplement what has already been acquired. Teaching should therefore focus on creating conditions for acquisition rather than learning. Third, the natural order hypothesis suggests that all learners acquire the TL in the same order but at a different pace (cf. Lightbown & Spada, 2006:82-96). The order of acquisition is independent of the learner’s age, L1 background, and conditions of exposure to the TL. Fourth, the input hypothesis, according to Mitchell and Myles (2004:49), is probably Krashen’s greatest contribution to SLA since it “stimulated a tradition of theorizing and empirical research on input and interaction”. The input hypothesis states that learners acquire L2 by exposure to comprehensible input. The learner makes progress along the natural order as long as there is input which is one step beyond (+1) the learner’s current stage of linguistic competence (*i*); comprehensible input is, thus, according to the formula $i + 1$. Fifth, the affective filter hypothesis states that learners with high motivation, good self-confidence, and a low level of anxiety are better equipped for L2 acquisition than those who have low motivation and low self-esteem, for example. In the latter case, with the affective filter up so to speak, language acquisition is impeded. Krashen has been criticized for his model of second

language acquisition since the hypotheses are untestable, i.e., his explanation lacks empirical validity. Nevertheless, his terminology is commonly used.

Following Krashen, Long (1981) suggested that interaction is a key element in L2 acquisition and he proposed what he referred to as the interaction hypothesis. He emphasized that TL input needs to be comprehensible to the learner and this is obtained via modifications during actual interaction. Later, Long (1987:349-350) said that this is something which is mainly achieved through the negotiation for meaning, where negotiation for (or of) meaning refers to speakers' adjustments of their speech (or other techniques) to avoid breakdown in communication, as described by Lightbown and Spada (2006:203). In the mid 1990s, Swain (1995) proposed the comprehensible output hypothesis which emphasized the importance of output for L2 acquisition. In her discussion of L2 acquisition, she stresses the importance of forcing learners to produce output. She later proposed the collaborative dialogue in *The output hypothesis and beyond* (Swain, 2000). The collaborative dialogue is described as knowledge-building dialogue, or dialogue that constructs linguistic knowledge. She says it is "what allows performance to outstrip competence" (Swain, 2000:97), where language use and language learning can co-occur, and where language use mediates language learning. She claims it is both a cognitive and a social activity. There are several empirical studies which support the output hypothesis (see e.g. R. Ellis & He, 1999; de la Fuente, 2002; Joe, 1995).

The following quote from Gass and Mackey (2006:3-4) sums up the interaction approach (or the interaction hypothesis)¹ very well:

the interaction approach considers exposure to language (input), production of language (output), and feedback on production (through interaction) as constructs that are important for understanding how second language learning takes place.

2.2 The role of awareness in language learning

The role of awareness in language learning is linked to "the interface issue" (R. Ellis, 2009:20), i.e., the extent to which learning interfaces with acquisition (see above). Thus, the role of awareness is related to *implicit* (cf. Krashen's "acquisition") and *explicit* (cf. Krashen's "learning") L2 learning. Furthermore, to a certain degree, awareness in L2 acquisition and implicit/explicit learning are connected with out-of-school learning, i.e., with extramural English (cf.

¹ The interaction hypothesis is very close to Lev Vygotsky's theories about language learning, including the zone of proximal development (see e.g. Vygotskij, 1999).

chapter 1 and section 3.1.1). Nick Ellis (1994:1-2) describes the two concepts, implicit and explicit learning, as follows:

Implicit learning is acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply and without conscious operations. Explicit learning is a more conscious operation where the individual makes and tests hypotheses in a search for structure. Knowledge attainment can thus take place implicitly (...), explicitly through selective learning (...), or, because we can communicate using language, explicitly via given rules (...).

Schmidt (1990) went to great lengths to examine what role implicit and explicit learning has in L2 acquisition while investigating his own learning of Brazilian Portuguese. One question he discusses is whether conscious awareness at the level of “noticing” is necessary for L2 acquisition (Schmidt, 1990:132). *Noticing* can be defined as ‘paying attention to specific linguistic features in the input’ (R. Ellis, 1994b:93) or as ‘conscious attention to the form of input’ (Robinson, 1995:284). For example, when reading, people are aware of (“notice”) the content of what they read, rather than the background noise outside or the music playing on a radio in the next room (Schmidt, 1990:132). Generally, learners are able to give verbal report about noticing, but noticing is also possible without learners’ verbal report because conscious experiences can be very difficult to describe (Schmidt, 1990:132). For example, learners may recognize a regional accent but be unable to describe it, just as people may have problems describing the difference between two wines even though they can taste it (Schmidt, 1990:132). In his paper, Schmidt (1990:133) also points to the importance of maintaining a distinction between “consciousness as awareness” and “consciousness as intentional behavior”. He argues that intentions may be either conscious or unconscious and people often become aware of things they did not intend to notice (Schmidt, 1990:133).

Schmidt (1990:149) concludes that subliminal learning is impossible. He claims that in order for learners to convert input to intake, noticing is necessary; in fact, he defines *intake* as ‘that part of the input that the learner notices’ (Schmidt, 1990:139).² Furthermore, he concludes that “incidental learning is certainly possible when task demands focus attention on relevant features of the input” (Schmidt, 1990:149). However, at least among adult L2 learners, incidental learning appears highly unlikely in the sense that learners would pick up TL forms from input when the forms do not carry information which is crucial to the task (Schmidt, 1990:149). Schmidt (1990:149) also considers

² The concept of *intake* has been defined in several ways in SLA. Schmidt (1990:139) provides a brief overview.

implicit learning possible. It is best characterized by a “gradual accumulation of associations between frequently co-occurring features” rather than by an unconscious induction of abstract rule systems. An important finding in his study was that memory requires attention and awareness. This finding agrees with Bowers (1984:236-237), who suggests that unattended material makes it into short-term memory but, if there is no opportunity for learners to selectively attend to it and notice it, the material cannot be processed into long term memory. Robinson (1995:319) argues in the same way: noticing takes place in short-term memory.

These findings are interesting, for example in light of the various EE activities that learners may be involved in, where, in order to be meaningful for the learner, some activities require more attention (on the part of the learner) than others do. Reading, for example, is pointless without attention. Watching TV, however, may serve its purpose even though the viewer’s attention might be low, such as when the purpose of watching TV is to relax.

2.3 Individual difference variables

Having discussed some key concepts in SLA and theories of L2 acquisition, I will now address individual differences among learners. The success of learners’ L2 acquisition depends on a number of factors, for example learners’ personality, aptitude, and motivation. Factors such as these are referred to as *individual difference variables* (or *ID variables*) in psychology research, and they are often prominent features of L2 acquisition (Dörnyei, 2006:42). There are many cases of SLA research where ID variables are discussed (see e.g. Hedge, 2000:16-24; Rubin, 2005:46-50; Wenden, 1987:107-109). In the present section I discuss ID variables such as personality and personality traits (e.g. shyness and anxiety), aptitude, motivation (in general) and self-efficacy (in particular), and their relevance to L2 acquisition.³

Regarding personality, it is generally argued that those who are more extrovert are better suited to language learning than those who are more introvert, but this is not always the case (Cohen, 1991:112). It seems to be valid only for certain aspects of L2 acquisition. Lightbown and Spada (2006:62), for instance, say that the positive relationship between a more extrovert personality and L2 acquisition holds for oral communicative skills but not for the acquisition of literacy or academic skills. The rationale behind this argument is that learners who are less inhibited (i.e. who are more extrovert) dare to take

³ For a summary of studies on various IDs, see R. Ellis (1994a:526-527).

risks, something which is necessary in the L2 acquisition of oral communicative skills. Dewaele (2004:148) argues in a similar fashion regarding the acquisition of colloquial vocabulary. In his study of Dutch L1 speakers, who were learners of French, Dewaele (2004) found that learners' degree of extroversion affected their use of colloquial words in oral communication. That is, those who were more extrovert used such words more freely, which contributed positively to their acquisition of French.

Shyness is a personality trait whose reactions are often triggered by social interactions. It is closely related to extroversion/introversion and may be divided into two components: the instrumental and the emotional (Buss, 1986:65). The instrumental component of shyness consists mainly of the relative absence of instrumental activity (e.g. withdrawal, reticence, inhibition of speech and gestures). The emotional component of shyness may be accompanied by an arousal of the autonomic nervous system (e.g. feelings of fear, vulnerability, drop of self-esteem) (Buss, 1986:65).

In a study on another ID variable, namely anxiety, and L2 acquisition (Xiu Yan & Horwitz, 2008), 21 learners of English (first year university students, China) representing a range of anxiety levels, were interviewed in order to study their perceptions of the role of anxiety in their own learning of English. The sample is very different from the one used in the present study but the results are interesting nevertheless. Among the most influential sources for anxiety was the variable "comparison with peers" (Xiu Yan & Horwitz, 2008:173). The fact that "comparison with peers" was considered highly influential is evidence of the importance of, for example, seating arrangement in classrooms and the set-up of peers in test situations. Another study on anxiety found that spending time abroad enhanced learners' self-confidence in speaking English, their motivation, and ultimately their grades (Matsuda & Gobel, 2004:32). A result from that study indicates that the English teacher's character and psychological state can affect classroom atmosphere. Thus, it is essential that teachers work to improve the classroom climate when necessary and learn to remain relaxed, in order for their students to be relaxed as well (Matsuda & Gobel, 2004:32). Time spent abroad was found to reduce anxiety also in an American study carried out among university level students (Onwuegbuzie, Bailey, & Daley, 1999). Finally, Pappamihel (2002) found that different types of English language anxiety seemed to be at work among the informants in her study (178 middle-school Mexican immigrants, the US). The type of anxiety was dependent upon the context of the interaction. Fears of negative evaluation, test anxiety, communication apprehensions, and identity issues reduced the informants'

feelings of self-efficacy (Pappamihel, 2002:348-349; also see below). Pappamihel (2002:348) also found a gender difference in the main-stream classes in the study: girls tended to be more anxious than boys. Finally, the most common strategy among the learners to reduce anxiety was avoidance.

Language anxiety can be said to be more of a temporary state (i.e., a moment-to-moment experience of anxiety) than a trait (i.e., a permanent predisposition to become nervous in a wide range of situations that relates to an individual's personality) (Marcos-Llinás & Juan Garau, 2009:96). Gardner (1993:5) defines language anxiety as 'the apprehension experienced when a situation requires the use of a second language with which the individual is not fully proficient'. It is, for instance, characterized by derogatory self-related feelings ("I can't do this") and sometimes with physiological responses (e.g. increased heart rate, blushing). Language anxiety may influence learners in interaction and test situations, for example. Thus, it affects L2 acquisition.

Regarding another ID variable, aptitude, Stansfield and Reed (2004:43) argue that it is closely associated with John B. Carroll, who was the leading developer of the Modern Language Aptitude Test (MLAT). For more than twenty-five years, Carroll pursued the notion that such a thing as aptitude for learning second languages does indeed exist (Carroll, 1981:83). What is more, he claimed that the concept of aptitude for language learning is relevant for any language (foreign, second, third etc.) provided that the language at hand is learned in some sort of instructional context.⁴ Carroll suggested that aptitude can be subdivided into four components.⁵ The four sub-components, as presented and described by Skehan (2002:71), are (1) phonemic coding ability, (2) grammatical sensitivity, (3) inductive language learning ability, and (4) associative memory. The phonemic ability refers to the capacity "to code unfamiliar sound so that it can be retained" (Skehan, 2002:71), i.e., the capacity to impose structure on unfamiliar sounds so that they become memorable and less transitory. The second sub-component, grammatical sensitivity, refers to the capacity to identify the functions of words in sentences. The third, inductive language learning ability, is concerned with learners' capacity to extrapolate from a corpus to create new sentences. In L2 acquisition it would, thus, refer to the capacity to understand grammatical rules from samples of the TL. Finally, associative memory corresponds to the capacity to form links in one's memory, or, to be more specific, the capacity to make associations between lexical items in one's L1 and the TL. In addition, Skehan (2002:80) suggests that there is a

⁴ Carroll spoke of aptitude for learning languages other than one's L1 (Carroll, 1981:83).

⁵ Skehan (2002:92) argues that this traditional view is still relevant to SLA.

relationship between variation in the speed of L1 learning ability and L2 aptitude.

In *The Good Language Learner*, Naiman, Fröhlich, Stern, and Todesco (1996:217) conclude that

perceived (but not measured) aptitude was less of a crucial factor than attitude to language learning, persistence, and willingness to adapt to varied learning situations over prolonged periods of time (...).

What is described in the quotation is basically the importance of motivation for language learning, i.e. another ID variable. McDonough (2007:369) suggests that motivation involves four elements: (1) the reasons why we want to learn, (2) the strength of our desire to learn, (3) the kind of person we are, and (4) the task, and our estimation of what it requires of us. Furthermore, motivation can be either *instrumental* (language learning for immediate or practical goals) or *integrative* (learning for personal growth) and research has shown that both types are related to successful L2 acquisition (Lightbown & Spada, 2006:63-64).⁶ Scholars are in agreement here. The importance of motivation in L2 acquisition is apparent not only in the quotation from *The Good Language Learner* above (only learners who are motivated are likely to show persistence, for instance), but also in other studies of L2 acquisition (see e.g. Horwitz, 1999; Rubin, 2005).

It was previously mentioned that fears of negative evaluation, test anxiety and so forth reduced learners' self-efficacy (Pappamihel, 2002). *Self-efficacy* is closely related to motivation and it refers to people's judgment of their own ability to carry out specific tasks (Dörnyei, 2001:22-23). It has been shown that learners' self-efficacy is a good predictor of their subsequent grades in English (Schunk, Pintrich, & Meece, 2008:54). Thus, learners who are motivated and believe in themselves in general achieve high values on measures of self-efficacy.

On the whole, previous research suggests that there is a relationship between learners' ID variables on the one hand, and success in L2 acquisition on the other. But, to quote Lightbown and Spada (2006:63), this "relationship is an intricate one (...) in that it is probably not personality alone, but the way in which it combines with other factors, that contributes to second language learning". As has been noted, there is no simple way of explaining successful L2 acquisition and it would be a mistake to claim that there was (cf. Gass, 2006:209). Learners differ in their ability to acquire an L2, both in terms of rate and ultimate attainment (Gass & Mackey, 2006:14).

⁶ The term *intrinsic motivation* is sometimes used in a similar way as *integrative motivation* (cf. R. Ellis, 1994a:515-516).

2.4 Gender and language learning

Gender differences in language learning have caught the attention of researchers in recent years. In schools, girls generally do better in languages than boys (see e.g. Björnsson, 2005; Carr & Pauwels, 2006; Klapp Lekholm, 2008) and in language studies in general, gender tends to produce significant differences in results (Dörnyei, 2008). Various explanations have been put forth as to why this is the case. Drawing on results from neural science research which point to sex differences in the brain (see e.g. Gorski, 2000), innate differences between genders have been put forward to explain why boys lag behind girls in school, for example, that a majority of women's brains develop at a faster pace than men's and, therefore, girls mature earlier than boys (see e.g. Dahlström, 2007; Sax, 2007). Because of differences such as these, Dahlström (2007:137) proposes separate classes for boys and girls in the teaching of foreign languages, and Sax (2007) argues that a combination of social and biological factors creates an environment – at home and in school – which is generally bad and demotivating for boys. In short, they pinpoint the importance of considering results based on biological differences as an explanation for gender differences in scholastic achievement. However, others have shown that such gender differences in scholastic achievement can be explained by other factors, see e.g. Klapp Lekholm (2008) below, and regarding learners' ability to self-assess their own language learning, Oscarson (1998:151) did not find a gender-related difference.

An empirical study based on data from adolescent learners in Australia, England, Wales, New Zealand, and Scotland (Carr & Pauwels, 2006) is relevant for the discussion of gender and language learning in general, as well as to the more specific discussion of innate gender-related differences in learning. For example, in the study, societal perceptions of what is considered appropriate male behavior had an effect on boys' motivation for and participation in second language courses. Fear of being different and not like "real boys" discouraged boys from studying languages. In addition, the teaching methodology used in the language classrooms often appealed more to girls than to boys; it suited girls better (Carr & Pauwels, 2006:202). Carr and Pauwels' (2006) conclusion was that boys are not as motivated for language learning as girls. This finding is probably generalizable also to a Swedish context, at least as indicated by statistics. For instance, girls in Sweden are enrolled in so-called B-language courses (i.e. the studying of an L2 in addition to English, which is a mandatory

language)⁷ to a higher extent than boys, and girls have higher grades than boys in both Swedish and English.⁸ To remedy gender-related differences observed in L2 acquisition, Carr and Pauwels (2006:202-203) suggest a major change in L2 pedagogy.

In light of what has been described above, it is very interesting to see that Klapp Lekholm (2008), in her study on the effects of student and school characteristics on grades and grade assignment (Swedish, English, Mathematics), found that girls were more motivated than boys for schoolwork and learning in general. This general interest in learning and motivation was found to mediate a large part of the difference in the language domains (Swedish, English), which, according to Klapp Lekholm (2008:77), means that girls achieve higher grades than boys due to their greater interest and motivation. Girls' higher achievement in school may be due to their coming to school better equipped for educational activities and the fact that "they have developed a learning approach that is more closely related to school practices and which is better rewarded by teachers in the grade setting practices" (Klapp Lekholm, 2008:86). Such a learning approach is culturally valued both by teachers and the curriculum, Klapp Lekholm (2008:86) argues. Her argument is empirically supported by Nycander (2006).

Nycander (2006) compared the final grades in all subjects and the overall grade on the national tests for the three core subjects in Sweden (English, Swedish, Mathematics), in a statistical study spanning over the period 1998-2005. Regarding English, when comparing grades on the national test with students' final grades, she found that the girls' grades improved to a higher extent, or were less frequently lowered, in comparison with the boys' (Nycander, 2006:11, 19). Furthermore, she found that for the national test in English, the difference in grades between boys and girls decreased year by year in favor of the boys. This might be due to changes in the test format during the time of the investigation, namely changes which aimed at minimizing gender-biased questions (Nycander, 2006:19), but there might be other reasons as well. Similarly, the gap between boys' and girls' final grades in English decreased during the time of the investigation (Nycander, 2006:11). These findings seem to indicate that Swedish boys are catching up with the girls in the subject English. In fact, English was the only subject in school where this happened, at

⁷ According to statistics from the National Agency for Education, <http://www.skolverket.se/sb/d/1636/a/15936#paragraphAnchor1> (accessed July 29, 2009).

⁸ According to statistics from the National Agency for Education, <http://www.skolverket.se/sb/d/1637> (accessed July 29, 2009).

least according to Nycander's (2006) study. In comparison, in the subject Swedish, there was a big difference between boys' and girls' grades, both regarding their grades on the national test and the final grade. This difference remained stable during the period investigated, i.e., the girls outperformed the boys. In Mathematics, there was no gender-related grade difference on the national test and the difference in final grades was negligible (Nycander, 2006:11).

2.5 Cultural capital

Cultural capital, a term coined by the French sociologist and writer Pierre Bourdieu (1973), is a sociological concept which refers to the education, knowledge, skills, and advantages which a person has that give that person a higher status in society. Cultural capital is provided to a child by its parents, who transmit the attitudes and knowledge needed in order for the child to succeed in the educational system, for example by the use of appropriate language and suitable behavior (Bourdieu, 1973:80, 84). The educational system requires familiarity with the dominant culture and only those who are endowed with "cultural capital" can receive and acquire the information and training offered by the educational system. In support of his arguments, Bourdieu (1973) used official statistics on a wide range of topics, such as various cultural activities among the French people, for example reading habits, going to museums, going to the theater, going to exhibitions, visiting art galleries, and listening to classical music. Different indices from statistics were also used: possession of radio/TV, educational background, religious background, and residency, only to mention some (Bourdieu, 1973:87-95).

In line with Bourdieu's argument, it is common to include background variables which tap into the concept of cultural capital in studies that are carried out on learner achievement in school. For example, the number of books in the learner's home is one variable that has been used as a measure of learners' cultural capital (see e.g. Öquist & Wikström, 2006). Others are learners' residency, for example rural versus urban residency (see e.g. Forsman, 2004; Klapp Lekholm, 2008), the educational level of the parents (see e.g. Klapp Lekholm, 2008; Öquist & Wikström, 2006), and experience of travels abroad (see e.g. Sundh, 2003; Sylvén, 2004). Regarding residency and the educational level of parents, when the latter variable was controlled for, Klapp Lekholm (2008:77) found that the proportion of parents who had an education higher

than upper secondary school was lower for students in rural schools than it was for students in suburban and urban schools.⁹ This is in line with Bourdieu's (1973:100) suggestions. In the present study, all four investigated background variables (sections 9.1.1-9.1.4) can be said to relate to the students' "cultural capital" to some extent, in particular one of them: the number of books in the home (section 9.1.3).

⁹ Schools in rural areas were situated more than 45 minutes from the nearest town with 3,000 inhabitants. Schools in suburban areas were 5-45 minutes from the nearest town (3,000 inhabitants); in urban areas there was a maximum of five minutes travel time to a town with more than 3,000 inhabitants (Klapp Lekholm, 2008:64-65).

3 Extramural English

The present chapter is about extramural English. The term *extramural* is an adjectival compound of Latin origin where the prefix, *extra*, means ‘outside’ and the stem, *mural*, means ‘wall’.¹ Hence, the term *extramural English* means ‘English outside the walls’ and, as I said in chapter 1, it refers to the English that learners come in contact with or are involved in outside the walls of the classroom. The present chapter has two sections. The first section (3.1) is about the out-of-class context and provides an account of relevant terminology and theory. I define *extramural English* and discuss the concept along with other terms used within this field of research. Findings from previous research are also presented with regard to language learning outside school. Moreover, the role of English is discussed from a general European perspective, and, more specifically, from a Swedish perspective. Also, English is discussed in the light of young people’s media habits. The second section (3.2) focuses on previous empirical studies which combine data on EE with learning outcomes.

3.1 The out-of-class context and language learning

According to Benson (2001:62), *out-of-class learning* refers to “any kind of learning that takes place outside the classroom and involves self-instruction, naturalistic learning or self-directed naturalistic learning.” He explains that *self-instruction* can be understood as a deliberate long-term learning project which is instigated, planned, and carried out by the learner without any intervention from a teacher, or as any deliberate effort by the learner to acquire language content or skills. *Naturalistic learning*, on the other hand, occurs through direct spoken interaction with users of the target language or through interaction with target language texts. Benson (2001:62) adds that the learners’ degree of deliberate intention to acquire language content or skills at the time of the learning event itself is what makes self-instruction different from naturalistic learning. He also introduces the concept of *self-directed naturalistic learning* in reference to situations in which learners create a naturalistic learning situation for themselves with the advance intention of learning the language. However, at the time of the learning event, learners’ focus of attention might be on communication or on learning something other than the target language. Finally, Benson (2001:62) concludes that most language learning research (as of the year 2001, when the book was published) had focused on learning in the

¹ According to the *Oxford English Dictionary Online* (accessed June 11, 2009).

classroom, whereas research on out-of-class learning was a relatively new field of research. It is within this context I place my study and just as with Benson's research, my study is linked to the theory of *learner autonomy*, originally defined by Holec (1981:3) as "the ability to take charge of one's own learning."

3.1.1 Towards a definition of extramural English

I choose to introduce and use the term *extramural English (EE)* in the present work. Originally, Moira Linnarud, professor emerita, Karlstad University, suggested that I use the term *extramural input in English*, which I also did for some time. Then, inspired by Sylvén's (2006) terminology, I changed it into *extramural exposure to input in English*. However, both these terms which I previously used failed to include the concept of output. Furthermore, I realized that *exposure* often has negative collocations and implies passivity. Thus, I eventually decided on *extramural English*, a term which covers aspects of both input and output and, hopefully, a term which is free of negative connotations and collocations. It is closely related to Benson's (2001) *out-of-class learning* (of English), and to *out-of-school learning* (of English) which is yet another term that is used for the same phenomenon (see e.g. Lamb, 2004; Yi, 2005). Extramural English is also related to the more colloquial expression *spare time English* (Swe. *fritidsengelska*) (see e.g. Lundahl, 2009:37).

Extramural English resembles Benson's (2001) description of self-directed naturalistic learning but there is an important difference. In extramural English, no degree of deliberate intention to acquire English is necessary on the part of the learner, even though deliberate intention is by no means excluded from the concept. But what is important is that the learner comes in contact with or is involved in English outside the walls of the English classroom. This contact or involvement may be due to the learner's deliberate (thus conscious) intent to create situations for learning English, but it may equally well be due to any other reason the learner may have. In fact, the learner might not even have a reason for coming in contact with or becoming involved in extramural English. For instance, a sudden encounter with a foreigner in the street, which leads to a conversation in English between the learner and the foreigner, is an example of an extramural English activity that the learner did not even know about before it actually happened. In her study on Finland-Swedish² EFL learners, Forsman (2004:173) uses the term *unintentional learning* in reference to "accidental learning

² *Finland-Swedish* refers to Finns with Swedish as a first language. Both Finnish and Swedish are official languages in Finland. Swedish is spoken by approximately 5.5% of the population (<http://www.ne.se/lang/finlandssvenska>, accessed July 2, 2009).

of information without the intention of remembering that information.” Forsman’s definition of unintentional learning is thus close to my definition of extramural English, but the exclusion of learners’ intention from her definition separates it from mine.³ Finally, I would like to add that contact with extramural English, or involvement in extramural English activities, is generally voluntary on the part of the learner. However, there is also the possibility that learners engage in specific extramural English activities because they feel pressured to do so, for example by their peers or parents. In sum, extramural English is broadly defined and may therefore function as an umbrella term for the other terms discussed above.

I mentioned in the previous section that my study is linked to the theory of learner autonomy. There are learners who come in contact with EE or become involved in EE activities because they want to improve their English; i.e., they take charge of their own L2 learning. In this respect, my study and EE is linked to the theory of learner autonomy. It is also possible that learners develop an interest in learning English via EE activities, an interest they might not have had to begin with.

3.1.2 Early out-of-class language learning studies

An early field of research pertinent to extramural English concerned the value of functional practice in language learning (Bialystok, 1981). Bialystok’s 1981 study is based on questionnaire data collected from 157 learners of French as a second language (aged 14-15 and 16-17). She points out the importance of exposure to the target language in any form because such exposure increases learners’ competence (Bialystok, 1981:25), and argues that the most functional situations are likely to occur outside the classroom. An important conclusion of her study was that functional practice proved to be critical to achievement on all tasks that were investigated (Bialystok, 1981:34). A weakness of her results is that they are based solely on questionnaire data, but her findings are intriguing nevertheless, because they pinpoint the importance of teaching students out-of-class learning strategies and the role of conscious strategies in L2 proficiency (cf. explicit/implicit language learning, discussed in chapter 2). The role of consciousness/awareness in L2 learning is also addressed by Schmidt

³ It should be mentioned that Forsman (2004) has an interesting in-depth discussion on terminology in chapter 2 of her thesis. Moreover, she points out the impossibility of determining in what way (i.e. with or without intention) students’ lexical knowledge (as tested in her study) was really learned. She also admits that the definitions and applications of the term *unintentional learning* “are not always clear-cut” (Forsman, 2004:84).

(1990:150), who emphasizes the need for further research on learner self-assessment and learner beliefs with regard to L2 acquisition.

Learner self-assessment has been investigated by, for example, Oscarson (1998:151), who found that learners are able to assess their own language proficiency in a reliable way. As for learner beliefs, Wenden (1987:112) shows that learners do think about the process of language learning and also that they are able to articulate their beliefs. On the same topic, Horwitz (1987:120) developed a questionnaire (The Beliefs About Language Learning Inventory, BALLI) which surveyed learners' beliefs about language learning.⁴ She used the BALLI questionnaire in several empirical studies and, among other things, found that a large proportion of language learning goes on outside of the classroom and is not subject to teachers' intervention. In the following I address the out-of-school context of English mainly from a European perspective, with a special focus on English in Sweden. The role of English in the media is also included in the discussion.

3.1.3 The role of English in Europe and Sweden

In 2009, English functions as a global *lingua franca*, i.e., it is the means of communication between speakers who have different first languages in various settings around the world (Berns, 2007:5). The spread of the English language in Europe belies its comparatively limited influence there until the 20th century (Berns, Claes, de Bot, Evers, Hasebrink, Huibregtse, Truchot, & van der Wijst, 2007b:17). In the 20th century, English grew increasingly important. This was not only due to British influence in politics. The role of the U.S. in ending World War I, for instance, and its standing as a world power, were highly important as well. The influx of American pop music (alongside British pop music) in the 1960s further strengthened the hold of English in Europe, as did the widespread use of English in science (Berns et al., 2007b:17-18). The use of English among adolescents in various countries in Europe is described in Berns, de Bot, and Hasebrink (2007a), with a special focus on media and youth in four countries: Belgium, France, Germany, and the Netherlands. By and large, this book provides a good picture of the status of English in Europe in the early 2000s.

ESL became part of secondary education in Europe after the 1950s and, eventually, also of primary education (Berns et al., 2007b:18). In Sweden, English has been mandatory subject from grade 4 in compulsory school since

⁴ An adapted version of BALLI was used in my pilot study (see section 6.1.1).

1962 (Malmberg, 2000a:8). In the 2000s, English holds a unique position among the modern languages taught in Swedish compulsory school because it can be taught already from grade 1, whereas other modern languages are not taught until grade 6 (Malmberg, 2000a:9; Viberg, 2000:30).⁵ This unique position of English in Swedish schools is a reflection of its status in Swedish society.⁶

English is very dominant in the media. On Swedish television, English-speaking programs and films are not dubbed but subtitled. It has been suggested that many children in countries where the original soundtrack is kept and subtitling is used know how to speak and understand a considerable amount of English even before they receive formal education (Berns, 2007:8), and there are studies which show that viewers draw information both from the visual (subtitles) and aural (spoken TL) input (Berns et al., 2007b:33; d'Ydewalle & Van de Poel, 1999:234).⁷ Moreover, a recent study by Webb and Rodgers (2009b) on the vocabulary demands of TV programs concludes that if learners know the 3,000 most frequent word families in English and watch at least one hour of TV a day, there is a potential for significant incidental vocabulary learning. Regarding the vocabulary demands of films, the same researchers show in another publication that a vocabulary size of 3,000 word families was necessary to reach 95% lexical coverage of American and British movies.⁸ For viewers who comprehend that ratio of words, there is a possibility of incidental vocabulary acquisition (Webb & Rodgers, 2009a:408).

English is also the dominant language on the Internet (see e.g. Crystal, 2001). The use of the Internet is widespread in Sweden, both among university-educated citizens (94% use the Internet) and among those who lack higher education (70%) (Findahl, 2007:13-14). The fact that seven out of ten Swedes who lack higher education use the Internet is in great contrast to other countries in Europe which are behind Sweden in this respect (Findahl, 2007:14). Moreover, English is the lingua franca used when playing video games online or when engaging in other online activities, for instance publishing materials on YouTube (Purushotma, 2006:127-128). And, as was mentioned above, English is also the dominant language in music. To sum up, English was

⁵ Each individual school decides in what grade the teaching of English is to begin.

⁶ Note also that it has been argued that English is one of the school subjects where there is most learning outside school (Skolverket, 2004b:33).

⁷ However, results from another study (Sachs, Bard, & Johnson, 1981) suggest that linguistic input from television is not enough for children to develop speech.

⁸ *Lexical coverage* is the percentage of known words in discourse (Webb & Rodgers, 2009a:408).

previously a foreign language in Sweden, but that is not the case anymore. English functions more like a second language (Viberg, 2000:28-30).

It is important to account for media habits among young Swedes, due to the presence of English in the media. Relevant statistics are available from the Media Council (Medierådet, 2008). The surveyed target group (N=2,000; aged 9-16) makes up a sample consisting of two subsets based on a division of the sample at the age of 12, where “children” are under 12 and “adolescents” 12 or older (Medierådet, 2008:5). Questionnaires were used and the total response rate was high (70%). In the results, those who reported spending three hours or more per day per media activity (i.e. 21 hours per week) are considered to be “high frequency users”.⁹ Of the investigated media-related activities, “the Internet” had the highest ratio of high frequency users (18%), but “TV” was still the most common media activity with 78% of the whole sample reporting watching TV on a daily basis (Medierådet, 2008:8). Furthermore, it is noted that about 62% of all respondents reported using the Internet daily. In comparison with previous reports, the ratio of daily Internet users had increased. Moreover, an interesting finding was that for the first time, the number of adolescents who reported having a computer in their own room surpassed the number of adolescents who reported having a TV in their own room, something which is interpreted as a paradigm shift (Medierådet, 2008:9). The report also reveals that the ratio of high frequency users of “the Internet” and “video games” had increased since the previous survey in 2006. Even though much spare time is spent on various media-related activities, the report shows that “meeting friends” was still the most common activity overall. With regard to gender, there were statistically significant differences in boys’ and girls’ use of mobile phones and the Internet. Girls are considered more advanced mobile phone users than boys because they reported texting people and receiving/sending pictures more frequently than boys did. Girls also downloaded music more often than boys. In addition, there was also a higher ratio of girls who reported using the Internet in their spare time (94%) than boys (89%). It should be noticed, however, that both genders appear to be very frequent users of the Internet. The difference between boys and girls was more evident in terms of their TV habits. For example, *The Simpsons* was very popular with the boys, but not at all with the girls, who preferred the Swedish soap *Andra Avenyn* or *Let’s Dance* (significant differences). There was also a significant gender difference in playing video games: more boys than girls reported playing video games. Boys

⁹ Swe. *högkonsumenter*.

and girls also had different favorite video games; the boys preferred *World of Warcraft* and *Counter-Strike*, whereas girls preferred *The Sims* and *The Sims 2*.

Thus, out-of-class English, or extramural English/EE which is the term I prefer, is omnipresent in Sweden. There are numerous sources which point to the potential of language learning with the help of EE. For example, there are learner self-assessment studies and learner retrospection studies which reveal that learners indeed value EE highly in L2 acquisition. Results from some such studies are presented in the following section.

3.1.4 Learner self-assessment and retrospection studies on EE

There are studies on learner self-assessment and retrospection in which language activities that can be labeled “extramural English” are in focus. These studies, which I account for in the present section, were carried out in countries other than Sweden and sometimes involved learners older than those in my study. Nevertheless, despite some dissimilarity, I find it important to present their main findings as background information.

Pickard (1995) conducted a case study of three highly proficient English language learners (German L1). Based on questionnaire and interview data, he found that, in addition to listening to the radio, very common activities in the spare time among these three learners were reading newspapers and novels outside the classroom. Another study on EE is Nunan (1991), who studied 44 so-called good language learners and their habits of EE with the help of a questionnaire and interviews. He found overwhelming evidence that formal classroom instruction was insufficient for learning English. That is, good learners applied their developing language skills outside the classroom as well. The EE activities which helped the good learners the most in learning English were out-of-class conversations with English speakers and access to the media (radio, TV, and newspapers). Since Nunan’s 1991 study, access to computers and video games has increased and, even though Gee (2003) is not a self-assessment or retrospection study, it is worth mentioning that he argues that playing video games develops a new type of literacy that runs parallel to literacy defined more traditionally, i.e. as reading books and writing pen-and-paper style.

In a study of language learning attitudes and activities among adolescents in provincial Indonesia during their first year in junior high school, it was found that young learners (N=219) learned English independently, i.e. without instructions from their teacher, both inside and outside the classroom (Lamb, 2004:229), a finding which supports the idea of learner autonomy. Their EE

activities mainly involved listening to music and watching movies and television. Reading books or magazines in English was less common, and access to computers was very limited. Similar interests were found among eight adult ESL learners in Japan, who were the informants in Murray (2008). He let these learners inform him about their experiences of learning English. The key finding of that study was that the learners all shared an interest in American pop culture (TV, movies, and music), which had been their source of motivation to learn English.

With the spread of IT and the Internet, new encounters with the English language became possible. Lam (2000), for instance, shows how a Chinese teenage learner of English discursively constructed his identity in English while engaged in synchronous and asynchronous computer-mediated communication (written correspondence) with a group of peers. Not only extramural writing, as was the case with the learner in Lam (2000), but also extramural reading in English seems to promote learning. Arnold (2009) found that university learners (N=8, native speakers of English) who were involved in an extensive reading program in German as a foreign language (online reading, no teacher pre-selection of books) improved their reading ability and motivation to read. This included pleasure reading in their spare time. Interestingly, some learners purposely sought out more difficult texts to challenge themselves. This particular finding is in line with Krashen's input hypothesis (see section 2.1) and Laufer's (1991:446) conclusion about L2 vocabulary acquisition, namely that learners must be challenged in order to develop their language ability beyond a certain threshold (see section 5.2). The fact that the learners in Arnold's 2009 study looked for more challenging texts to read was interpreted as an indication of their growing intrinsic motivation and self-efficacy (Arnold, 2009:360). In another study on extensive reading, learners (N=4) revealed in self-reports and interviews that they felt that their aural and oral proficiency in English had improved thanks to the reading. The researchers hypothesize that the learners' self-assessed improvement was at least partly due to their increased vocabulary, as measured in pretests and posttests (Cho & Krashen, 1994:666-667).

In sum, the results of these self-assessment and learner retrospection studies point in one direction, namely that various EE activities contribute to L2 acquisition. Furthermore, in all the studies there is an underlying general assumption that there is a causal relationship between learners' EE and their proficiency in English. This assumption is troublesome for several reasons. For one, which is the chicken and which is the egg? Do learners become more proficient thanks to EE, or do proficient learners engage in EE activities to a

larger extent than less proficient learners? Very few researchers address this problem, even though there are exceptions, for example Pickard (1995:37). Nevertheless, apparent relationships between learners' involvement in EE activities and level of proficiency suggest that EE has great potential in L2 acquisition.

3.1.5 Other studies on out-of-class learning

There are other studies on out-of-class learning than the ones mentioned above. For instance, Freed (1995:137) investigated studying abroad and its effects on learners' oral fluency and found that one oral fluency measure, rate of speech, was significantly better among students who had studied abroad in comparison with those who had not studied abroad. It should be noted that studying English abroad per se is not equivalent to extramural English, since the studying takes place in an educational setting. However, in the spare time while studying abroad, there are golden opportunities for learners to interact in the target language, i.e., to engage in extramural English activities. In contrast to Freed (1995), Sundh (2003:139), who examined whether time spent abroad had an impact on his informants' English (as measured in that study), could not see any positive effects of such experience.

The out-of-class context is obvious in the rather famous "hole-in-the-wall" experiments carried out by Sugata Mitra and his colleagues (Mitra, Dangwal, Chatterjee, Jha, Bisht, & Kapur, 2005). Their work has shown that children, irrespective of who or where they are, can learn to use computers on their own in an out-of-class context. The "hole-in-the-wall" refers to public computers that are made available to children in urban slum areas and rural locations in India. The technology offers children unique intellectual experiences and opportunities and the researchers were able to conclude that in addition to the acquisition of computer literacy, the children also benefit in other areas, such as socially and academically (Mitra et al., 2005:23). The hole-in-the-wall experiment is set in an out-of-school context very far from the conditions in the present study. Nevertheless, its implications are general. Children who really become interested in something (cf. a hole-in-the-wall/an EE activity) challenge themselves and take charge of the situation. This is the fundamental principle of learner autonomy (with specific regard to language learning, see e.g. Holec, 1981, and Reinders, 2000). However, empirical studies are needed, in particular studies on EE and language proficiency based on data from reasonably large samples of students, because some of the studies

mentioned above are very small. My study is one that aims to fill this void in research, but there are others as well, which I discuss in the next section.

3.2 Empirical studies combining data on EE with learning outcomes

For some time, researchers in applied linguistics have recognized the importance of access to and participation in second language communities as an essential aspect of language learning (see e.g. Black, 2009:690). However, little research has focused on the links between instructed contexts of L2 learning and L2 use in other contexts (Higgins, 2009:401-402). Furthermore, Higgins (2009:402) argues that this relationship between instructed language learning and out-of-class use of the target language is “radically under-theorized”, a statement I fully agree with and something which made the undertaking of writing the current chapter particularly challenging. Nevertheless, in the present section I give an account of empirical studies which combined data on extramural English with learning outcomes.

Pearson (2004) did a longitudinal study of the English learning habits among eight Chinese students enrolled in a twelve-week English for Academic Purposes course and combined it with a questionnaire distributed to all 106 students in the same course. The aim of the study was to document and verify what learners did outside of class. The students who participated in the longitudinal study kept written records and evaluations of their studying of English in the university self-access center and library, and they met every two weeks with the researcher to discuss their out-of-class activities. The questionnaire was distributed at the end of the course and included questions with regard to, for example, what materials were used outside of class, why the informants used these materials, and what extramural activities they had been involved in to improve their English (11 predetermined activities plus a final open category). Among other things, Pearson (2004:7) concludes that learner motivation and awareness were important aspects of out-of-class language learning. Moreover, he found that students’ language proficiency (as measured by an initial placement test) was one factor which influenced what type of extramural activities the informants engaged in, and how frequently. For instance, less proficient informants used the self-access center less frequently than those who were more proficient. In addition, informants with high intrinsic motivation seemed more enthusiastic about their out-of-class learning and were involved more actively in using English than those with lower intrinsic motivation. Results of the questionnaire revealed that the informants were most frequently involved in the following five out-of-school English

language activities (in rank order): (1) listening to/watching news on the radio or TV; (2) studying in the library; (3) reading books, magazines, or newspapers; (4) watching TV programs, videos, or movies; and (5) listening to music or the radio.¹⁰ Results also showed that several informants rated out-of-class language learning more highly as regards both enjoyment and learning than their formal classroom-based learning (Pearson, 2004:4). Unfortunately, Pearson's 2004 study did not attempt to identify or quantify potential language gains specifically from out-of-class activities.

Out-of-class video game playing was not investigated in Pearson (2004), but others have focused on that activity, for example Piirainen-Marsh and Tainio (2009). Based on data from video recordings of game interaction recorded in the players' homes (Finnish boys, aged 10-14, playing *Final Fantasy X*), the researchers argue that the informants developed their linguistic and interactional competence in English thanks to the lexical and prosodic repetitions that were integral features of the video game.¹¹ The game enabled the players to reproduce, practice, perform, and play with different characters' styles and accents of English (Piirainen-Marsh & Tainio, 2009:165). On the whole, repetition was found to offer a flexible resource through which players participated in the play actively, and this practice was found to be inextricably linked to learning (Piirainen-Marsh & Tainio, 2009:166). Similar findings on the importance of repetition is found in Folse (2006) and Hulstijn (2001), but with regard to vocabulary. Folse (2006:273) found that the number of word retrievals was the most important condition (out of three tested conditions)¹² for learners' word retention, whereas Hulstijn (2001) argues that word retention depends on the quality of processing, i.e. learners' elaboration of aspects of a word's meaning and form, and on the amount of rehearsal. Such elaboration processes and repetition are thus important for L2 vocabulary growth (see also de la Fuente, 2002; Kim, 2008; Laufer & Hulstijn, 2001). Vermeer (1992:159) further claims that words to be learned should be embedded in various meaningful redundant contexts in which they have an unambiguous meaning and in which they are frequently paraphrased and repeated, preferably supported visually. Such an approach is likely to lead to varied and meaningful links between concepts, Vermeer (1992:159) says, and to guarantee correct

¹⁰ One activity, "listening to the radio", is in fact included in both (1) and (5), something which weakens the result. "Playing video games" was not listed among the predetermined activities.

¹¹ *Final Fantasy X* includes voice-over dialogues, making spoken English an integral feature of the game (Piirainen-Marsh & Tainio, 2009:158).

¹² Target vocabulary was practiced in three types of written exercise conditions: (a) one fill-in-the-blank exercise, (b) three fill-in-the-blank exercises, and (c) one original-sentence-writing exercise, where (b) proved to be best for target word retention (Folse, 2006:273).

connections between concept and word.¹³ Thus, it appears that EE activities such as playing video games and using the Internet (e.g. for chatting, publishing, and surfing) should provide suitable opportunities for language learning since certain linguistic features are repeated several times and presented in various ways in those mediums.

Video game playing, as in the Finnish study above, is one way for language learners to engage in second language communities; another is to interact with native speakers, which is something that Norton (2000) studied. Her longitudinal study of immigrant women in Canada may seem irrelevant to my study of Swedish ninth graders, but it is not. Her findings on power relations between language learners and target language speakers are interesting and perhaps also applicable to power relations between L2 learners at varying proficiency levels. Even though no learning outcomes are reported, I will briefly comment on Norton (2000) here. Norton's (2000:122-124) data suggest that in natural learning situations, a learner's anxiety is associated with the learner's oral skills rather than his/her literacy skills. Unlike their Canadian colleagues who were fluent in English, the informants in her study reported that they could not take customers' orders and speak to customers at the same time. Thus, the problem was the online processing of speech. Another perhaps more interesting finding was that the informants, who were all immigrants, reported feeling more comfortable and fluent in speaking English when interlocutors did *not* constantly show their linguistic and cultural superiority. Native speakers' comments on their foreign accent, for example, were perceived as discouraging by the informants. Such comments increased their level of anxiety and led to mistakes in speech that could otherwise have been avoided (Norton, 2000:123). Together the findings suggest that anxiety is not an inherent trait of language learners but one that is socially constructed within and by the lived experiences of language learners (Norton, 2000:123). It is worth pointing out that the informants in Norton's study did not necessarily have the personality traits that are typical of so-called good language learners (see Naiman et al., 1996). That is, a good language learner would not have become as discouraged or anxious in a situation such as the one described above. This discussion serves as a reminder of the fact that language learning is indeed idiosyncratic. Even so, it is reasonable to assume that power relations between peers in one school class might have negative effects on at least some learners' speech, similar to what was described in Norton (2000). However, provided that there are no

¹³ A study which confirms Vermeer (1992) is Chun and Plass (1996).

discouraging power relationships between interlocutors, interaction in the target language is generally considered good for L2 acquisition (see chapter 2).

In her licentiate's thesis, Forsman (2004) studied the situation in English language classrooms in Finland-Swedish schools (secondary level of the nine-year compulsory school). Her aim was to describe and discuss classroom practice, and in addition to that to focus on the influence of extracurricular (cf. extramural) activities in English on students' linguistic and cultural awareness and attitudes. By choosing schools from two different regions in Finland, "South" (mainly urban area) and "West" (rural area), she was able to compare the results depending on the students' residency. She did so due to a previous Finnish study which had shown that students' residency (urban vs. rural) was a more decisive factor for students' level of success than, for example, differences between municipalities (Forsman, 2004:70).

Empirical data included one questionnaire (N=330) and interviews, both with students (N=20) and teachers (N=8). One question in the questionnaire addressed extramural activities in English. Ten predetermined EE activities were listed, plus a final open one. For each activity, the informants were to estimate how much time was spent on it per day or per week (the researcher later transformed all data into "hours per week"). Although at least one informant's estimates were clearly off the wall (surpassing 24 hours per day) and others wrote time intervals (e.g. "4-6 hours/week"), Forsman (2004:85) concludes that she had no reason to doubt that the reported figures gave "enough information" about the relative time spent on the various EE activities; i.e., the data was considered reliable. However, in order to increase reliability, she stipulated rules for the encoding procedure. For example, time intervals were changed into averages (e.g. "5 hours/week" in the example above) and the total which surpassed 24 hours per day was cut to half its value. Finally, she points out the possibility of concurrent EE activities, e.g. listening to music while using the Internet (Forsman, 2004:85). Results showed that the total amount of EE was 51.1 hours per week for informants from the south/urban area and 36.7 hours per week for those from the west/rural area (Forsman, 2004:92). The difference was statistically significant. Most time was reported for "listening to music", followed by "watching TV", "playing computer games", and "using the Internet". Very low values were reported for "reading books" or "reading magazines" in English (on average less than one hour for each of the activities). For the four most popular activities, there were significant differences between informants from urban and rural areas, where those from urban areas reported higher values.

Based on the responses to a question about the informants' word choice preference in American and British English word-pairs (e.g. *gas* vs. *petrol*), Forsman (2004:100, 173) concludes that the correlation between the amount of time spent on extracurricular activities and American English word choices indicates a strong influence of media input both on what students learn and on their attitudes towards different language varieties. Furthermore, with regard to gender differences, she concludes that girls in general expressed greater preference than boys for standard language used in school and for British English (which was the variety mainly taught in the schools). Boys showed evidence of unintentional learning to a greater extent than girls because they opted for more choices of American English words than girls did (Forsman, 2004:177). This final conclusion seems to be far-fetched, I think. For instance, it is possible (or even likely) that girls to a larger extent than boys opted for the British English word just because they knew that this was the unmarked form in the classroom. It is commonly known that girls are more likely than boys to follow teachers' rules (cf. e.g. Björnsson, 2005; Carr & Pauwels, 2006; Sax, 2007).

A final source on extramural English and learning outcomes that I would like to comment on is Sylvén (2004). Even though extramural English was not the main focus of her thesis, she arrived at interesting results with regard to the importance of EE activities. For instance, the boys in her study (upper secondary school) had a larger amount of English input from media than the girls, and the boys also scored higher on vocabulary tests (Sylvén, 2004:226). The boys' habits of involvement in video games and role plays proved to be an important factor contributing to their results on vocabulary. Sylvén (2004:226) also found that those who were involved in out-of-class reading scored higher on vocabulary tests than those who did not read. The effects of time spent in English-speaking countries were difficult to determine. On the one hand, among the students who improved the most throughout the study, the effects of time spent abroad seemed minimal. On the other hand, there were students who had spent between three and twelve months in English-speaking countries and they indisputably had the highest vocabulary scores (Sylvén, 2004:226).

To summarize, in the present chapter, I defined the concept of extramural English (EE). Moreover, I discussed out-of-class learning and included a section about the role of English from a general European perspective, and, more specifically, from a Swedish perspective. English was also discussed in the light of young people's media habits. In addition, I gave an account of learner self-assessment and learner retrospection studies, since they provide a

background to my own study. Their results pointed in one direction, namely that various EE activities contribute to L2 acquisition. The present research topic is under-researched both with regard to theory and empirical studies, in particular with regard to empirical studies that combine data on learners' EE with their learning outcomes. There are, however, some studies which cover both aspects, and in the final section, I presented and commented on the results from those.

4 Oral proficiency in learner English

Since the beginning of the 1980s, studies on oral proficiency and the assessment of oral proficiency in learner English have become more frequent than they used to be. The communicative approach to language teaching, which saw its breakthrough during the decade, functioned as a motivator for these types of study (see Lindblad, 1992). The present chapter, which consists of three sections, focuses on oral proficiency in learner English. The first section (4.1) includes a general discussion of language proficiency and oral proficiency, followed by four sub-sections on basic types of learner speech (4.1.1), the characteristics of learner speech and communication (4.1.2), previous research on L2 oral proficiency (4.1.3), and, finally, previous research on L2 oral fluency (4.1.4). The second section (4.2) discusses accommodation theory and the role of the interlocutor. Both concepts are of interest in the thesis since all speech data in my study were collected from students in dyads. The final section (4.3) is about assessment of oral proficiency and oral fluency.

4.1 On language proficiency and oral proficiency

The *Longman Dictionary of Language Teaching and Applied Linguistics* (2002:292) provides the following definition of *language proficiency*: “the degree of skill with which a person can use a language, such as how well a person can read, write, speak, or understand language”. Taylor (1988:164-166) has a similar definition. He claims that proficiency incorporates what learners know (i.e. their competence, see section 2.1) and their ability to use their knowledge in actual communication. Proficiency is seen as a dynamic concept, closely linked to process and function. Therefore, a very general definition of *oral proficiency in English* is ‘the learner’s ability to speak and use English in actual communication with an interlocutor’.¹ This is the definition of *oral proficiency (OP)* that I adopt in the present study.

Given that oral proficiency deals with a learner’s ability to speak and use a language in actual communication with another person, described as a two-way negotiative effort by Kramersch (1986:368), the close relationship between speaking and listening becomes apparent (see also Levis, 2006:259). In fact, measures of learners’ listening comprehension are at times used to make claims about the same learners’ levels of oral proficiency (see e.g. Hellgren, 1982:12). As is generally known, speaking and listening make up two of the four

¹ My definition is very close to what Pollitt (1996:4) calls “oral interaction ability.”

traditional language skills. Speaking and listening are interrelated since both fall within the oral/aural mode of language, whereas the other two skills, reading and writing, belong to the written mode. But even though speaking and listening are interrelated, they are clearly different. For example, speaking is observable, whereas listening is non-observable, at least in direct performance. In regards to research on speech, it is a person's performance that is observable – provided that there is an audio or video recording that preserves the output (D. H. Brown, 2004:118).

Often speaking is considered the most complex and difficult skill to master for a learner (Tarone, 2005:485). Speaking is commonly referred to as the productive skill of the oral mode. In contrast, listening is referred to as a receptive skill. It is, however, too simplistic to label listening a receptive skill because listening also involves a certain degree of productivity on the part of the listener. Paying attention to what someone else says includes simultaneous cognitive processes of reception and production. In short, it is reasonable to assume that learners' aural comprehension also affects their oral proficiency (Bachman, 1990:103; D. H. Brown, 2004:119). Thus, speaking is more or less intertwined with listening.²

Regarding oral proficiency, it is a multifaceted concept. That is, oral proficiency involves several aspects of language, such as vocabulary, grammar, pronunciation, prosody, fluency, and interactional skills. There are studies which aim to investigate the concept of OP as a whole, i.e., as a variable consisting of several aspects of language (see e.g. Bialystok, 1991; de Jong & van Ginkel, 1992; de Jong & Verhoeven, 1992), and others which concentrate on only one (or more) of the aspects, for example interactional skills (e.g. Riggensbach, 1998), fluency (e.g. Lennon, 1990; Mizera, 2006; Wood, 2001), vocabulary (e.g. Afitskaya, 2002; Altman, 1997; Miliander, 2003), or repair strategies (e.g. van Hest, 1997; Plejert, 2004). Needless to add, there are also studies which do both; i.e., they examine the concept of OP as a whole as well as some aspect(s) of OP (see e.g. Iwashita, Brown, McNamara, & O'Hagan, 2008). My study is of the latter kind. I begin by investigating oral proficiency as a whole for all participants in my sample. For a selection of ten students, I also investigate two specific aspects of OP, namely fluency (see section 6.3.3) and vocabulary (see section 6.3.4).

² I might add that a person also listens to himself or herself.

4.1.1 Basic types of learner speech

As is clear from my definition of oral proficiency in section 4.1, the focus is on learner speech produced in communication with an interlocutor.³ There are several types of learner speech that may be elicited in order to investigate learners' oral proficiency. Brown (2004:141-142) presents a taxonomy which includes five types of speaking: *imitative speaking* (i.e. the ability to parrot back language, pronunciation tasks), *intensive speaking* (e.g. read-aloud tasks, sentence and dialogue completion, sentence level translations), *responsive speaking* (e.g. giving instructions and directions, standard greetings and small talk, simple requests, limited exchanges), *interactive speaking* (e.g. interview, role-play, discussions and conversations, multiple exchanges), and *extensive speaking* (e.g. oral presentations, picture-cued story-telling, speeches, informal monologues). In studies of oral proficiency, researchers choose the tasks that suit their aims.

It is fairly common that tasks used to elicit extensive speaking are adopted in oral proficiency studies (cf. e.g. Iwashita et al., 2004; Sundh, 2003) since such tasks are of a monologic nature, something which facilitates recording and subsequent analysis. However, intensive and interactive speaking tasks are also used (cf. e.g. Buckwalter, 2001; Foster & Skehan, 1996; Hasselgren, 2004), even though interactive tasks (which involve a minimum of two speakers) put greater demands on both recordings and analyses. However, the benefit of using tasks that resemble natural communication supposedly makes up for the potentially more problematic follow-up work.⁴

4.1.2 Characteristics of learner speech and communication

It is generally accepted that language exists primarily as speech and that conversation is the most basic form of human communication. Speech is time bound and dynamic as opposed to writing, which is space bound and static. The differences between these two modes of language are due to how language is produced and perceived. In a conversation many conversational traits arise from the fact that speech is typically spontaneous. Speakers need to plan and execute their utterances in real time and as a natural consequence so-called *normal dysfluency* occurs (Biber, Johansson, Leech, Conrad, & Finegan, 1999:1048). Examples of normal dysfluency are pauses, hesitations (*er, um*), and repetitions of single words (or phrases) at a point when one's need to keep talking runs ahead of one's need to plan what to say. Normal dysfluency is

³ See section 6.2.3 as to why a prepared speech, during which no interaction took place, was used as part of Test 4.

⁴ In my study, a majority of the speaking tasks involve interactive speaking (see Table 6.2).

common also in L1 speech; thus, it is a feature of all types of speech. Other characteristics of speech are reduced structures, such as contractions (*didn't* instead of *did not*), and vague instead of specific referencing (e.g. saying “that kind of thing” instead of giving a more detailed description of something). Ellipsis, non-clausal utterances, vague nouns (e.g. *stuff, thing*), and hedges (e.g. *sort of, kind of, like*) are also common characteristics of speech, as is the so-called *add-on strategy*, where the speaker produces long utterances made up of finite clause-like units (Quaglio & Biber, 2006:702-703).

As was previously mentioned, there are many similarities between L1 and learner speech. However, there is an important difference in terms of accuracy. Due to the “online” cognitive processing involved in speech, learners generally do not have time to consciously consider or apply grammatical rules (Krashen, 1981:3) and they may therefore produce ungrammatical structures. In contrast, such conscious formal considerations are redundant for L1 speakers since L1 grammar is internalized (Chomsky, 1957:15).⁵

According to the interaction hypothesis (see chapter 2), L2 learning takes place in interaction when learners need to negotiate for meaning and are forced to produce output. Several studies point to the fact that learner-learner communication is fruitful for L2 acquisition (see e.g. Gass & Mackey, 2006:13). However, in a study on repair sequences in Spanish L2 dyadic discourse, Buckwalter (2001:391) found that learner-learner discourse in the TL rarely led to breakdowns in communication due to misunderstandings. Consequently, there were few opportunities which triggered the participants to interact in line with the interaction hypothesis. Similarly, Foster and Skehan (1996:301) show that relatively few informants in fact account for most negotiation of meaning in learner interaction; i.e., most learners do *not* engage in negotiation of meaning at all. This is especially common where the L1 is shared. But, I might add, Buckwalter (2001:392) did find that learners collaborated regarding the lexicon, which might be an indicator of L2 learning taking place despite the fact that there were few opportunities for negotiation of meaning.

Various aspects of learner communication have been studied. For instance, a study on Spanish ESL students (university level) (Naughton, 2006:172) revealed that learner speech contains fewer follow-up questions than L1 speech. Another interesting finding was that learners who were *not* oriented towards exploiting learning opportunities that arose during conversation often took the easy way out. That is, they mixed their L2 English with L1 Spanish,

⁵ Conscious considerations regarding grammar may, of course, also occur in L1 speech.

something which did not lead to interlanguage development (Naughton, 2006:178-179). Mixing languages in such a way is not unusual among Swedish learners either (see e.g. Malmberg, 2000b:119-132). In Naughton (2006:177), it is also suggested that rules of socially acceptable behavior had a greater influence on interaction patterns than lack of ability on the part of the learners. This is an interesting suggestion, I think, because it sheds some light on the ongoing rather complex conversational process between interlocutors (see section 4.2).

Not only the rules of socially acceptable behavior may affect learners' speech, but also the amount of time allowed for planning. In a study from Hong Kong (Littlewood, Liu, & Yu, 1996), the researchers report that learners (university level) who had grown accustomed to passive speech roles in previous English classes, with limited experience of actual speaking activities, had low confidence while speaking in English, especially when speech was unplanned. This finding is in line with results from other studies on planning and speech production. Mehnert (1998:83) shows that as little as one minute of planning time before speaking tasks commenced led to improved accuracy in learner speech (university students, L2 German). An earlier study on the same topic revealed similar results for the relationship between planning and performance, including gains also in oral fluency (Foster & Skehan, 1996:317-318). As a consequence, it seems like a good idea to allow some time for planning in studies on oral proficiency, such as mine (cf. section 6.2.3).

4.1.3 L2 oral proficiency studies

It was mentioned in chapter 1 that there is limited research on OP among adolescent learners. Nevertheless, in the current section, I present the results of some studies that are pertinent to my own study. For example, in 2003, a Swedish thesis was published which focused on the assessment of OP in English among students (N=29, aged 19-20) in their last year of senior high school (Swe. *gymnasieskola*) (Sundh, 2003) (see also section 4.3). The results of that study indicated that students who were awarded high overall grades on OP spoke at length and discussed abstract and demanding topics. In addition, a factor that appears to have affected students' grades was the actual order of the tests since, in general, students improved from their first test to their third.⁶ Also, there was a gender difference regarding vocabulary, with boys using abstract nouns more frequently than girls. On the other hand, girls used

⁶ The participants took three tests in one day.

modifiers of adjectives (e.g. *very*) and the discourse marker *you know* more often than did boys (Sundh, 2003:263). Finally, another interesting finding was that students with low grades in English from school could perform equally well on the OP tests as students with high English grades, an indication of “uneven profiles” among learners, according to Sundh (2003:263). That is, some learners may be better at speaking than writing, for example.

In another study, Lorenzo-Dus and Meara (2004) examined OP based on data from tasks involving role-play, i.e., interactive speaking. It was found that role-play provides micro-contexts of interaction in which the learners (adults, L2 Spanish) indeed displayed aspects of their L2 speaking ability that went beyond lexis and grammar. For instance, subjects displayed the ability to establish and change topics, hold and yield the floor, interrupt, collaborate, and use backchannels.⁷ The researchers argue that the results are an indication of learners’ management of topics and turns, both of which are important features of interactional competence (Lorenzo-Dus & Meara, 2004:93). Thus, tasks involving role-play seem to be useful in studies on OP and some such tasks were included in the speaking tests in my study (cf. section 6.2.3).

In a Dutch study (de Jong & van Ginkel, 1992), one aim was to assess whether the relative contribution of subskills (i.e. aspects of OP) changed with increasing oral proficiency. In general, considerable weight is given to pronunciation and fluency (referred to as “channel-bound aspects”) when speakers’ OP is evaluated, both by laymen and by scholars (de Jong & van Ginkel, 1992:187). The reason for this specific weighting is the fact that pronunciation and fluency are very important in successful communication. For learners at the lower end of the OP scale, pronunciation was found to contribute almost 50 per cent to overall ability (as assessed by an overall grade), whereas fluency contributed very little. However, as the level of OP went up, fluency became more important. Accuracy and comprehensibility took on a middle position. Moreover, for learners at the higher end of the scale, it was found that each of the subskills provided “about an equal contribution to the overall ability” (de Jong & van Ginkel, 1992:202). The authors suggest that accuracy and comprehensibility constitute two perspectives of what is a single dimension, the perspective of the speaker and the perspective of the hearer:

⁷ According to Wolf (2008:280), *backchannels* (or *backchannel cues*) are the listener’s/interlocutor’s verbal (e.g. *uh-huh*, *yeah*) and nonverbal (e.g. head nods, smiles) messages to the speaker in oral interaction.

The degree of accuracy achieved by the speaker in encoding the message has a direct effect on the comprehensibility of the message to the hearer. Aspects of channel control, on the other hand, such as pronunciation and fluency constitute two separate dimensions of speech production. The fluency aspect seems to discriminate effectively between subjects receiving lower and higher global ratings. Apparently, in superficial evaluation or global ratings of oral proficiency, this aspect of channel control dominates (de Jong & van Ginkel, 1992:202).

The three studies discussed in the present section examined the concept of OP as a whole as well as some aspects of OP. All three studies were also concerned with the assessment of OP. The assessment of OP is discussed further in section 4.3, where I will return to two of these studies (de Jong & van Ginkel, 1992; Sundh, 2003). In the following section, focus is on one aspect of oral proficiency: fluency.

4.1.4 L2 oral fluency studies

As I mentioned at the beginning of this chapter, the communicative approach in language teaching and learning had its breakthrough in the 1980s. Its emphasis on interactive tasks and the concurrent development of oral assessment materials (see e.g. Lindblad, 1992) initiated a scholarly interest in the study of oral fluency in learner speech; an interest which, in fact, had not received a great deal of attention earlier (McCarthy & O'Keeffe, 2006:95).

To achieve oral fluency, it is crucial to have control of linguistic processing in real time and to pay attention to relevant and appropriate information (Bialystok, 1991:71-72). In fact, the key difference between fluent and non-fluent speakers is claimed to be the ability to formulate utterances smoothly under time pressure (Levis, 2006:265). Nevertheless, even though everybody seems to agree that fluency is important, it is an elusive concept, to learners as well as to teachers, researchers, and others. The term is difficult to define, as is pointed out by several researchers (see e.g. Freed, 1995:123; Kormos & Dénes, 2004:145-146; Levis, 2006:263). As a matter of fact, Hasselgren (2002a:147) even says that the notion of fluency “haunts” linguists, for example language testers:

We [language testers] have a feeling that we know what it is, and that it is too significant to be left out when assessing spoken interaction. What is more, the people we test for – learners and teachers, and the public at large – seem to share this view. Yet pinning down and describing fluency with a degree of consensus is notoriously difficult (...).

Even though defining and assessing fluency is difficult, Hasselgren (2002a) points to features that do indeed seem to correlate with oral fluency

ratings, such as learners' higher speech rate and quantity, accompanied by fewer disruptive pauses (cf. Lennon, 1990), or learners' access to ready-made chunks of speech (cf. Erman & Warren, 2000; Fillmore, 1979). But the question remains, how can oral fluency be defined? In the aforementioned study, Hasselgren (2002a:148) provides the following definition of *oral fluency*:

The ability to contribute to what a listener, proficient in the language, would normally perceive as coherent speech, which can be understood without undue strain, and is carried out at a comfortable pace, not being disjointed, or disrupted by excessive hesitation.

It should be pointed out that we speak of fluency in its more *narrow sense* here, as "one, presumably isolatable, component of oral proficiency" (Lennon, 1990:389). The opposite is fluency in its *broad sense*, where it is used as a cover term for oral proficiency, representing the highest point on a scale that measures spoken command of a second language.

Angela Hasselgren, University of Bergen, has carried out a great deal of research on oral fluency among adolescents in Norway. In addition to oral fluency, her focus has been on so-called *smallwords* (e.g. *you know, okay, right*).⁸ Hasselgren draws on data from the speaking test part of the EVA (Evaluation of English as a school subject) diagnostic test material, which was developed in the 1990s (Hasselgren, 1996a). One of the speaking tests in the present study (namely Test 2) was from the EVA test material. The main purpose of the EVA project was to provide teachers and their students (grade 8) with information about the students' communicative language ability, both strengths and weaknesses, so that learning activities could be better adapted to learners' specific needs.

Hasselgren (2002b:103) examined 19 smallwords both quantitatively and qualitatively. The quantitative analysis covered tokens of smallwords used as well as the range (types).⁹ Moreover, the data were investigated with respect to the position of smallwords in a turn: turn-initial, turn-final, *loner* (i.e. free standing, relative to number of turns), and turn-internal (relative to number of words) (Hasselgren, 2002b:105). She demonstrated that smallwords contribute significantly to fluent speech in English (Hasselgren, 2002b:103). The more fluent pupils (grade 5.0 or higher on a 1-6 scale; see section 6.2.4) used smallwords in a more native-like way than the less fluent pupils (grade 4.0 or

⁸ *Smallwords* are defined as "small words and phrases, occurring with high frequency in the spoken language, that help to keep our speech flowing, yet do not contribute essentially to the message itself" (Hasselgren, 2002b:103).

⁹ *Tokens* are the total number of word forms in a speech sample whereas *types* are the total number of different words in a speech sample (see e.g. Vermeer, 2000:66; see also chapter 5).

lower), concerning quantity and distribution across turns. Furthermore, the group which included the more fluent students used more smallwords than the other group.¹⁰ Another interesting finding was that the group of more fluent students used smallwords in loner positions to the same degree as the students in a control group, which consisted of adolescent native speakers of English (NSs).

The qualitative analysis involved group comparisons of the way smallwords were used for signalling various functions. Here, the control group of NSs functioned as a yardstick. Hasselgren (2002b:122) concludes that as pupils gain in fluency, “they appear to have the means to send all the basic signals.” However, this might just be a brilliant disguise, because there is still a real difference between the ranges of smallwords used by more fluent learners and NSs.

Use of formulaic expressions has been claimed to facilitate oral fluency (Fillmore, 1979:92). In a study of ten native speakers’ and ten non-native speakers’ oral fluency (Lin, 2006), two different speaking tasks were used in order to elicit speech: picture-telling (i.e. extensive speaking) and problem-solving (i.e. interactive speaking). Lin (2006:113-115) found that native speakers (NSs) and non-native speakers (NNSs) did *not* differ in their use of formulaic sequences in the problem-solving task. This finding is very interesting, even though it should be noted that the result was not statistically significant. Even so, Lin argues that in a problem-solving task, speakers – irrespective of whether they are native or non-native speakers – have time to plan their speech while their interlocutor holds the floor. This time for planning makes the speaker’s access to formulaic structures less important in order to speak fluently. In contrast, in the picture-telling task, there was indeed a statistically significant difference between the groups. That is, NSs outperformed NNSs concerning fluency and use of formulaic expressions. Lin (2006) suggests that the difference between the groups here is due to the fact that the picture-telling task, which is monologic, puts greater demands on learners’/NNSs’ cognitive processes, making access to formulaic structures crucial. In sum, knowledge of formulaic expressions seems to facilitate oral fluency in cognitively demanding tasks. In the present study, tasks involving interactive speaking were used more frequently than tasks involving extensive speaking (see Table 6.2). However, in Test 2, picture-cued story telling (i.e. extensive speaking) was used, similar to one of Lin’s (2006) tasks. Thus, at least based on her results, it can be assumed

¹⁰ Both groups were gender-balanced.

that the students in my sample who had already acquired a certain degree of knowledge of formulaic language when the study began had an advantage over students who lagged behind in this respect.

The cognitive dimension of tasks is something that not only Lin (2006) has focused on. Robinson, Ting, and Urwin (1995), for example, also report interesting findings regarding cognitive load and fluency. They examined the difficulty level of speaking, writing, and listening tasks on three dimensions: (1) amount of cognitive load imposed, (2) amount of planning time allowed, and (3) amount of prior information supplied. One generalization that emerged was that harder tasks on the dimension of cognitive load tended to encourage greater accuracy, whereas easier tasks tended to encourage fluency on all three dimensions. Evidently, task type matters for L2 performance, as is also shown by Foster and Skehan (1996:312-313) who, for example, conclude that there are strong effects of planning on fluency in learner speech (see also Mehnert, 1998).

Thus, several studies have shown that planning has an effect on oral fluency. There are also studies which point to a correlation between L2 reading and oral fluency. In a study on young learners (N=57, 3rd grade, L1 Spanish, L2 English), informants were divided into three groups according to their performance on a diagnostic reading test: low, intermediate, and high L2 readers (Peregoy & Boyle, 1991). L2 oral proficiency was individually assessed using an interactive, simulated science lesson. The speech samples were then rated on four "oral variables": grammatical complexity, well-formedness, informativeness, and comprehension. One finding in the study was that significant differences on all four oral language variables were identified, which suggests that L2 proficiency encompasses both oral and written language (Peregoy & Boyle, 1991:45). This is an interesting finding, even though it should be pointed out that nothing can be said regarding a causal relationship between L2 reading on the one hand, and L2 oral fluency on the other. To make such a claim, much larger studies are needed.

Oral fluency can be measured both qualitatively and quantitatively. In quantitative studies, it is common to measure oral fluency in relation to temporal variables or various hesitation phenomena.¹¹ Empirical research on temporal variables reveals that speech rate is an excellent indicator of oral fluency. Another is pauses, which can be investigated in terms of placement, frequency and/or length. According to Wood (2001:575-576), placement of pauses in particular tells us something about the nature of oral fluency. Like

¹¹ For an overview of fluency studies, see Ellis and Barkhuizen (2005:156).

native speakers, highly fluent L2 speakers tend to pause at sentence and clause junctures, or between non-integral components of clauses as well as between clauses themselves. That is, fluent speakers avoid pausing in intra-positions/within phrases because such behavior leads to dysfluency. As for frequency of pauses, it has been found that the number of unfilled pauses correlates with proficiency level, as does pause length (total pause time) (Iwashita et al., 2008:41). Pause length can be measured as total pause time or as mean pause length beyond some threshold, commonly set at one second (cf. Foster & Skehan, 1996:310). I find it troublesome that many studies never discuss why a particular threshold is used. In the present study, mean intra-utterance pause length *beyond one second* is used to measure oral fluency and the reason for that particular threshold is given in section 6.3.3.

Temporal variables were also used in a Hungarian study, which included 16 learners at two distinct levels of OP (Kormos & Dénes, 2004:145). Speech samples were compared and temporal and linguistic measures were correlated with oral fluency scores awarded by six raters (3 NSs, 3 NNSs). In this study also, speech rate turned out to be a good predictor of oral fluency scores. The number of pauses (unfilled and filled), however, were not found to influence perceptions of fluency (Kormos & Dénes, 2004:161). Finally, accuracy of output also played an important role in fluency judgments.

Wood's (2001:580) overview of oral fluency yet again pinpoints learners' use of formulaic expressions as being of paramount importance for oral fluency. Formulaic structures are stored as wholes in long-term memory. When learners have access to such structures, spontaneous speech is facilitated due to the lessened cognitive load in encoding and decoding. This is in line with de Jong and van Ginkel (1992:204) (discussed above), who suggest that whereas pronunciation might have to do with automatization of a single motor skill, oral fluency would be automatization of the same motor skill *and* sufficient automatization of a cognitive skill. While learners' access to formulaic expressions certainly appears crucial for fluency in speech, the interlocutor (or interlocutors) might also affect speech. The role of the interlocutor is the topic of the following section, along with a theoretical framework known as "accommodation theory".

4.2 Accommodation theory and the role of the interlocutor

Communication accommodation theory is a framework designed to explore phenomena of accommodation. Accommodation is defined as "the adjustment of one's speech or other communicative behaviors vis-à-vis the people with

whom one is interacting” by Giles (2001:193). In other words, it has to do with the way we shift our language depending on whom we are talking to. In communication accommodation theory, two importance concepts are convergence and divergence, where *convergence* has to do with the way a speaker adapts his/her speech to that of the interlocutor and where *divergence* has to do with the way a speaker increases perceived dissimilarities of speech and nonverbal differences between himself/herself and the interlocutor(s) (Giles, 2001:194-195).

There are several studies which touch upon accommodation theory and investigate the role of the interlocutor in oral interaction (see e.g. Iwashita, 2001; Lazaraton, 1996; Lindeman, 2006; O'Loughlin, 1995; Sundh, 2003). In dyadic test situations, a speaker's interlocutor is particularly important, regardless of whether the interlocutor is a peer or an instructor. In a learner-learner set-up with an instructor as a passive bystander (which is the case in the present study), learners might accommodate their speech not only to one another but also to the instructor (see e.g. A. Brown, 2003). Based on data from 58 transcribed interviews (Cambridge Assessment of Spoken English), Lazaraton (1996) shows that the test instructor's comments affect learner productions, i.e., learners accommodate to the instructor. In another study (O'Sullivan, 2002), it was found that the learners varied their language depending on how well they knew their interlocutor (familiar/unfamiliar). In addition, results revealed that the learners received higher scores on OP when they interacted with a friend than when they were paired up with a person who was unfamiliar to them (pair-work, language elicitation tasks) (O'Sullivan, 2002:291). What is perhaps even more interesting was that there appeared to be “a sex-of-interlocutor effect”, which means that when female learners interacted with unfamiliar female interlocutors, accuracy improved to a statistically significant degree. However, the same female learners' accuracy worsened (although not to a significant degree) when the interlocutor was an unfamiliar male (O'Sullivan, 2002:291). It should be pointed out that the sociocultural context in the study, which was Japanese, probably contributed significantly to variability in learner performance (O'Sullivan, 2002:291). Thus, the results cannot be directly transferred to a Swedish context. Nevertheless, it is perhaps possible that the degree of familiarity with one's interlocutor (in the present study – the degree of familiarity with a class mate) has at least some effect on the speaker in his/her speech accommodation processes.

In my final comment on accommodation theory and the role of the interlocutor, I address what has become known as *the cooperative principle*. Yule

and Tarone (1991:167) and Warren (2006:223) both stress the interdependence of speaker and listener in successful communication, described in Paul Grice's so-called cooperative principle. This principle means that you make your "conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grundy, 1995:37). In short, it is assumed that speakers intend to be cooperative when they talk.

4.3 Assessment of oral proficiency and oral fluency

This final section is about the assessment of oral proficiency and oral fluency. I begin with one general problematic area of L2 language assessment: What should be the norm in assessment of oral proficiency and fluency in a second language? Usually, learner speech is compared to that of native speakers. However, there are a number of standard varieties of native-speaker English: British English, American English, Australian English, and so forth. Thus, which variety to use as the norm is a problem in its own right. Still, the native speaker of English generally becomes the yardstick, i.e. the norm (Arnaud & Savignon, 1997:168), despite that fact that the norm is inherently imprecise. In particular, high proficiency speakers are judged against the norm of the native speaker, whereas low proficiency speakers are judged more on accuracy and on how things are said (Pollitt & Murray, 1996:4). Accordingly, OP is understood in different terms at different levels. Currently, the NS yardstick is under attack in L2 research and suggestions are put forward to measure learner speech more on an intelligibility scale than on a native-like scale (see e.g. Cook, 2005; Modiano, 2005). Cook (2005:49-50) even questions if the NS target can ever be attainable, since L2/non-native speakers per definition can never become native speakers. Thus, it is a complex matter what norm to use in the assessment of L2 oral proficiency and fluency, but it is essential that some sort of yardstick be used that provides descriptors or grading criteria, both for different levels of OP and for various aspects of OP (see D. H. Brown, 2004, chapter 7).

Once a yardstick has been decided upon, it is possible for raters to qualitatively assess learners' OP and/or oral fluency. It is common procedure to quantify qualitative assessment; i.e., the adopted descriptors or grading criteria in a study are turned into quantifiable scores or grades on a scale. Sometimes the terms *factorial* and *overall* (or *global*) *scores/grades* are used (see e.g. Hasselgren, 1996a, 1996b; Sundh, 2003), where the factorial grades refer to the assessment of some aspect or aspects of OP and the overall grade to the assessment of the performance as a whole. In the present study, I use these terms (see section

6.2.4). Another term used is *holistic scoring* (see e.g. de Jong & van Ginkel, 1992:9), which is used in reference to the procedure where raters assess the whole performance in one overall score or grade. Finally, the term *subskill score* refers to the score for a particular subskill examined in a study. For example, Hasselgren (1996b:24) uses three descriptors for the subskill “pronunciation” where raters assess whether the testee’s pronunciation was “very good, with no sounds that could be misinterpreted”, “good enough to get the message across”, or “difficult to understand (or would be for a non-Norwegian!)” (see also Hasselgren, 1997).

For ethical reasons, participants in studies on oral language must be informed beforehand that their output is to be assessed (or what purpose there might be for collecting the speech data). This means that all participants are aware of the fact that they are part of a research study. The mere fact of knowing such a thing might have psychological effects on the participants that, in turn, might have a bearing on the outcomes of the study. In literature, such phenomena are known as *the Hawthorne Effect*.¹² That is, in studies on OP, which is the topic here, there might be a positive effect on participants’ spoken output just because they are part of a study. In the following, I briefly describe and present relevant findings from five studies on oral proficiency and/or fluency. The studies are presented in chronological order.

Lennon (1990) investigated oral fluency using speech data from four female informants (L1 German) who were in England for six months to study English. Speech data were assessed by ten raters (teachers, NSs) twice, once upon arrival and once shortly before returning home. As test materials, the same set of a six-picture sequence to elicit narration (i.e. extensive speaking) was used on both occasions, but the informants were unaware of the study design. All in all, 12 quantifiable performance variables considered to be objective indicators of oral fluency were measured. In the study, values per informant and per recording were obtained and expressed as frequency rates or as proportions in order to make possible comparisons between recordings. In order to establish for which linguistic features improvements were manifest, informants’ scores on the two different recordings were compared. Statistical analyses were carried out using *t* tests (see sections 6.3.6.1-6.3.6.2) for each

¹² The name of the effect, “Hawthorne”, refers to the Hawthorne works of the Western Electric Company in Chicago. A series of studies on the productivity of some workers manipulated various conditions (e.g. pay, light levels, rest breaks) and resulted in a rise in productivity which was true of each of the individual workers as well as of the group mean. The variables the experimenters manipulated were clearly not the only causes of increased productivity, nor the dominant ones. The conclusion was that the important effect was the feeling of being studied, hence “the Hawthorne Effect” (*Encyclopaedia Britannica*).

variable between sample means for the first and second recording. Significant improvement (one-tailed, .05 level) was found for three variables: (1) speech rate, (2) filled pauses per T-Unit, and (3) percentage of T-Units followed by a pause, filled and unfilled (Lennon, 1990:388).¹³ In his discussion of rater assessment, Lennon (1990) suggests that disagreement may have derived from individual raters' different reactions to various aspects of fluency. Phonological accuracy and, in particular, perceived levels of "phonological confidence", contributed to a large extent to raters' perception of fluency (Lennon, 1990:409). Lennon's 1990 study, then, shows clear connections between temporal variables (speech rate, pausing) and perceived fluency, but, albeit pioneering and often referred to, I would like to point to the fact that it was a very small study based on data from only four informants.

In a Dutch study, which included 25 learners of L2 French (penultimate year in secondary education) (de Jong & van Ginkel, 1992), OP as a whole was investigated. The performance of each student was assessed by three raters, who were strictly guided by written instructions. First, several factorial grades were awarded on tasks that covered, for example, intensive speaking (reading aloud), extensive speaking (picture description), and interactive speaking (expressing one's opinion, role-play). After that, an overall grade was decided upon, using a scale 1-10, where 5.5 was the lowest passing grade.¹⁴ The study primarily aimed to evaluate the dimensionality of the OP test used, and also to assess whether the relative contribution of subskills changed with increasing proficiency. Regarding the assessment of OP, the researchers suggest that an overall grade of OP might often prove to be too simple and instead attainment along three dimensions of OP should be accounted for separately: pronunciation, accuracy, and fluency (de Jong & van Ginkel, 1992:203). However, they also caution against the generalizability of their test results due to the small sample size and due to the task type used for pronunciation (i.e. reading a passage aloud), which clearly differs from producing one's own speech in real time. That is, even though their study was much larger than Lennon's (1990), they hesitate to draw any major conclusions from their results. I might add that in a discussion on the validity of holistic scoring by de Jong and Verhoeven (1992:9), the criticism of such scoring is repeated. However, at the same time the researchers note that for practical purposes an overall grade

¹³ Speech rate was measured as pruned words per minute, where *pruned words* are basically all produced words minus self-corrected words, repeated words, and words that may be classed as "asides" (e.g. comments to the rater or on the narrating task itself). A *T-Unit* is defined as "one main clause and all its attendant subordinate clauses and nonclausal units". A *filled pause* includes nonwords, e.g. *erm*, *er*, and *mm* (Lennon, 1990:406).

¹⁴ The scale follows the grading system traditionally used in the Netherlands.

might often be necessary. Moreover, such an overall grade might very well work in many contexts, in particular for high proficiency learners where it has been shown that subskills grow undistinguishable (de Jong & Verhoeven, 1992:9-10).

In Sundh (2003), the Swedish study mentioned in section 4.1.3, the informants participated in an OP assessment project carried out at Gothenburg University. The informants took three speaking tests which were similar in structure but different in content. Each test included two tasks of extensive speaking and one of interactive speaking. Nine experienced test instructors, who also functioned as raters, were in charge of the actual testing sessions (Sundh, 2003:103-105). A rotation system was adopted so that each student met as many different instructors as possible (Sundh, 2003:110-111). Student productions were graded on a five-point scale according to a set of grading criteria immediately after each test was finished. The instructors filled in an assessment form in which the students were first awarded an overall grade for the performance (i.e. holistic scoring); this was followed by the awarding of three factorial grades: (1) communication and fluency, (2) grammar and vocabulary, and (3) pronunciation. This procedure is opposite to the one used in de Jong and van Ginkel (1992) and the present study, where the raters first awarded factorial grades and then the overall grade (see also section 6.2.4). Sundh (2003:261) concludes that the overall and factorial grades were closely related; i.e., the raters often assigned the same overall and factorial grades to an individual student's performance.¹⁵ This is, perhaps, not very surprising (see sections 7.3.4 and 8.4 for the factorial grades in the present study).

Lorenzo-Dus (2007:220) investigated vocabulary in speech and showed that vocabulary is often assessed as part of a larger construct, namely the ability of the learner to interact with, for example, an examiner or his/her peer. As a consequence, the validity of the vocabulary assessment is affected by factors that go beyond the learners' lexical output per se. It should be noted that in my study, the validity of the vocabulary assessment (students' use of polysyllabic words in speech, see section 6.3.4) might have been affected in a similar way. Lorenzo-Dus (2007:226) also found that examiners are sensitive to the number of different words (types) that learners produce. In fact, the high-scoring learners had the highest number of rare words per examination. Rareness of vocabulary hence proved to have a significant effect (Lorenzo-Dus, 2007:227; see also Lorenzo-Dus & Meara, 2005).

¹⁵ I would have appreciated the inclusion of the values of the correlation coefficient regarding interrater reliability in the study.

Iwashita et al. (2008) investigated the relationship between detailed aspects/features of learners' L2 English and learners' holistic scores on a new international test of English for Academic Purposes, TOEFL ® iBT.¹⁶ From a large pool of pilot test data, 200 performances were selected and assessed, using a scale with five levels (1-5). The learners varied in terms of age and L1 but all were preparing for tertiary study in the USA (Iwashita et al., 2008:29). Learners did five OP tasks, all of which were of monologic nature (i.e. extensive speaking).¹⁷ In speech data, three conceptual categories were analyzed, the first two with several accompanying aspects of OP: (1) linguistic resources (aspects: grammatical accuracy, grammatical complexity, vocabulary), (2) phonology (pronunciation, intonation, rhythm), and (3) fluency (a conceptual category which was regarded as a single aspect).

Results were reported in terms of inferential statistics (see section 6.3.6) for most aspects using, for example, analysis of variance, ANOVA (see section 6.3.6.3) (Iwashita et al., 2008:34). It was found that learner speech varies according to proficiency level. Significant differences across levels of speakers in the expected direction were found for grammatical accuracy (all measures), grammatical complexity (verb phrase complexity, mean length of utterance), vocabulary (both token and type), pronunciation (target-like syllables), and fluency (speech rate, unfilled pauses, total pause time) (Iwashita et al., 2008:41). Furthermore, large standard deviations were found for most measures. That is, there was an indication of broad variation among learners assessed at any one level. There was also overlap between levels, or as the authors put it: “while the differences across level [sic] were real, that is, not attributable to chance, their impact on the overall level assigned to the test taker was not particularly strong” (Iwashita et al., 2008:41). Even more relevant for the present investigation is the fact that certain measures had a greater relative impact on overall scores, namely measures of grammatical accuracy (global accuracy), vocabulary (word type and token), pronunciation (target-like syllables), and fluency (unfilled pauses, total pause time, speech rate). Of these, the strongest were for vocabulary (token) and fluency (speech rate). In other words, *vocabulary and fluency seemed particularly important for the overall score* which was awarded. The authors conclude that “a combination of aspects determines the assessment of the overall proficiency of the learner” (Iwashita et al., 2008:43). Their findings

¹⁶ TOEFL ® Internet-based Test; the test measures skills in reading, listening, speaking and writing, and requires the test-taker to combine two or more of these skills to respond to a question (<http://www.ets.org/toefl/>, accessed Aug. 6, 2009).

¹⁷ This is a drawback of the study and the authors call for further studies investigating interactional data (Iwashita et al., 2008:45).

are clear evidence of the multi-faceted nature of OP. Finally, they also saw that problematic areas of speech were not homogeneous among learners. This means that at the macro level, some learners' pronunciation might be weak but vocabulary knowledge good, to give one example. From a rater's point of view, such learners' comprehensibility is problematic.¹⁸ Pronunciation can, therefore, act as a "first level hurdle" in assessment (Iwashita et al., 2008:44).

In sum, this chapter focused on OP in learner English and discussed relevant terminology, basic types of learner speech, and characteristics of learner speech and communication. Furthermore, results from previous research on OP and oral fluency were presented. Since speech was collected from students in dyads in the present study, a discussion of accommodation theory and the role of the interlocutor was also included. In the final section, assessment of OP and oral fluency was discussed and results from some studies presented. As was mentioned in chapter 1, my study aims to see whether EE has an impact on learners' OP and vocabulary. This chapter serves as a background to OP and chapter 5 as a background to L2 vocabulary.

¹⁸ *Comprehensibility* refers to the identification of a speaker's meaning, whereas *intelligibility* refers to the identification of a speaker's words (Lindeman, 2006:46)

5 Vocabulary in learner English

In the field of applied linguistics, a great deal of attention is given to research that addresses L2 vocabulary (Webb, 2009:405) and my study is yet another contribution. In the present chapter I give a theoretical background of ESL vocabulary and an account of current research on the topic. In the first section (5.1), there is a brief discussion of learner vocabulary in speech and writing, where differences in vocabulary due to the mode used are pointed out. After that, in section 5.2, the topic is receptive and productive vocabulary. I discuss what is meant by these terms and provide some empirical findings that are relevant to my own study. Section 5.3 accounts for various ways in which researchers have measured vocabulary in learner speech and writing. It includes two subsections, one which discusses problems of defining what a word is (5.3.1) and another which presents previous empirical studies on measuring second language vocabulary (5.3.2). The final section (5.4) presents English vocabulary from a frequency perspective; here I discuss core (frequent) and peripheral (infrequent) vocabulary. Core and peripheral vocabulary correlate not only with frequency, but also with the syllable structure of words, which is described in subsection 5.4.1, and with the etymology of words, the topic of subsection 5.4.2.

5.1 Learner vocabulary in speech and writing

The products of learners' speaking and writing can be directly observed by a teacher or researcher provided that the output of speech is recorded and the pieces of writing collected (D. H. Brown, 2004:118). In order to investigate learner vocabulary, quantitative statistics has commonly been used for vocabulary in writing, whereas it has less often been used for vocabulary in speech (Read, 2000:209). One reason for this is that in speech, full transcripts need to be produced before the analysis can even begin. Furthermore, the analysis itself is time consuming because the lexical items must be identified and classified. And, as Read (2000:213) also points out, once the calculating is done, the question remains what the statistics mean. Furthermore, the limited amount of research in the field of vocabulary in speech makes the interpretation of learners' results challenging. A reference point which can be used to interpret the results, however, is raters' judgments of the overall quality of learners' speech with the help of a rating scale of some sort (Read, 2000:213). In other words, such qualitative measures can be used to assess learners' vocabulary use along with other aspects of their spoken (or written) language.

The vocabulary used in speech differs in certain respects from the vocabulary used in writing. For instance, people use a smaller vocabulary in speech than they do in writing (see Nation, 2005:586; Read, 2000:236-237). What vocabulary is used depends on the context and also on the topic of conversation. For instance, in everyday conversations, most speakers would opt for a short native form when faced with a choice (e.g. *food* rather than *nourishment* or *nutrition*), whereas more formal contexts would include a greater proportion of foreign loanwords (Jackson & Amvela, 2000:46). Another difference between spoken and written vocabulary is the use of personal pronouns and verbs of communication, both of which are more frequent in speech than in writing (Biber et al., 1999:333-335, 371-375). In addition, in school settings, a learner may be able to utter an infrequent word in a sentence when required to do so by the teacher, but be reluctant to adopt the same word when left to his or her own devices, instead choosing to use a simpler word with a similar meaning. Such reluctance may very well be due to learners' lack of confidence (Laufer & Nation, 1999:37). It is also a fact that specific words need to be "fluently available" for learners, if fluent conversation (or reading) is expected (Laufer & Nation, 2001:7-8). That is, if too much time is spent on retrieving word meanings, comprehension will suffer (Laufer & Nation, 2001:9).

A characteristic of spoken vocabulary is the use of formulaic sequences, i.e. conventionalized language (Schmitt, Dörnyei, Adolphs, & Durow, 2004:55).¹ Even though the relationship between the size of learners' "individual word lexicon" and "formulaic sequence lexicon" is not straightforward, research has shown that relatively proficient learners (university level) know a great many formulaic sequences to begin with, and enhance their knowledge while studying English (Schmitt et al., 2004:64-69). In their study, Schmitt et al. (2004:64-69) found that the university students learned to master formulaic sequences productively, sequences which were previously only known receptively. Another finding was that some of the more transparent sequences were most likely explicitly taught (e.g. *first of all*, *on the contrary*), whereas less transparent ones (e.g. *in spite of*) had probably been acquired through exposure (Schmitt et al., 2004:64; cf. chapter 3). The use of formulaic sequences in speech is essential because they contribute to oral

¹ The term *formulaic sequence* is one of many which are used for the same concept. Other terms used are, for example, *lexical bundles* (Biber & Barbieri, 2007:263), *prefabs* (Erman & Warren, 2000:31), *multi-word expressions* (Knutsson, 2006:2), and *multi-word lexical items* (Read, 2000:231). For a review of formulaic language in SLA, see Weinert (1995).

fluency (Read, 2000:234; see section 4.1.3). It is worth noticing that vocabulary size also contributes to fluency in speech, and learners often associate progress in language learning with the acquisition of new words (Laufer, 1998:256), even though lexical knowledge and lexical use do not necessarily develop in the same fashion (Laufer, 2005:584-585).² The fact that learners associate progress with the acquisition of new words might be a mirror of many teachers' frequent use of vocabulary tests in ESL classrooms; it is easier to measure and assess vocabulary than to measure and assess language proficiency.

5.2 Receptive and productive vocabulary

In discussions of learner vocabulary, the terms *receptive* and *productive* may be used (see e.g. D. H. Brown, 2004). In the present section, focus is on the terms applied to vocabulary in a second language, and in such a context the two terms cover all the aspects of what is involved in knowing a word. The most general level of knowing a word involves the form, meaning, and use of the word (Nation, 2001b:26). A learner's receptive knowledge of a word implies, for instance, that the learner is able to recognize the word in its spoken and written modes. Moreover, the learner would know what the word means in a particular context and perhaps also recognize its typical collocations (Nation, 2001b:26).³ According to Nation (2001b:28), a learner's productive knowledge of the same word implies that he or she is able to say or write it. Furthermore, the learner would be able to use appropriate and correct forms of the word in both the spoken and written mode, and to use it in different contexts, only to mention a few characteristics of receptive and productive knowledge of a word (for a thorough description, see Nation, 2001b:26-30). It is worth pointing out that not all researchers include a "correctness of form criterion" in the productive knowledge of a word in the way Nation (2001b) does (for a discussion, see Read, 2000:25-28). For example, learners' appropriate use of a word in speaking and writing is interpreted as "enough" to show their capacity of productive knowledge of a word, i.e., even if the "correct form" is not used by the learner. In fact, minor spelling and grammatical mistakes of lexical items were *not* marked as incorrect in Nation (and Laufer's) own Productive Levels Test (Laufer & Nation, 1999:38-39). This means that in order for learners to get

² Laufer's conclusion is based on vocabulary in writing.

³ *Collocations* are words which co-occur within a certain distance of each other (see e.g. D. Gardner, 2007:257), e.g. *fast* is a collocation of its collocates *food*: *fast food* (i.e. not *quick food* or *speedy food*), (examples from Nation, 2001:56). Note that collocations do not necessarily have to be adjacent to each other: *commit a crime*, *commit an awful crime*, and *commit an extremely awful crime*.

credit for productive knowledge of a word in that test, it was indeed enough to show the capacity of producing the targeted word.

Sometimes, the terms *passive* and *active* vocabulary are used instead of receptive and productive vocabulary. The two sets of term both imply a clear cut distinction between on the one hand receptive/passive vocabulary, and on the other productive/active. Hedge (2000:116-117) suggests that this distinction should be removed and replaced by “a scale running from recognition of a word at one end to automatic production at the other, through intermediate stages of making greater sense of the word and how it might be used in different contexts.” The scale is described as a continuum, and Hedge (2000:117) says that knowledge of some words will remain at the recognition end (i.e. receptive/passive vocabulary) and will be called on in, for example, reading and listening, but might never become part of the learner’s productive/active ability. The view of learners’ vocabulary as a continuum is also adopted by, for example, Laufer (1991, 1998). It should be mentioned, however, that even the notion of the receptive-productive continuum is a simplified description of second language vocabulary development. Learners’ acquisition of lexical items “seems more complex and untidy” than that (Read, 2000:249; see also Vermeer 1992:148). In sum, L2 vocabulary acquisition is a very complex matter that cannot be easily described or explained.

Based on the results from several empirical studies, Nation (2001b:370-371) concludes that learners’ receptive vocabulary size is greater than their productive vocabulary size. Moreover, as learners’ vocabulary increases, the proportion of receptive vocabulary becomes greater. This means that the gap between receptive and productive vocabulary becomes greater at the lower-frequency levels. Still, a large proportion of high frequency vocabulary is known both receptively and productively. Finally, increases in learners’ vocabulary size (as measured by decontextualized vocabulary tests) are not necessarily reflected in an increase in vocabulary use (as measured by the proportion of low frequency words used in written composition) (Nation, 2001b:371). In other words, although different kinds of vocabulary knowledge are certainly related to each other, they develop in different ways. However, according to Laufer (1991:445), there appears to be a threshold for learners’ active vocabulary. Among advanced learners (university level students), she found that the productive lexicon grew only until it reached the average level of the group in which the students were required to function. In other words, the learners who had not yet reached their active threshold showed an increase in vocabulary, but this was not the case for the other students who did not need to increase their

lexicon in order to function in the group. Thus, the growth in active vocabulary for the group of students she studied was determined not so much by comprehensible input (cf. Krashen's input hypothesis) but by the sheer needs of the learner; a conclusion which rhymes well with Swain's output hypothesis (see section 2.1). The results of Laufer's (1991:446) study suggest that it is the necessity of using specific lexical items that drives learners to vocabulary enrichment and, therefore, she would like teachers to create situations in which learners find it necessary to use such items. Explicit vocabulary instruction and testing are suggested as potential incentives for learners. Finally, Webb's (2008) study on the receptive and productive vocabulary sizes of second language learners (83 Japanese students enrolled in English, university level) confirms several of Nation's (2001b) conclusions above, but Webb also found that receptive vocabulary size might be an indication of productive vocabulary size. In other words, learners with a larger receptive lexicon are also likely to know more of those words productively than learners with a smaller receptive lexicon (Webb, 2008:91).⁴ It is worth pointing out that Webb (2008:91) was able to conclude that receptive knowledge of the meaning of a word precedes productive knowledge.

The expected size of learners' vocabulary differs depending on what learners are investigated. Nation (2001a:176-177) reports that in countries where English is taught as a "foreign" language (e.g. Sweden), after five years of instruction of approximately five hours per week for about 40 weeks per year, learners would be expected to have become familiar with the high frequency words of English and to have worked on strategies for dealing with infrequent vocabulary. Students in such countries tend after five years to have a vocabulary of less than 2,000 word families.⁵ Laufer (1998) reports that high school learners in Israel (grade 10) had learned 1,900 English word families after six years of instruction. Then their passive vocabulary increased by an impressive 1,600 word families in one year of school instruction (grade 11). Similar progress was seen in controlled active vocabulary size, but the learners did not put this knowledge into use. Instead, free active vocabulary reached a plateau in the same year, something which Laufer (1998:266) suggests could be the result of a "lack of incentive to use more advanced, infrequent, and error-prone

⁴ Another study (Leiko-Szymańska, 2000) on receptive and productive vocabulary knowledge in advanced learners English (100 1st year and 69 4th year students of English, University of Łódź) only partly confirms Webb's results (2008) and instead shows that an increase in the receptive lexicon does not imply a growth of the productive vocabulary.

⁵ A *word family* consists of a head word, its inflected forms, and its closely related derived forms (Nation, 2001b:8) (see also section 4.3.1).

words.” There was, thus, a non-linear progress in second language passive vocabulary, which is in great contrast to the linear progress of approximately 1,000 word families per year in L1 acquisition (Laufer, 1998:266). For Norwegian learners in grade 8, Hasselgren (1996a:30) estimates a passive vocabulary of approximately 1,600 word families, which is similar to Nation’s (2001a) conclusion and also sounds reasonable in comparison with Laufer’s (1998) figures. In Swedish official documents in the 2000s, there is no explicitly stated number of English words that learners should supposedly have acquired by a certain age or grade (Håkansson, 2001:65; Skolverket, 2000), but it is reasonable to assume that Swedish learners of English are comparable to Norwegian and Israeli learners. That is, based on the discussion above, Swedish ninth graders can be expected to have a passive vocabulary of about 1,600–1,900 word families.

5.3 Measuring vocabulary

There are different ways to measure learner vocabulary depending on, for example, the object of inquiry and whether it is vocabulary in spoken or written English. In this section, I discuss measures of learners’ lexical variation, lexical sophistication, lexical richness, receptive vocabulary, and productive vocabulary. It should be mentioned that much of the following refers to vocabulary in written production, while my study on polysyllabic words is based on speech data and, as was pointed out in section 5.1, vocabulary in speech is generally less advanced than vocabulary in writing.

Measures of *lexical variation*, also called *lexical diversity*, focus on the range of learners’ vocabulary and the avoidance of repetition in production. Lexical variation is commonly measured with the type/token ratio; a transformation which compares the number of different lexical items (*types*) with the total number of lexical items (*tokens*) (Biber, 1988:238).⁶ Note, however, that this type of measure is unstable because it is affected by differences in text length/the number of words in the sample (Laufer & Nation, 1995:310; McKee, Malvern, & Richards, 2000:323; Vermeer, 2000:68).

Another object of inquiry might be students’ *lexical sophistication*, which is commonly defined as the percentage of “advanced words” in the text, where what is classified as “advanced” would depend on the level of the learners tested (see e.g. Laufer, 1991:442). For lexical sophistication, the type/token

⁶ The research community seems to agree on this definition of tokens and types because it is found in several sources, see e.g. Read (2000:17-18), Nation (2001b:7), and Malvern et al. (2004:19). However, what counts as a “word” varies in different studies (cf. discussion in section 5.3.1).

ratio is an insufficient measure. It needs to be supplemented with information about the quality of the words being used (Malvern et al., 2004:123).

The quality of words is closely linked to what is generally referred to as *lexical richness*. Measures of lexical richness attempt to quantify the degree to which a learner is using a varied and large vocabulary (Laufer & Nation, 1995:307). Laufer and Nation (1995) have proposed their *Lexical Frequency Profile (LFP)* as a measure of lexical richness in L2 writing. The LFP shows the percentage of words at different vocabulary frequency levels that a learner uses in his or her writing. Even though it was originally designed for ESL, the LFP has also been applied in other contexts (Malvern et al., 2004:128). The theory behind the LFP is that less proficient learners are less likely to use rare vocabulary than advanced learners. The LFP has been shown to be a reliable and valid measure of lexical use in writing because it correlates with other established measures of lexical knowledge, for example the Productive Levels Test (Laufer & Nation, 1995:313, 319). Laufer and Nation (1995:319) conclude that it is reasonable to expect learners' vocabulary size as measured by a vocabulary test to be reflected in the learners' productive use of the language (see also Vermeer, 2000:70).

Lexical richness also includes learners' use of infrequent vocabulary ("difficult words") because infrequent vocabulary is closely related to the concept of word quality. In a second language context, the use of low frequency words is, not surprisingly, interpreted as mastery of "difficult" words in the target language (Vermeer, 2000:79). Based on spontaneous speech data from 70 Dutch native and 76 ethnic minority children aged 4–7, Vermeer (2000) shows that the type/token ratio mentioned above lacks both validity and reliability as a measure of lexical richness. Instead, he suggests that the degree of the difficulty of words used by informants should be measured by relating the words to their frequency ranking in various corpora (Vermeer, 2000:79). In the same study, it is also shown that simple counts of types and lemmas are the best measures of lexical richness.⁷ However, especially in the early stages of acquisition, Vermeer (2000:78-79) questions a lemma-based analysis because learners would be likely to store lexical items such as *go*, *went*, and *gone* separately (as three words) rather than together (as one word). Thus, a word-type-based analysis (i.e., where *go*, *went*, and *gone* count as three words) could be a solution (Vermeer, 2000:78; see also Vermeer, 2004:175). Furthermore, Vermeer (2000:79) claims that it is essential to control the tasks that are used in studies whose purpose is to

⁷ A *lemma* is "a set of lexical forms having the same stem and belonging to the same major words class, differing only in inflection and/or spelling" (Francis & Kuera, 1982:1). The concept of *lemma* is discussed in section 5.3.1.

measure vocabulary based on speech data. That is, it is important that the topics are similar for all informants in a study. In addition, a written vocabulary task (or test) should be included in all studies on spoken language, so that results may be compared and validated. As a matter of fact, including several measures of vocabulary in a study provides a more comprehensive and useful picture of learners' vocabulary knowledge (Nation, 2007:42). Two written tests which are commonly used in studies on vocabulary are the so-called Levels tests, first developed by Paul Nation, and then by Paul Nation in collaboration with Batia Laufer.⁸

The Vocabulary Levels Test (VLT) was originally designed for diagnostic purposes to be used in the teaching of English. This test measures the size of learners' receptive (passive) vocabulary (Nation, 1983:14; 2001b:416-424). It is a monolingual test; i.e., it is all in English. The first version of the VLT was divided into five levels, where each level in turn was connected to a certain word frequency level. The first level in the VLT tested words from the second most frequent 1,000 word families ("2,000 Level") and the second level tested words from the third most frequent 1,000 words ("3,000 Level"). The third level tested words at the "5,000 Level", which was considered to be on the boundary of high and low frequency vocabulary. The fourth level tested specialized vocabulary used in university textbooks; the test items came from the University Word List.⁹ The fifth level, finally, was aimed at testing infrequent vocabulary ("10,000 Level") (Nation, 1983:14-15). Thus, learners are likely to know less at each succeeding level of the test since the words become more and more difficult/infrequent (cf. Fitzpatrick, 2007:126). In the original as well as in current versions of the VLT, each section consists of six words and three definitions and test-takers are supposed to make the right combinations. The lexical items put together in each section are not related in their meanings (Nation, 1983:14-15).¹⁰ The original version of the VLT has been improved and made available in several equivalent forms (see Nation, 2001b:416-424). The first section of the first level ("2,000 Level") today looks like this:

⁸ Other researchers were also involved in the production of the tests, see Nation (2001b:416).

⁹ Nation (1983:16) describes the *University Word List (UWL)* as a list which contains the specialized vocabulary frequently used in university texts (e.g. *complement, emerge, generate*) but not elsewhere. The UWL was published in 1984 (Xue & Nation, 1984) and it contained 836 words (Laufer & Nation, 2001:10). In 2009, the UWL includes 850 head words of word families and is available online at <http://www.lexrutor.ca/> (retrieved May 29, 2009). Coxhead (1998) has compiled the Academic Word List which includes 570 head words of word families (available at the same site).

¹⁰ A shortened version of the VLT was used in the present study (see section 6.2.5 and App. 14).

1	copy	
2	event	_____ end or highest point
3	motor	_____ this moves a car
4	pity	_____ thing made to be like another
5	profit	
6	tip	

The Productive Levels Test (PLT) was modeled on the VLT and designed to measure students' productive (active) vocabulary ability (Laufer & Nation, 1999:36). More specifically, it measures learners' "controlled productive ability", which is the ability to use a word when compelled to do so by a teacher or researcher, whether in an unconstrained context (e.g. a sentence-writing task), or in a constrained context (e.g. a fill-in task where a sentence context is provided and the missing target word has to be supplied) (Laufer & Nation, 1999:37). The latter format is used in the PLT.¹¹ The first word to be tested is *opportunity* and the provided constrained context is as follows: *I'm glad we had this opp_____ to talk.* The constrained context used in the PLT is also used in standard cloze tests, i.e., a text passage in which certain words have been deleted according to a fixed ratio (e.g. every seventh word) and the test-taker is supposed to fill in the missing words in the blanks.¹² There is, however, an important difference between the PLT and a standard cloze test in English because the cloze test supposedly measures overall English proficiency, not only productive vocabulary (D. H. Brown, 2004:9; Read, 2000:102).

As mentioned above, the PLT was modeled on the VLT. It samples 18 items at each of the word frequency levels described above, from the 2,000 Level to the 10,000 Level. Based on learners' results on the various frequency levels, it is possible to decide whether a learner has satisfactory mastery of a particular level. Such a decision, however, depends on the relevant educational level of the learner (Laufer & Nation, 1999:41). Both the VLT and the PLT are, thus, useful in teaching, but they may also be used for the purpose of research since they have been validated and found to be reliable tools of measure (Laufer & Nation, 1999:44; Nation, 1983:370). Even so, there has been some criticism of the tests. For instance, Fitzpatrick (2007:119) considers the PLT extremely context-dependent, and Paul Nation himself shows that the monolingual format of the VLT might be too difficult for lower proficiency learners, because lower proficiency learners in fact scored higher on a bilingual version of the

¹¹ A shortened version of the PLT was used in the present study (see section 6.2.5 and App. 12).

¹² There are also so-called rational cloze tests in which there are selective deletions of words (Read, 2000:106-107).

VLT which he piloted, than they did on the monolingual version.¹³ He stresses, however, that bilingual versions of the VLT would only work as a measure in settings where all learners share the same first language (Nation, 2007:40-41).

Despite some criticism, both Levels tests are suitable to use for comparison and validation of measures of learners' vocabulary based on speech data in the way Vermeer (2000) advocates. Learners' vocabulary may be compared with frequency rankings so that the quality of the words used may be assessed (Vermeer, 2000:79). Another, perhaps more direct, way to identify the quality of learners' vocabulary is suggested in a study by Biber (1988). In addition to the type/token measure, Biber (1988) also examined word length (measured by the mean number of orthographic letters) in both speech and writing.¹⁴ Word length reflects usage of words with more specific meanings; i.e., it is connected with the quality of lexical items, with advanced/infrequent vocabulary (Biber, 1988:104-105, 238-239) (cf. section 4.4.1 on polysyllabicity). In addition, it marks high density of information and very precise lexical choices, something which results in exact presentations of informational content (Biber, 1988:104). Word length per se is, thus, seen as a "proxy measure for rarity by appeal to Zipf's Law that shorter words are more frequent and more general in meaning" (Malvern et al., 2004:124; see also Gardner, 2007:253).¹⁵

The present section opened with a discussion of the measurement of lexical variation and closes on the same topic. According to Malvern et al. (2004:125), there are intrinsic and extrinsic measurements of lexical variation. *Intrinsic measures* rely exclusively on information contained within the language sample; the type/token ratio and its transformations are examples of such a measure. *Extrinsic measures* relate the contents to external criteria, for example frequency data. Rarity measures, that is, measures involving infrequent vocabulary, can be either intrinsic or extrinsic. An *intrinsic rarity measure* draws on words that occur only rarely in the text/speech sample itself, for example rare words as a percentage of types (Malvern et al., 2004:126). An *extrinsic rarity measure* defines words through reference to external criteria such as word frequencies in the BNC corpus, which was used by Afitskaya (2002) to distinguish between basic and advanced vocabulary in speech among adult

¹³ The piloted version of the VLT was "easier" than the original version because it included test items at the 1,000 Level. Because a definition of a test item cannot include vocabulary beyond the level of the test item, students' first language had to be used in the definitions, hence the bilingual format (Nation, 2007:40-41).

¹⁴ Many other measures were also used in Biber's (1988) study, but they are not relevant for the present study.

¹⁵ The American linguist George Kingsley Zipf (1902-1950) did some of the earliest investigations of statistical regularities in language, e.g. that word length is inversely proportional to frequency of usage (Malvern et al., 2004:202).

learners (see section 5.3.2). Regardless of what type of measurement is used in vocabulary research, it is usually important to establish how “a word” is defined, which is the topic of the following section.

5.3.1 *What is a word?*

Word frequency studies cannot be done without deciding what counts as a word. The definition of a word is generally equated with head words in a dictionary. A *head word* may also be termed *lexeme* or *lexical item*. A *lexeme* is a unit of lexical meaning which exists regardless of any inflectional endings it may have or the number of words it may include (see e.g. Crystal, 2003:118). An example is the lexeme *go*. The three words *goes*, *going*, and *gone* would all be assigned to the lexeme *go* and thus not count as three words but as one. This means that none of *goes*, *going*, and *gone* would occur as a head word in a dictionary but instead be listed under the lexeme *go*. But what about the past tense form of the same verb, i.e. *went*? Should it be assigned to the lexeme *go* or should it be its own lexeme and thus be counted as another word? To cover such issues, another concept of “word” called *lemma* may be used in word frequency studies. The definition of a lemma is very close to that of the lexeme. A *lemma* is said to be “a set of lexical forms having the same stem and belonging to the same major words class, differing only in inflection and/or spelling” (Francis & Kučera, 1982:1). What lies behind the use of lemmas as the unit of counting is the idea of learning burden (Nation, 2001b:7). The learning burden of a lexical item is the amount of effort required to learn it. An irregular verb form such as *went* (also referred to as *suppletive inflection*), which was mentioned above, has a heavier learning burden than a regular verb form and is, therefore, counted as another lemma (another word) separate from the base word, in this case *go*. For the same reason, it is reasonable to count *went* as a separate lexeme/word in L2 vocabulary, at least for young learners. In research, the term *lemma set* is sometimes used, and an example of such a lemma set would be *climb*, *climbs*, *climbing*, and *climbed* (all verbs), whereas the noun *climber* would be considered another lemma (D. Gardner, 2007:44). Regardless of what terminology is used (lexemes or lemmas), it is essential to state clearly on what basis words are defined/counted in word count studies. Crystal (2003:123) describes the problem of defining words in his discussion of the size of William Shakespeare’s vocabulary. In some accounts, Shakespeare is claimed to have had a vocabulary of almost 30,000 words, a figure based on the fact that the famous author used 29,066 different words (i.e. types) in his works and 884,647

words in all (i.e. tokens). However, when Shakespeare's lexemes are counted, his lexicon falls to less than 20,000 words (Crystal, 2003:123).

Compounds in writing may pose a problem in word frequency counts because compounds are two or more words that are combined to form a unit of meaning. Moreover, compounds may be represented as two orthographic words (*window shopping*), as one unbroken orthographic word (*database*), or as a hyphenated word (*eye-witness*) (Biber et al., 1999:325-327), so spelling does not always facilitate one's decision. Compounds in speech may also pose problems in frequency counts. For example, in the case of noun plus noun sequences, it may be difficult to decide whether it is a compound made up of two nouns (which would count as one word in a frequency count) or rather a sequence of a noun modifier followed by a noun head (i.e. two words in a frequency count). In such noun plus noun cases, stress placement is helpful. A compound is usually characterized by initial stress (*'heart attack*, primary stress on *heart*) whereas primary stress on the second element is an indication of a noun modifier followed by a noun head (*glass 'bottle*) (Biber et al., 1999:589-590).

Thus, it is anything but an easy task to define what a word is. Nevertheless, as I mentioned above, it is essential to do so in vocabulary studies where word counts are relevant. For instance, one aim of the present study is to investigate learners' use of peripheral (advanced/infrequent) vocabulary in speech. Therefore, it is necessary for me to define what counts as a peripheral word (see section 6.3.4).

5.3.2 Previous empirical studies measuring vocabulary

In a study of thirty non-native speakers of English enrolled at the University of the West of England (15 intermediate level; 15 advanced level), Afitskaya (2002) assessed vocabulary richness in learner speech. In her investigation, she made use of the 4,500 words Rank Frequency List (BNC corpus) provided in Leech, Wilson, and Rayson (2001). Given that the subjects were non-native speakers and there was a need to discriminate between basic and advanced words in the analysis, the dividing line between basic and advanced vocabulary was set to 1,000 words in that study. That is, the 1,000 most frequent words of the above-mentioned corpus were considered basic and all other words advanced (cf. section 5.4, Core and peripheral vocabulary). Afitskaya's aim was to find a reliable measure to estimate vocabulary size. She took both quantitative and qualitative properties of words into account while examining two transformations of lexical sophistication (LS): LS* equaled advanced tokens per all tokens, and LS** equaled advanced tokens per basic tokens. Afitskaya

(2002:8) found that the two measures were virtually identical and that the advanced students scored significantly higher than the intermediate students. In the present thesis, investigations of learners' lexicon are also linked to the notions of basic/core and advanced/peripheral vocabulary (for methods, see section 6.3.4 and for results, sections 8.1-2 and 8.7).

There are other empirical studies which involve investigations of learner vocabulary similar to the present study. Linnarud (1986), for example, did pioneering work in her thesis where she used written compositions to study second language learners' content words that were exclusive to one writer, which resulted in measures of *lexical individuality* (originality of content words) and *lexical sophistication* (difficulty of content words) (Linnarud, 1986:5). In addition to the study of content words, she included three non-lexical factors in the investigation, namely the number of words in the text, the number of errors in the text, and the number of words per sentence. Similar to Linnarud, I will also address a certain type of rare words, namely students' "own" polysyllabic types (for details, see section 6.3.4). One obvious difference between Linnarud's (1986) study on learner vocabulary and mine is that my study is based on speech data. Similar to Linnarud (1986), Malvern et al. (2004:123) also investigated lexical sophistication in writing. They point out that lexical sophistication may be reflected in the proportion of rare words, which is described as one dimension of lexical richness in a second language.

A final empirical study that I would like to comment on is based on speech data from 34 British secondary school students taking their oral examination in French for the General Certificate of Secondary Education (Malvern & Richards, 2002). In one part of that study, focus is on assessment of lexical richness. It was found that the teachers who rated their own students' oral productions did not discriminate between vocabulary use and other areas of proficiency. However, at the same time, the researchers conclude that teachers are likely to respond to aspects of lexical richness such as the use of low frequency words, or words which are at least less common in foreign language classrooms (Malvern & Richards, 2002:95).

5.4 Core and peripheral vocabulary

Minkova and Stockwell (2006:463) discuss English words and their frequency using the terms *core* and *peripheral vocabulary*. The *core* is made up of the most frequent words, words that are absolutely essential and without which basic communication and sentence composition would be impossible. Outward from that core lies *peripheral* vocabulary, that is, words of decreasing frequency and

familiarity. The core-periphery distribution of the English vocabulary can be described with the help of concentric circles with the nucleus as the core (see Fig. 5.1).

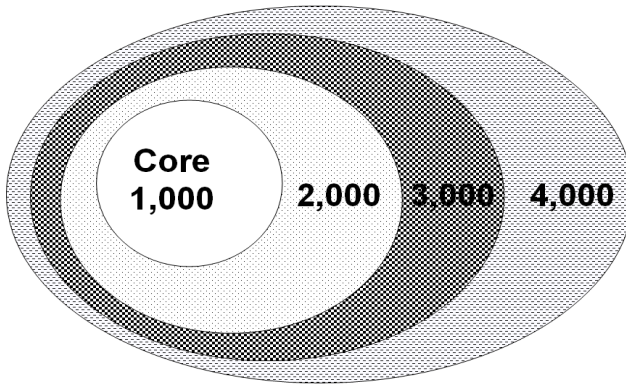


Figure 5.1. *The core-peripheral distribution of the lexicon (based on Minkova & Stockwell, 2006:464).*

The majority of the words in the core are native English words; i.e., they are the oldest words in English, found in written accounts as early as in the Old English period. Minkova and Stockwell (2006:463-464) describe core vocabulary as “the structural backbone of syntax”, consisting of articles, conjunctions, prepositions, auxiliaries, pronouns, quantifiers, and determiners. Necessarily, such items rank highest in all word frequency studies independent of which corpus the studies are based on.¹⁶ Frequency rankings are headed by function words (e.g. *the, be, of, and, a, and in*) (Minkova & Stockwell, 2006:463-464). The content words in the core cover basic objects, actions, and concepts of daily life, such as *hand, food, wife, stone, go, sing, eat, see, sleep, good, wise, cold, and sharp* (Williams, 1975:44). These concepts are so basic that they can most likely be found in any language. The theory is that such words are passed on from generation to generation for centuries, and even though pronunciation and spelling of individual words may change over time, they are basically the “same” words whose meanings have remained fairly stable (Jackson & Amvela, 2000:23-24; Williams, 1975:43-44). Peripheral vocabulary, in contrast, is more

¹⁶ Commonly used corpora in frequency studies are, for example, the Brown Corpus (American English; 1,014,000 words from 500 samples of texts from a broad range of genres), the Lancaster-Oslo (LOB) corpus (British English; comparable in size and text types to the Brown Corpus), and the British National Corpus, BNC (a hundred times larger than the Brown Corpus and LOB, and covering several varieties of both spoken and written English) (Leech et al., 2001).

advanced. In peripheral layers, there are words from, for example, the realms of ideas, art, science, and technology. These words are more often loanwords than words in the core are, and they may also appear more learned (Jackson & Amvela, 2000:34-35; Minkova & Stockwell, 2006:467). Moreover, the meaning of such words may change over time in a way that the meaning of native core words generally does not do. For instance, the word *computer* was adopted in the 17th century from French and meant ‘a person who makes calculations or computations’. However, three hundred years later *computer* has taken on another meaning, albeit related to the original one: ‘an electronic device (or system of devices) which is used to store, manipulate, and communicate information, perform complex calculations, or control or regulate other devices or machines, and is capable of receiving information (data) and of processing it in accordance with variable procedural instructions (programs or software)’ (*Oxford English Dictionary Online*, 1989; accessed May 11, 2009).

Minkova and Stockwell (2006:465) stress that there can be no absolute ranking of words, and other researchers have come to the same conclusion (see e.g. Nation, 2001a:180). In the foreword to *Word Frequencies in Written and Spoken English Based on the British National Corpus*, the authors make the same point (Leech et al., 2001:xi) and add the importance of acknowledging different frequency profiles for different varieties of language, such as spoken and written language (cf. section 5.1). Even so, word frequency counts based on different corpora are similar enough to support the use of the concepts *core* and *peripheral vocabulary* (Minkova & Stockwell, 2006:466; Nation, 2001a:179). There is, however, always a problem as to where to draw the line between core/basic and peripheral/advanced vocabulary.

5.4.1 The syllable structure of English words

The notion of core and peripheral vocabulary does not only correlate with frequency of usage, but also with syllable structure. In other words, with regard to the number of syllables in English words, the majority of the words in the core are made up of either one syllable (a monosyllabic word) or two (a disyllabic word). In contrast, the proportion of words that are monosyllabic in peripheral layers of the lexicon is relatively low. Consequently, the proportion of words with two syllables or more in peripheral layers of the lexicon is larger than the proportion of monosyllabic words (Minkova & Stockwell, 2006:465-467).

The model of *Consonant-Vowel tiers* (*CV-tiers*) is one way to describe what a syllable is. According to CV-tier theory, the V element represents the syllable

nucleus, i.e. the peak of sonority (Katamba, 1989:156-159). The C element represents what is called the syllable *onset* (less sonorous sound preceding the nucleus) or the syllable *margin* (less sonorous sound following the nucleus); that is, the C element is not the peak of sonority (Katamba, 1989:157). Crystal (2003:246) uses a simplified model of CV-tiers in his discussion of syllables. Even though there are some English words or word-like sounds that consist of only one sound, for example *I*, *eye*, and *oh* (each with the syllable structure V), most English words are combinations of Vs and Cs. For instance, the word *go* has the syllable structure CV and *up* has VC, whereas *cat* corresponds to CVC and *screw* to CCCV. Despite the different number of letters in these words – *go*, *up*, *cat*, *screw* – they are all examples of *monosyllabic words*, i.e., words that consist of one syllable only (Crystal, 2003:246). Each word has one nucleus, one peak of sonority (V).

Most words in the language have more syllables than one – they are *polysyllabic*. A word with two syllables, for example *copy* (CVCV), is commonly referred to as a *disyllabic word* (see above), whereas a word with three syllables is referred to as a *trisyllabic word*: *basketball* (CVCCVCCVC). With regard to my investigation of vocabulary in learner speech, the analysis that relates to the syllable structure of words focuses on words which include *three syllables or more*. By choosing to study words with three syllables or more, I bring peripheral vocabulary to the center of attention since peripheral vocabulary tends to be polysyllabic. If the method works, it would be a fairly straightforward way of identifying advanced vocabulary in learner speech. Thus, the design of my vocabulary investigation takes a so-called intrinsic rarity measure of lexical variation into account; it draws on words that occur only rarely in the text/speech sample itself (cf. Malvern et al., 2004:126).

In a final comment on syllables I would like to point out that sometimes it can be difficult to decide on the number of syllables in a word. Crystal (2003:246) uses the word *meteoric* as an example. Does *meteoric* consist of three (*me-teo-ric*) or four (*me-te-o-ric*) syllables? The number of syllables in *meteoric* may depend on accent, speed of speech, level of formality, and context of use, Crystal concludes. My approach to identifying which words in my speech data include three syllables or more is described in section 6.3.4.

5.4.2 The etymology of English words

By adding an etymological dimension to my study of polysyllabic words, I hope to deepen my analysis (see section 6.3.4). The notion of core and peripheral vocabulary correlates with etymology. This means that core words are

etymologically near-homogeneous, while peripheral words are more diverse in terms of their origin (Minkova & Stockwell, 2006:465-467).

As was discussed above, the core vocabulary is predominantly native. Williams (1975) conducted a large-scale vocabulary study on the etymological composition of English in relation to the relative frequency of words. The results were based on a corpus of more than fifteen million running words. Over half of the words were recorded in business and personal correspondence which reflected ordinary everyday activities.¹⁷ Based on frequency counts, the first one thousand words made up the core, and 83% of the words in the core were descendants of Old English, a Germanic language (Williams, 1975:67) (see Table 4.1). Minkova and Stockwell (2006:466) point out that the percentage of Old English words in the core is not static; core vocabulary has been changing in favor of borrowings. Nevertheless, the core is made up of function words and words used in everyday conversations. Native words are, therefore, stable in the core. In peripheral vocabulary, that is, from the 2,000 layer and onwards, the proportion of native words is much more unstable. In fact, there is a dramatic drop in the ratio of English words already in the 2,000 layer, from 83% to 34 (Williams, 1975:67). In contrast, the non-Germanic language French makes up 11% of the core vocabulary but as much as 46% of the next layer. It is a fact that in peripheral vocabulary, the proportion of native English words decreases and the proportion of words adopted from the two Romance languages Latin and French, or other languages, increases (see Table 4.1). Words assigned to “other languages” mostly include words of mixed or doubtful origin, but also words from languages such as Greek, Dutch, Italian, Spanish, and German.¹⁸

¹⁷ Data were originally compiled by A. Hood Roberts (1965).

¹⁸ The influx of foreign words into the English language is well described in, for example, Jackson and Amvela (2000).

Table 4.1. The origin of words in core (the first most frequent thousand words) and peripheral vocabulary (the second most frequent thousand words, the third most frequent thousand words, and onwards) (based on Williams, 1975:67-68).

Vocabulary	Origin		
	English (%)	Latin and French (%)	Other languages (%)
Core	83	13	4
2,000	34	57	9
3,000	29	60	11
4,000	27	62	11
5,000	27	64	9
6,000	27	61	12
7,000	23	62	15
8,000	26	59	15
9,000	25	58	17
10,000	25	60	15

To summarize, this chapter has given a background to the parts of my study which relate to vocabulary. Among other things, I discussed differences between vocabulary in speech and writing as well as differences between receptive and productive vocabulary. Moreover, this chapter accounted for measures of lexical variation, lexical sophistication, lexical richness, receptive vocabulary, and productive vocabulary. As suggested by Vermeer (2000:79), all informants should have similar topics in speaking tasks/tests when their output is to be analyzed and compared, a conclusion my research design will adhere to. Moreover, the chapter presented the notion of core and peripheral vocabulary. Core and peripheral vocabulary relate to word frequency, syllabicity, and etymology. I found that for my investigation of learner vocabulary in speech, a suitable approach would be to identify and analyze learners' use of peripheral (advanced/infrequent) vocabulary. In the present study, words with three syllables or more will be targeted and used in subsequent word counts and etymological analyses. Moreover, written vocabulary tasks/tests for all informants should accompany tests of speech, so that results may be compared and validated (Vermeer, 2000:79). For that purpose, the Vocabulary Levels Test and the Productive Levels Tests (Laufer & Nation, 1999; Nation, 2001b) are considered suitable in the present study. In the next chapter, I will account for the methods in depth.

6 Material and methods

In this chapter I present the design and procedure of my study. Furthermore, I comment on how my sample was selected and the generalizability of my results. The material which served the basis for my study is presented, followed by an account of the methods of analysis. In conjunction with the presentation of the types of material and the various methods of analysis used, concerns with regard to reliability and validity are discussed. The chapter closes with a section on the statistical methods used in the present study.

6.1 Design and procedure

In this section, an account of my pilot studies is presented, followed by an outline of the main study and the criteria for participation. I then describe my sample. A brief section on how the study was administered comes next, followed by a discussion of ethical considerations regarding the current project. The section ends with a discussion of the mixed methods research design.

6.1.1 *Pilot studies*

Two pilot studies were carried out in preparation of the main study, the first in spring 2005 and the second one year later. In the first, I interviewed four students in grades 11 and 12 (upper secondary school) who had excelled in English but not in other subjects, at least not according to their subject grades. Five topics were covered in the interviews: (1) views on English in school and (2) outside school (cf. Pearson, 2004); (3) acquisition of new vocabulary; (4) dreaming in English; and (5) beliefs about language learning.¹ The purpose of the interviews was three-fold: (1) to serve as an aid in formulating the research question for the main study, (2) to practice conducting interviews and to determine whether interviews would be a suitable method for the main study, and (3) to see what the informants had in common with regard to the five topics. The results of the pilot study made it possible to formulate the research question. Also, interviews proved to be a useful method, particularly with regard to learner strategies and EE, but actual EE data obtained from interviews were considered insufficient and unreliable as a sole measure of

¹ The question about dreaming in English was included for two reasons. First, I thought that it would be a topic which would make the informants talk. Second, based on my personal belief that it is only possible to dream in English for a language learner if a high enough level of proficiency has been reached, I was simply curious to see how these informants would respond. The informants' responses regarding their beliefs about language learning were collected with the help of a questionnaire, BALLI (cf. Horwitz, 1987), which they filled in at the end of the interview (see section 3.1.2).

informants' EE. Thus, interviews would need to be complemented by additional data, hence the inclusion of a questionnaire and a language diary in the main study (see sections 6.2.1–6.2.2). I also found that the informants' interest in English had begun very early, even before starting school, and they had all been engaged in various EE activities over the years.

The second pilot study involved trying out test versions of the questionnaire and the language diary. In addition, various speaking tests were piloted, as were a number of written vocabulary tests. The overall aim was to identify measuring tools which would be suitable for inclusion in the main study. Another purpose was to practice using the technical equipment and transcription software for the speech data, Transana (Fassnacht & Woods, 2006). In the following, I describe how the second part of the pilot was carried out and what conclusions I was able to draw from it.

The questionnaire included 27 questions and was designed to provide general background information about the informants along with more specific information about their contacts with EE, involvement in various EE activities, motivation for learning English, and views on English.² Many of the questions were identical with or similar to items used in previous studies (Oscarson & Apelgren, 2005:110-118; Sylvén, 2004:280-288). It was piloted with five 9th-grade students who volunteered to fill it in and comment orally on it.³ The questionnaire worked well in the sense that the informants understood the questions and only had comments with regard to the layout.

The language diary was developed in collaboration with Liss Kerstin Sylvén, Gothenburg University, and was tried out in a group of 29 ninth graders.⁴ It comprised eight pages: the title page followed by seven pages, one for each day of the week, where students filled in time spent on seven activities plus an open final category (cf. e.g. Forsman, 2004; Pearson, 2004), for linguistic activities both in English and in Swedish. The diary was in Swedish and I gave the group oral instructions about how to fill it in and provided each informant with one diary and a pen, plus an envelope with my name, address and prepaid postage. They were encouraged to fill in the diaries on a daily basis, preferably in the evening before going to bed, and to post them at the end of the week. They were also told that I would visit them again for follow-up questions. The result was satisfactory with regard to how the diaries were filled

² Liss Kerstin Sylvén, Gothenburg University, was consulted in the development of the questionnaire. I am very grateful for all her help.

³ A teacher I know taught the informants; they filled the questionnaire in during an English class.

⁴ A teacher I know taught the group, but it was not the same teacher as the one referred to in footnote 3.

in but the response rate was considered too low (45%). At the subsequent visit, the informants more or less unanimously suggested that the diaries should be handed in directly to the teacher instead because such a procedure would increase the response rate: “It would feel more like homework even though it is not”, as one informant said. Another suggestion was to add “other languages” to cover all spare time language use, and also to add the possibility of filling in whether subtitling was used (and if so, in what language) when watching films or TV. Since all suggestions were good, the language diary was revised accordingly. I also decided to have the teachers collect the diaries in the main study, even though it would mean extra work for them, something which I was initially reluctant to suggest. However, if the teachers’ collection of diaries would increase the response rate, it was essential to have them do so.

The nationwide goals regarding speaking in English that Swedish students should have attained by the end of grade 9 are that they should:

- understand clear speech (...) concerning familiar conditions and their [i.e. the students’] own areas of interest,
- be able to actively take part in discussions on familiar topics and with the help of different strategies communicate effectively,
- be able to orally relate and describe something they have seen, heard, experienced or read, as well as express and give their reasons on how they understand a topic that is of personal importance.⁵

In order to measure oral proficiency (cf. my main research question) in line with the national goals to be attained, speaking tests with a dyadic set-up and interactive tasks were considered suitable. In addition, in order for my study to appeal to both teachers and students, I included tests that could have been part of any teacher’s regular planning. That is, if it were not for my study, the teachers might have used the speaking tests anyway. However, normally, it would have been impossible for the teachers to listen carefully to each dyad due to time constraints. Thus, I tried to design the study so that participation would be a win-win situation for teachers and students: teachers would have access to more speech data for each individual student than they normally would have (thanks to the recordings, see section 6.1.2) and students would have several

⁵ From <http://www3.skolverket.se/ki03/front.aspx?sprak=EN>, accessed Aug. 20, 2009.

opportunities to practice speaking in English and show their speaking skills to their teachers.

One type of test that meets the stipulated criteria is the national test of English used in Sweden. In general, English teachers in Sweden prepare their students for the national test by letting them practice using old versions in class. Teachers consider the national test helpful in assessing students' various performances and needs (Erickson & Börjesson, 2001:262).⁶ Four old versions of the national test (three intended for grade 9, one intended for English A/grade 10) were included in my pilot study along with a national test of English used in Norway (grade 8) (see Hasselgren, 1996b), and a test I prepared myself called "Expository speech".⁷ In total, eleven test sessions were carried out thanks to twelve students who volunteered to participate. The tests were taken while the students were in school but their participation was viewed as an extracurricular activity. However, they had the option of deciding whether a copy of their recording should be given to their English teacher or not. All informants were in ninth grade and their parents' written consent had been collected beforehand. Each test session ended with a short interview in which I asked for the students' opinions on the level of difficulty of the test and if anything needed to be clarified. Some students took two or more tests. In those cases, I asked them to compare different tests, to see what test(s) they preferred and why.

With regard to written vocabulary tests, four types of test were piloted. Two were shorter versions of the Productive Levels Test and the Vocabulary Levels Test from Nation (2001b) (see section 5.3). The third test was a rational cloze test (see chapter 5, footnote 12) called "Two creatures of the past – mammoths and mastodons" (multiple choice) provided free online by the University of Cambridge as practice material for the First Certificate in English in spring 2006.⁸ From the same source, the fourth test called "Find the correct form" (rational cloze test, base forms of the targeted words provided) was also retrieved. The four tests were tried out at two different schools among five groups of students in grade 9 and one group in grade 8. Their English teachers administered the tests and were given written instructions to ensure similar testing conditions. The Productive Levels Test was taken by 96 students and

⁶ Old tests are available at http://www.ipd.gu.se/enheter/sol/nafs/exempel_provuppgifter/ or via <http://www.skolverket.se> (both accessed July 6, 2009).

⁷ The project leader of the national tests of modern languages in Sweden at Gothenburg University, Gudrun Erickson, was consulted in the test selection process. I am very grateful for her thoughtful advice.

⁸ University of Cambridge, ESOL examinations: <http://www.cambridgeesol.org/exams/general-english/fce.html> (accessed July 6, 2009).

the mean score was 21.6 (max: 45). Based on the teachers' comments on the test, their students' test scores, and my own judgment, I decided to include the test in my main study. However, I found two items difficult to score and decided to replace them in the version to be used in the main study.⁹ I also decided to replace a third item (which had a Swedish cognate) with another target word from the same word frequency level.¹⁰ The mean score on the Vocabulary Levels Test was 60.0 (max: 90; N=62) and scoring was unproblematic. Again, for the same reasons as the ones mentioned above, I decided to include the piloted VLT (as it was) in my main study, and also to provide the teachers with some enlarged copies of the test in case there were students who needed large print.¹¹ The rational cloze test (multiple choice) was taken by 45 students and the mean score was 9.6 (max: 15), whereas the rational cloze test (with base forms of target words provided) was piloted in one class only (N=22). The mean score was low (2.6/10), which indicated that the test was too difficult for Swedish ninth graders. I decided to exclude both cloze tests from the main study and instead use the cloze which is part of the mandatory national test of English (see section 6.2.6).

6.1.2 Outline of main study

The main study was carried out during one school year, 2006-2007. It was begun during the very first week of school (August, 2006) with the students filling in a questionnaire, and ended forty-three weeks later (June, 2007) with the collection of the students' leaving certificates. Thus, it is a longitudinal study stretching over approximately ten months. It should be mentioned that I had met the students already in eighth grade. I visited all four classes in February, 2006, soon after I had established contacts with the participating schools and teachers. The main purpose of those meetings was to introduce myself (in English), inform the participants orally about my study (in English and Swedish), and to explain why the consent of their parents was needed (in Swedish). By visiting the classes personally at an early stage, I also hoped to minimize external attrition (see section 6.1.3). My main research question was never disclosed; I referred to my project as "a study on oral proficiency in English". However, from the outline of the study and the form of consent, the students must have understood that I was also interested in what they did in

⁹ The items which were difficult to score correspond to #16 (*slight*) at the 2,000 level and #6 (*entry*) at the 5,000 level in Nation (2001b:425-426).

¹⁰ Item #1 (*career*) at the 3,000 level (Nation, 2001b:425).

¹¹ Large print may facilitate reading for people who suffer from dyslexia; one such copy was used in my study.

their spare time and what vocabulary they had acquired. The outline of the main study is provided in Table 6.1.

Table 6.1. Outline of the main study.

Week/Year	Week of study	Data collection/Activity	Comments
34/2006	1	Questionnaire	Start of study
35	2	Speaking test 1	“Life today”
36	3	Language diary (fall)	
37	4		
38	5	Productive Levels Test	
39	6		
40	7	Speaking test 2	“Three-part test”
41	8		
42	9		
43	10		
44	11	Fall break	
45	12		
46	13		
47	14	Speaking test 3	“TV life”
48	15		
49	16		
50	17		
51	18		
52	19	Christmas break	
1/2007	20		
2	21	Instructions speaking test 4	
3	22	Speaking test 4	“Expository speech”
4	23		
5	24		
6	25		“The world around us”, national test, part A, any time weeks 3-22. ¹²
7	26		
8	27		
9	28	Winter break	
10	29	Speaking test 5 (classes 1-2)	
11	30	Cloze test: “Working for Change”	National test, part B
12	31		
13	32	Written essay	National test, part C
		Language diary (spring)	
		Vocabulary Levels Test	The VLT,
14	33		any time weeks 14-19.
15	34	Easter break	
16	35		
17	36	Speaking test 5 (class 3)	
18	37	Speaking test 5 (class 4)	
19	38		
20	39		
21	40	Student interviews	
22	41	Results of the national test	
23	42	Last week of school	
24	43	Summer break	
		Leaving certificate	
		Teacher interviews	End of study

¹² The national test guidelines stated that the speaking part of the test, i.e. part A - “The world around us” (speaking test 5 of the present study), could be taken any time between weeks 3 and 22. The teachers in my study and I agreed that it would be taken *after* speaking test 4 in the study, which was in week 3. As it turned out, classes 1 and 2 took speaking test 5 around week 10, class 3 took it in week 17, and class 4 in weeks 18 and 19.

6.1.3 Criteria for participation and selection process

After discussing criteria for participation in the study at a seminar, it was agreed that a suitable and reasonably manageable sample to aim for should include a minimum of 60 students. Furthermore, the participants should be from varied backgrounds (different communities, schools, and classes) to increase the possibility of having results that could be generalized. Another criterion was that the participants should be in grade 9; i.e., they were 15 or 16 years old (cf. section 1.2). Since I was to do all data collection myself, for practical reasons my sample had to be found within a geographical region referred to as Western Svealand (see App. 3).

I selected four schools that matched the criteria described above and contacted them by sending emails to ten English teachers whose addresses I either already had or which I obtained from the schools' homepages. Four teachers at three schools responded that they were interested in participating. The first three who responded taught classes that had 89 students in total. With the aim of including a minimum of 60 students (and for practical reasons the number of students could not grow much beyond that) it was considered enough to involve these teachers' classes and exclude the fourth teacher and her class. The teachers were all females in their 50s with long teaching careers. When asked why they wanted to help out in the present study they all responded that they thought it would be interesting. No one had previously taken part in a project like this. I arranged a meeting at Karlstad University at the end of the spring semester 2006 for the teachers. The purpose of that meeting was to present the outline of the study and to discuss practical and administrative matters. Ethics was also discussed and the teachers were introduced to two of my supervisors.

When I had established a connection between each teacher and myself, I approached their principals (first via letter and then by phone) in order to introduce myself, present the aims of my study, and have their approval. All three responded in the affirmative, provided that the study was carried out in line with regulations stipulated by the Swedish Research Council, for example that the anonymity of the students was guaranteed. The principals also gave me

permission to copy materials related to the participating students’ national test in English.¹³

6.1.4 Three schools, four classes – and generalizability of results

First of all, of the three schools represented in the study, two schools were situated in small towns (Swe. *småstad*) and one was in a medium-sized town (Swe. *större stad*), see Table 6.2.¹⁴ Geographically, the schools were in Western Svealand. School 1 was the largest school in the study. It had almost 600 students in grades 7 through 9. About five per cent of the students had an immigrant background. In this study, “students with an immigrant background” is equivalent to students who are enrolled in the subject Mother tongue. At school 1, about one third of the students lived outside the town where they went to school. These students used school buses or public transport (train and/or bus) to go to and from school. The mean overall grade (Swe. *meritvärde*, see section 9.4) in grade 9 was on par with that of all ninth graders in Sweden.¹⁵ The two classes in my sample from school 1, i.e. classes 1 and 2 (both taught by the same teacher), were what in Swedish is generally referred to as *resandeklasser* (Eng. literally *traveling classes*). That is, almost all students in classes 1 and 2 went to school by bus.

Table 6.2. The schools in the study.

Descriptor	School 1	School 2	School 3
Class in study	1 & 2	3	4
Teacher	1	2	3
Type of location	Small town	Small town	Medium-sized town
Number of students	600	300	400
Grades	7-9	7-9	6-9
Immigrants (%)	5	0	40
“Traveling” students (%)	33	5	20
Overall grade compared with “Sweden”	On par	Above average	On par

School 2 was, as was previously mentioned, also in a small town. However, the socioeconomic status of the population in the town where school 2 was situated was higher than that of school/town 1. School 2 had around 300 students in grades 7 through 9. In this school, too, some students went to

¹³ The Swedish National Agency for Education is responsible for secrecy and publishing regulations with regard to the national test of English, but principals at local schools are responsible for actually permitting researchers to copy the material (Tommy Lagergren, Skolverket, personal communication).

¹⁴ That is, there is no school from a big city (Swe. *storstad*) in the study.

¹⁵ According to statistics for years 2003-2006 available from www.skolverket.se (retrieved March 28, 2007).

school by bus, but they were much fewer at this school, only about five per cent and none in class 3 (taught by teacher #2). There were very few students with an immigrant background at school 2 and no one in the whole school was enrolled in Mother tongue classes. The mean overall grade in grade 9 was above par when compared with the Swedish mean. Finally, school 3 had approximately 400 students in grades 6 through 9 and was, as I said, situated in a medium-sized town. A majority of the students lived close to the school, but about one fifth went to school by school bus and some of them (nine students, to be specific) were in class 4 (taught by teacher #3). Of the schools in the study, School 3 had the highest proportion of immigrant students, namely 40%. Similar to School 1, the mean overall grade in grade 9 was on par with that of all ninth graders in Sweden.

It should be pointed out that none of the students who traveled to school by bus in my study lived in extremely remote areas, which could have been the case in more sparsely populated parts of Sweden. I would like to consider my sample representative of ninth graders in *Västra Svealand*. Thus, the results may be generalized to that statistical population.

6.1.5 Participants and participant attrition

At the outset of the study, a total of 89 students were approached. Eighty students agreed to participate, i.e. 90% (see Table 6.3). Of the 80 students who started the study, 74 students completed it (92%). To count as a participant who completed the study, it was necessary that the following data had been obtained: the questionnaire, at least one language diary, five speaking tests, the Productive Levels Test, and the Vocabulary Levels Test. On sample level, both the external attrition rate (10%) and the internal attrition rate (8%) were low (see Table 6.3). This means that the rate of participation was high, particularly considering the fact that it was a longitudinal study (cf. Dörnyei, 2007:82-83). The reason why so many students decided to participate can only be speculated upon, but both students and parents expressed a certain amount of curiosity when the project was introduced. Some also said that the speaking tests would be good practice – something which the teachers in fact emphasized in their informal talks with the students. It is also possible that I managed to introduce the project in an appealing and trustworthy manner, so that students were not scared off from the very beginning. For instance, I think visiting the classes in the spring of grade 8 was indeed important. Also, the plan to have the lion's share of data collected by Easter 2007 was good because it is troublesome to get hold of students during the notoriously hectic latter part of the spring

semester.¹⁶ Another explanation might be that the participants generally seemed to like and stand by their teachers. Thus, if the teacher was positive towards my project, so were the students.

Table 6.3. External and internal attrition rates for the four classes and in total.

Class	Approached	Agreed	External attrition		Completed study	Internal attrition	
			Dropouts/ approached	%		Dropouts/ Agreed	%
1	21	20	1/21	5	19	1/20	5
2	19	18	1/19	5	18	0/18	0
3	22	22	0/22	0	21	1/22	5
4	27	20	7/20	26	16	4/20	20
Total	89	80	9/89	10	74	6/80	8

Some facts are known with regard to the external attrition rate. Two of the nine students in fact moved in the summer after eighth grade. Of the remaining seven, one wanted to participate but the form of consent was never filled in by the parents, something which automatically excluded the student from the study, and one disliked speaking aloud in English. Both boys and girls dropped out. With regard to internal attrition, one student moved and another received special education in the home. The remaining four dropped out for unknown reasons. Again, both boys and girls dropped out. Thus, both external and internal attrition were random and not systematic. As Table 6.3 shows, class 4 had higher attrition rates than the other classes. I have no reasonable explanation why. It might be a coincidence, or it might be because, relatively speaking, the students in class 4 were more independent of their English teacher than the other students in my sample were of theirs; at least that was my impression.

Most students in the sample had Swedish as a first language. Eight students had other first languages: Kurdish (3 students, class 4), Spanish (1, class 4), Norwegian (1, class 1), a Finnish-German bilingual (class 1), and an English-Swedish bilingual (class 4).¹⁷

¹⁶ Since it was the teachers' job to carry out the last speaking test, i.e. Test 5 (the mandatory national test), I knew that it would be carried out in due time before the end of the study.

¹⁷ The English-Swedish bilingual did not complete the study, but all other students with other first languages than Swedish did.

6.1.6 Parental contacts

I informed the parents regularly about the study via letters (email and regular mail). Four letters were sent out: the first together with the form of consent, the second at the end of grade 8, the third in the middle of the spring semester in grade 9, and the fourth at the end of the study. I sent separate reminders to parents whose children forgot to hand in the language diary. I also introduced myself and explained my study to all parents at parental meetings which took place at each school at the beginning of the school year. At those meetings, the parents could ask questions about the study, and so forth.

6.1.7 Ethical considerations

The informants of the present study were all minors, aged 15-16. Thus, it was necessary to collect written forms of consent from their guardians. After I had introduced myself to the students in the spring semester of eighth grade, a letter was posted to the parents (see App. 1) along with a form of consent (see App. 2) to be signed with a “yes” or a “no”. The fact that participation was voluntary was stressed both at my meetings with the students and in contacts with the parents. Thus, the principle of voluntary participation in research projects was emphasized both orally and in writing. I also pointed out that students might benefit in various ways from participating in my study, in the way that they would receive extra attention and be involved in a number of speaking tests, whose recordings their teachers would watch/listen to. In addition, the students would be allowed to watch/listen to themselves in English as well.¹⁸ However, I also pointed out that students might not benefit from participating in the study. They might become nervous due to the presence of a video camera, for example, and in fact perform worse than normally.

The forms of consent were collected by the students’ mentors or English teachers. Parents of students whose forms of consent were missing by Easter were contacted again by mail. At the end of the school year, forms of consent had been collected from all but one student in the participating classes. As was previously mentioned, the rate of external attrition was 10%, which was considered satisfactory (see Table 6.3).

The present research project was planned and carried out in accordance with the ethical regulations stipulated by the Swedish Research Council. The

¹⁸ The possibility of watching/listening to one’s own recordings depended on the willingness of the English teacher to arrange such sessions for the students.

study was approved by the Ethical Committee at Karlstad University before it started.¹⁹

6.1.8 Mixed methods research design

Dörnyei (2007:42) refers to research which combines qualitative and quantitative methods as *mixed methods research*, something which is a suitable description of the present study. The quantitative part of my study is larger and carries more weight than the qualitative part, but both parts are important nevertheless. One purpose of a mixed methods research design is to achieve a fuller understanding of a target phenomenon than if a mixed design had not been used (cf. Dörnyei, 2007:164). By adopting the current design in which EE is investigated both quantitatively (questionnaire, diaries) and qualitatively (interviews), the concept of EE is, hopefully, better understood. Likewise, oral proficiency is investigated using a mixed methods research design, and so is vocabulary. Mixed methods research is closely linked to the concept of *triangulation*, which refers to the generation of multiple perspectives on a single phenomenon by using a variety of data sources, theories, or research methods (Dörnyei, 2007:165). Examples of method triangulation in the present study are to have three raters (i.e. three perspectives) assessing the same speech samples and to measure vocabulary in different ways (two written tests as well as the use of peripheral vocabulary in speech, see section 5.3). Similarly but at another level, by combining and integrating both quantitative and qualitative components in my study, the overall design itself is triangulated and can therefore be claimed to have good design validity (cf. Dörnyei, 2007:63).

Before I describe how data were collected and analyzed, it is necessary to address quality criteria for research because these have a bearing on my methodological decisions and claim to have conducted a “disciplined inquiry”, which is the basic definition of scientific research (Dörnyei, 2007:48; Seliger & Shohamy, 1989:9-11). I rely heavily on Dörnyei (2007) in the present section and present quality criteria in the same order as he has, starting with a discussion on criteria in quantitative research, followed by criteria used in qualitative research.

Three terms that relate to quality criteria in quantitative research are reliability, measurement validity, and research validity. *Reliability* indicates “the extent to which our measurement instruments and procedures produce consistent results in a given population in different circumstances” (Dörnyei,

¹⁹ See § 8 in the minutes of the meeting with the Ethical Committee, Karlstad University, May 30, 2006.

2007:50). Dörnyei (2007:50) emphasizes that it is important to remember that it is not the test or the measuring instrument that is reliable or unreliable. Instead, reliability is a property of the scores on a test for a particular sample and it is, therefore, necessary to estimate and report on reliability in all research studies. *Measurement validity* is generally summarized in a single phrase: a test or a procedure is valid if it measures what it is supposed to measure (see e.g. Dörnyei, 2007:51; Seliger & Shohamy, 1989:188). For example, in order for a vocabulary test to be valid, it should measure vocabulary and nothing else, such as reading comprehension or spelling. However, this general definition is too simple. Dörnyei (2007:51-52) points out that the concept of validity has gone through major changes over the last few decades and that it is necessary to view validity not only as an attribute of a test. He suggests that measurement validity depends on the quality of the interpretation and not of the test or the test scores, something which I think is a sensible suggestion. Furthermore, the best a researcher can do is to provide evidence that his/her validity argument is more plausible than other potential competing interpretations. Validity is also specific to a particular situation, Dörnyei (2007:52) continues, and a unitary concept which can be supported with many different types of evidence. Finally, *research validity* is broader than measurement validity because it concerns the overall quality of the research project (Dörnyei, 2007:52). It has to do with the meaningfulness of the interpretations made in a study and the extent to which these interpretations generalize beyond the research study. Main threats to research validity include, for example, participant attrition, the Hawthorne effect (see chapter 2), and practice effect (participants' performance may improve simply because they gain experience in taking repeated tests, and thus experience of various tasks). In the present study, quality criteria – both with regard to the quantitative and the qualitative part of my study – are discussed in conjunction with relevant topics and not in separate sections.

Quality criteria are more problematic in qualitative research than those in quantitative research because qualitative studies are inherently subjective (Dörnyei, 2007:54). This means that in a qualitative inquiry, the results depend on individual perceptions. Needless to say, this situation gives rise to concerns as to the methodology because the researcher is in fact the instrument in a qualitative study. Even though the analytical procedure might be rigorous and systematic in a qualitative study, “at heart of any qualitative analysis is still the researcher’s subjective sensitivity, training, and experience” (Dörnyei, 2007:28). Another concern in qualitative research is *anecdotalism*, which refers to the problem of knowing whether the findings of qualitative research are genuinely

based on critical investigation of all data and not due to some well-chosen examples (Silverman, 2005:211). Despite problems such as these, qualitative research is necessary because subtle meanings associated with individual respondents are always lost when only quantitative methods are used (Dörnyei, 2007:28) and in language learning studies, it is often particularly interesting to include the perspectives of individuals.²⁰

With regard to *reliability*, in qualitative research this refers to the degree of consistency with which a researcher goes about categorizing the events or activities described (Silverman, 2005:224). The *validity* of qualitative research, on the other hand, is often in the hands of the researcher because at the end of the day, readers form their opinion not based on various validity arguments but on the investigator's overall research integrity (Dörnyei, 2007:59). To increase reliability and validity, Dörnyei (2007:60-61) suggests the use of validity/reliability checks during the course of a study, such as respondent feedback (e.g. letting respondents read drafts or express their views on tentative results) and peer checking (e.g. letting a colleague perform some aspect of the researcher's role).

6.2 Material

My study is based on data from eight different types of material which I account for in this section. The section begins with the questionnaire, which was the first type of data collected, and closes with a description of the interviews, which ended the study. I would like to point out that even though I account for assessment data in section 6.2.4, i.e. below the heading of the present section ("Material"), raters' assessments also constitute an analyzing tool in the present study.

6.2.1 Questionnaire

As was mentioned above, the questionnaire was designed to provide general background information about the informants, their motivation for learning English, and their views on English. The questionnaire also aimed to measure extramural English. Several questions were identical with or similar to items previously used by Sylvén (2004:280-288) and Oscarson and Apelgren (2005:110-118), where they had proved to yield valid information. The inclusion of questions from the latter source makes it possible to compare the responses

²⁰ For an overview of arguments for qualitative research methods in L2 acquisition studies, see Seliger and Shohamy (1989:118-120).

of my sample with those of a much larger sample, namely all Swedish ninth graders, which was the statistical population in Oscarson and Apelgren (2005). My questionnaire (see App. 4) is similar to the one used in the pilot study except for some minor adjustments and additions with regard to the following questions: 1:b, 5:c, 13, 24:e-j, 27:a-b, and 30. Question 30 (the number of books in the home), for example, aimed to capture at least one aspect of Bourdieu's cultural capital (cf. chapter 2).

I administered the distribution of the questionnaire myself during an English lesson (see Table 6.1). The participants were encouraged to take their time and respond honestly. It took about 20 minutes to fill it in. The English teachers administered questionnaires to absentees.

All responses were inputted into the statistical software SPSS by me.²¹ To increase coding-reliability, all values were systematically double-checked for every eighth student. Despite this precaution and other efforts to ensure coding-reliability, such as working meticulously and avoiding fatigue, it is still possible that some mistakes were made.

6.2.2 Language diary

EE data were collected with the help of two one-week language diaries per student (see App. 5). As mentioned above, the diary was developed in collaboration with Liss Kerstin Sylvén, who has also used it in research on EE (see e.g. Sylvén, 2006). The language diary was revised and improved in accordance with the suggestions articulated in the pilot study. The diary was filled in by the informants twice: the first time in order to register one week of EE during the fall semester and the second time to measure a week in spring. Both weeks were regular school weeks (see Table 6.1). The diary aims to measure students' extramural English but also to provide information about their extramural activities in Swedish. Moreover, informants also filled in information on any other language (or languages) that they might have come in contact with. The language diary was in Swedish and it contained 20 pages in total, including two explanatory pages about how to fill it in.

When the students received their first language diary, i.e. in September, I was present in the classroom to instruct them on how to fill it in. I gave both written and oral instructions. Students were encouraged to carefully fill in the time spent on seven extramural activities for each language

²¹ See <http://www.spss.com/> (accessed July 14, 2009). SPSS (Statistical Software for the Social Sciences) is the most commonly used software package in applied linguistics and educational research (Dörnyei, 2007:198).

(English/Swedish/Other): reading books, reading newspapers/magazines, watching TV, watching films, surfing the Internet, playing video games (Swe. *TV-/Dataspel*), and listening to music.²² There was also an eighth open category: “other” (Swe. *annat*) (cf. Forsman, 2004; Pearson, 2004). Students were encouraged to fill in the diary on a daily basis, preferably in the evening before going to bed. Should they forget to fill it in one day the students were instructed to fill it in as soon as possible after that. To support and motivate the students in this extracurricular task, their English teachers reminded them of the diary in class. There was also a memo written on the whiteboard in their homeroom. In addition, the parents were informed of the diary via letters or email, with the purpose of further motivating and supporting the students in their task and ultimately to limit internal attrition. Monday after the “diary week” was due day and the diaries were handed in to the teacher, who in turn was provided with a check list and a small gift for each student who completed the task. Not all students remembered to bring their language diaries right away. Reminders were therefore sent to the parents of the students who forgot to bring the diary to school, something which sped up the delivery. I should add that I took the opportunity to remind certain students who had forgotten to hand in their diaries when I visited the schools for the speaking tests. The procedure of collecting the language diaries was identical in the spring except for the fact that the teachers distributed the diaries this time.

Sixty-nine students (86%) handed in two diaries each and eleven students (14%) handed in one diary each (four in the fall; seven in the spring). For students who handed in two diaries, average figures are used in the analyses. For students who only handed in one diary, the figures reported for that week are used in the analyses. The most common reason for not handing in a diary was that the students forgot either to fill it in or to bring the diary to school. I should mention that for the spring language diary, I appealed to the seven students who had failed to hand one in the previous semester to actually do so this time. I mentioned the importance of having language diary data from *all* students. At this point in time, the students knew me and they also knew that they had promised to fill in language diaries when they agreed to participate in the study. My appeal seems to have had an effect since all of them finally handed in a diary and, thus, in the end I actually had EE diary data from all 80 students.

²² As suggested by Gee (2003:1-3), the term *video games* is adopted to cover both games played on video game consoles (e.g. PlayStation, Nintendo, and Xbox) and those played on computers.

Students' comments in the diaries and informal discussions with both teachers and students led me to the conclusion that the vast majority of the students took the assignment of filling in their language diaries seriously. They seemed to give information to the best of their ability. Nevertheless, using a language diary as a measuring tool undoubtedly involves a certain margin of error. In other words, memory is unreliable so it is difficult for students to remember exactly what they had done and for how long. When writing about their various extramural English activities, students, therefore, were bound to make estimates. Most of these estimates are probably very close to what indeed happened, but, needless to say, some might be less accurate. At times students filled in, for example, titles of films or TV-shows but forgot to fill in the time. In such cases, I used the following stipulated times per category: reading books 15 minutes, reading newspapers/magazines 15 minutes, watching TV 30 minutes, watching films 30 minutes, surfing the Internet 15 minutes, playing video games 30 minutes, listening to music 15 minutes, and "other activity" 15 minutes. For students who wrote "a few hours" (Swe. *några timmar*) I used a stipulated value of three hours. All stipulated times are based on what I considered reasonable after having studied the data. Finally, I should mention that it would have been possible to retrieve SPSS-calculated estimates for each missing value through regression analysis, but I decided that the number of missing values was too small to merit such a procedure.

A language diary template was created in Microsoft Excel into which EE data for each student were entered (in minutes/EE activity/day) by me. To increase coding-reliability, data for each student were double-checked. Excel calculated the total times (in minutes/week) for each EE activity and the total amount of EE for each student. An A3-sized paper copy of each student's language diary was then printed as reference. The total times for each student that had been obtained for each EE activity and overall with the help of Excel were entered into SPSS by me. Again, to increase coding-reliability, the totals for each student were systematically compared with the corresponding printed A3-diary. Thus, reliability checks were built into the whole coding process but, nevertheless, it is possible that some errors occurred.

6.2.3 Speaking tests

As was mentioned in section 6.1.1, speaking tests with a dyadic set-up and interactive tasks were considered suitable to measure oral proficiency in line with the national goals for speaking in grade 9. Five speaking tests were included in the study and all were proficiency tests, i.e. they aimed to test global

competence in a language (D. H. Brown, 2004:44).²³ The tests were spread out over the school year (see Table 6.1). Two of the piloted old versions of national tests were included, “Life today” and “TV life”, along with the speaking part of the national test in 2007, “The world around us”, which was the mandatory “real” test for the participants. These three tests are very similar and scores therefore comparable. The remaining two tests in my study were a Norwegian national test, which I refer to as the “Three-part test” (see Hasselgren, 1996b), and the one that I created myself, “Expository speech” (see App. 7).²⁴ Both tests were enthusiastically endorsed by the students in the pilot study. The former was included in the study mainly because it made all (including shy) students produce speech, something which perhaps was the result of the test design which explicitly aimed to minimize any possibility that a strong partner might dominate or a weak partner “spoil” the test (Hasselgren, 1996a:29). The latter test was included mainly because some of the piloted students adamantly stressed the importance of including at least one test that could be prepared in advance at home. Table 6.4 provides an overview of the five speaking tests, based on Brown’s (2004:141-142) taxonomy for oral production (cf. section 4.1.1).

Tests 1, 3, and 5 follow the same format. The first part of the test encourages the test takers to exchange information related to a given topic. In the second part of the test, the interlocutors are supposed to interact with each other by discussing various matters. This discussion is initiated by more or less provocative statements provided on small cards. The test takers take turns drawing cards, then they read the text on the card aloud and respond to statements such as “There is nothing wrong with junk food” (Test 1), “TV ruins family life – people have stopped talking to each other” (Test 3), and “Money makes people happy” (Test 5). These statements are subsequently supposed to be discussed with the interlocutor. In Tests 1 and 5 there is a third part in which cards with questions are used in order to elicit speech from both interlocutors (e.g. “What can you do to protect the environment?” from Test 5). Such questions may elicit what Davies (2003:119) refers to as “personal stories”, which, he argues, allow speakers to display linguistic expertise. The test takers are encouraged to speak English all the time, to interact with each other, to state their own opinion(s), and also to discuss what other opinions there might be. I used a manuscript for the test introductions and conclusions in

²³ It is possible to argue that Test 4, “Expository speech”, is more of an achievement test (cf. D. H. Brown, 2004:47-48) than a proficiency test because it was limited to one topic, i.e. the students’ own.

²⁴ I had previously used the “Expository speech” test successfully when teaching grades 9 and 10.

Tests 1 and 3, but otherwise the tests are self-explanatory. The English teachers served as the instructors in Test 5.²⁵

Table 6.4. An overview of the five speaking tests according to Brown's (2004) taxonomy.

Test	Name of test	Types of speaking tested				
		Imitative	Intensive	Responsive	Interactive	Extensive
1	Life today					
	Part 1				x	
	Part 2		x		x	
	Part 3		x		x	
2	Three-part test					
	Task 1		x	x	x	x
	Task 2				x	
	Task 3, part 1		x	x		
	Task 3, part 2		x	x		
3	TV life					
	Part 1				x	
	Part 2		x		x	
4	Expository speech					
	Part 1					x
	Part 2				x	
5	The world around us					
	Part 1				x	
	Part 2		x		x	
	Part 3		x		x	

Tests 2 and 4 were different from the other three tests, and also from each other. Test 2 was developed in Norway (Hasselgren, 1996a). It is about twice as long (average recording time: 22 min 25 s) as the Swedish tests and much more structured.²⁶ There is, for instance, a detailed manuscript which guides the test instructor, and consequently also the speaking dyad, through the test. After initial small talk, the test instructor reads all instructions aloud throughout the test. The test covers picture-cued story-telling, an information gap task, reading aloud, and role-play. Whereas Test 2 is thus more structured than the three national tests, Test 4 is the only test where the students were free to choose the topic themselves (see App. 7). They were to give a short speech after which a question-answer session followed (see App. 8). The students were also encouraged to bring props. The English teachers gave the instructions for

²⁵ I was present when class 3 took Test 5 because I assisted with the recordings; I was not present in any of the other classes for Test 5.

²⁶ Average recording times for the Swedish national tests were 10 min 20 s (Test 1), 10 min 33 s (Test 3), and 15 min 6 s (Test 5). For Test 4, the average recording time was 15 min 6 s.

Test 4 one week before the actual test date (see Table 6.1). This means that there was plenty of time for planning.

The random dyads to be used in all five tests were organized by me at the beginning of the study.²⁷ In Tests 1-4, Student A always received test materials on sheets of laminated blue paper, whereas Student B always received a yellow copy. The reasons for doing so were to improve face validity, secure consistency, and facilitate the administration while making consecutive recordings. Furthermore, the dyads were arranged so that each student was sometimes Student A and sometimes Student B, and I made sure that no student was paired up with the same person twice. I was strict about keeping the predetermined random dyads throughout the study, but on a few occasions they had to be changed due to someone's absence. At least one minute of planning time was allowed in Tests 1-3 and 5, something which has been shown to improve accuracy in speech (Mehnert, 1998). Time for planning has also been shown to improve oral fluency (Foster & Skehan, 1996). With regard to the collection of speech data, both video and (back-up) audio recordings were used in Tests 1-4, whereas only audio was used in Test 5, because that test had to be taken under similar conditions as other ninth graders in Sweden took the test (i.e., with audio recordings only). A total of 199 recordings of tests were collected, making up 45.6 hours of speech data. (See App. 6 for a technical description of the collection of speech data.)

A total of 19 tests were transcribed, of which 11 were of Test 5. The transcripts of Test 5 were used for analyzing oral fluency and use of polysyllabic words in speech for ten of the students in the sample. Transcriptions were made using the software Transana (Fassnacht & Woods, 2006). The transcription notations used in the present study, partly based on Sandlund (2004:107-108), are shown in Table 6.5. I will not include any direct samples of student speech in the results, but nevertheless find it necessary to show the level of transcription I used so that it is possible to replicate my study. The transcription notations are particularly relevant with regard to the analysis of oral fluency. A coding system of digits was used to secure the anonymity of the students, both in SPSS and in Transana. For instance, transcripts are identified with the help of nine digits, where the first digit identifies the test (1-5), the following four digits the code for Student A in the dyad, and the final four digits the code for Student B. Of the four digits in the student code, the first digit represents the class (1-4), the following two is the number of the student

²⁷ I had received information from the teachers beforehand with regard to dyads that had to be avoided due to problems beyond my control, such as previous instances of bullying between students.

(01-n), and the final digit indicates gender (1=boy, 2=girl). For example, transcript 510411072 was Test 5 taken by Student A, a boy (number 04) in class 1, and Student B a girl (number 07) in the same class.

Table 6.5. Transcription notations used in the present study.

Notation	Explanation
(.)	Pause < 1 second.
(1.4)	Length of pause in seconds.
#	Laughter.
.hh	Audible inhalation.
hhh	Audible exhalation.
°talk°	Whispered talk.
(talk)	Transcriptionist's doubt or inaudible parts of talk.
(())	Transcriptionist's comment, usually on body language.
<i>italics</i>	Other language than English, often speaker's L1.
/ /	Change of batteries or tapes.

6.2.4 Assessment data

The five speaking tests were assessed by four external raters, all experienced teachers of English.²⁸ Raters 1 and 2 were non-native speakers (NNS) of English and raters 3 and 4 were native speakers (NS) of English. Three were women (raters 1, 2, and 4) and one was a man (rater 3). Rater 1 assessed the students in all five tests, whereas the other raters assessed the students in either three or four tests, as shown in Table 6.6. The design of the assessment procedure means that each student was assessed by three different raters for each one of the five tests.

Table 6.6. Assessment totals for the four raters.

Test	Rater				Total
	1 (NNS)	2 (NNS)	3 (NS)	4 (NS)	
1	79	79	79	-	237
2	77	-	76	77	230
3	76	76	-	76	228
4	74	-	74	75	223
5	75	73	74	-	222
Total	381	228	303	228	1,140

In their assessment work, the raters were provided with a set of written instructions, almost identical to the ones described in Hasselgren (1996b; see

²⁸ While this section is placed under the heading "Material", it should be stressed that the assessment data also constitute an analyzing tool in the present study.

App. 9). They were also equipped with one so-called performance profile scheme per student and test (see App. 10). The performance profile scheme was similar in content to the one used in Hasselgren (1996b:35-37) but slightly altered in its layout.²⁹ For instance, for each of the ten criteria (a-j) I added two boxes (level “2” and “4”) between the three boxes that matched the criterion descriptors (levels “1”, “3”, and “5”). I did so instead of letting the raters check the space in between boxes, which was suggested in Hasselgren (1996b:7). The raters worked independently, watching and listening to the DVD recordings, assessing one dyad at a time.

The raters were instructed to use the performance profile scheme as a guide in order to match each spoken performance with a grade on each of the two scales “message and fluency” and “structures and vocabulary”, where 1 was the lowest grade and 6 the highest. The scales were described in detail in the instructions (App. 9). This means that each student was awarded what I call a *factorial grade* for message and fluency (henceforth *fluency*) and another one for structures and vocabulary (henceforth *vocabulary*). A final *overall grade* (1-6) was awarded for the whole performance. The general principle was that the overall grade should be the grade which best reflected the student’s performance on the two criteria scales. Should the decision be difficult to make, raters were told that fluency should be weighted more heavily than vocabulary, since fluency is considered difficult due to the online processing in speech, with little (or no) time to monitor one’s production (cf. Hasselgren, 1996b:8).

To measure interrater reliability, i.e., the degree of consistency between different raters’ assessments of students’ performances, Hasselgren (1997:243-244) used the Pearson correlation coefficient (r) (see also section 6.3.6.5). She considered a minimum value of r at .4 as “reasonable” for overall grades.³⁰ In my study, there were normal distributions for factorial and overall grades for each rater. I investigated interrater reliability for both factorial and overall grades using the Pearson correlation coefficient (r) and reliability was satisfactory or high, ranging from .451** to .703** for the overall grade (see App. 11).³¹ Analyses using Cronbach’s alpha confirmed that interrater reliability was high, for example .906 for the overall grade.

²⁹ Angela Hasselgren gave me permission to use her material for my dissertation study, for which I am very grateful.

³⁰ In behavioral and social sciences, r -values of .4 and above are indeed high (Hugo Wikström, personal communication).

³¹ Factorial grades were missing for two students on some tests for one rater. I keyed in estimated values for them, based on what the other two raters had awarded these students on the test. The same grade was inputted when the other two raters agreed on the factorial grade. When the raters disagreed (by one point, no more), I keyed in the higher of the two, giving the student the benefit of the doubt, so to speak.

The raters' qualitative assessments, quantified into factorial grades (1-6) and an overall grade (1-6), were inputted into SPSS.³² For coding-reliability, keyed in data for approximately every eighth student were double-checked.

6.2.5 Vocabulary tests

Both the Productive Levels Test (PLT) and the Vocabulary Levels Test (VLT) are considered valid and reliable measures of learners' vocabulary, where the former measures learners' productive vocabulary ability and the latter learners' receptive vocabulary (see section 5.3). A shortened version of the original PLT (Nation, 2001b:425-428) was used in the present study (see App. 12). There was no need to test Swedish ninth graders extensively on vocabulary from the 5,000 level and beyond (cf. section 5.2), so only a few items were included from the 5,000 level and the University Word List. In the pilot study, students completed the PLT in less time than the VLT. Therefore, because it was the shorter test of the two and perhaps perceived as less "terrifying", the PLT was scheduled before the VLT (see App. 3).

Due to difficulties in the scoring process in the pilot study (see section 6.1.1), two of the original PLT items were replaced with new target words from the same word frequency level. The new target words (corresponding to #15, *abroad*, and #18, *disturbing*) were picked from the English word frequency list available at the Compleat Lexical Tutor website and the sample sentence for each word from the *Macmillan English Dictionary for Advanced Learners of American English* (2002).³³

The VLT used in the present study is also based on Nation (2001b:416-421), but I saw no reason to include infrequent vocabulary beyond the 5,000 level. Thus, I used a shortened version (see App. 14).

6.2.6 The Swedish national test of English 2007

The Swedish national test of English for grade 9 in spring 2007 consisted of three parts: A (oral interaction and production), B (receptive skills: reading and listening), and C (written production).³⁴ Part A was described above (section 6.2.3). One section in part B included a rational cloze test called "Working for

³² Data for each criterion (a-j) were also fed into SPSS, where 1 was the lowest and 5 the highest.

³³ The Compleat Lexical Tutor, <http://www.lex tutor.ca/> (accessed July 7, 2009). Note that the spelling, *compleat*, is archaic. This might allude to the best-selling 18th century English textbook by John King (alias Hans König), *The Compleat Grammar, English and High-German*. For more information on that book, see Werner Hüllen's text: <http://faculty.ed.uiuc.edu/westbury/Paradigm/hullen.html> (accessed Aug. 19, 2009).

³⁴ Written permission was obtained from the National Agency of Education to include Test 5 in the study and to describe it in my publications.

Change”. The passage had approximately 200 words and there were a total of twelve gaps (equivalent to 12 points). I decided to use it as an additional measure of vocabulary in the present study, even though cloze tests are supposed to be measures of overall proficiency rather than vocabulary (D. H. Brown, 2004:9). Guidelines for correcting and assessing the test were provided by the test constructors. A student’s grade on each part of the test was indicated by a tick on a scale, which was in the form of a continuum consisting of ten “boxes” (i.e. levels):



This system allows teachers to discriminate between various learner performance within the span of one grade; for example, there are three levels for “pass” (i.e. “G”; Swedish grades are explained in the following section, 6.2.7). The scores on all parts were combined into a scoring profile and an overall grade for the whole test by the teachers. The scoring profile and overall grade for each participant was collected along with copies of “Working for Change” and the essay (i.e. part C).³⁵

Students’ grades on part A, i.e. oral proficiency and production, were quantified 1-10 (i.e., matching the ten boxes/levels). In chapter 7, these grades, which thus represent the teachers’ assessment of the students’ OP, will be compared with the grade for oral proficiency (OP grade) described in section 6.2.4, which are my raters’ assessment of the students’ OP.

6.2.7 *Leaving certificate*

In Sweden, students’ leaving certificate from ninth grade is based on the sum of grade points from sixteen subjects. This sum is called the *overall grade* (Swe. *meritvärde*).³⁶ A *pass* (Swe. *godkänt/G*) in a subject is worth 10 grade points, a *pass with distinction* (Swe. *väl godkänt/VG*) 15 points, and a *pass with special distinction* (Swe. *mycket väl godkänt/MVG*) 20 points. The maximum overall grade one can get is 320. Copies of participants’ leaving certificates were collected for use in correlation analyses and for class comparisons.

³⁵ It was not within the scope of the study to analyze the essays.

³⁶ See the Swedish National Agency for Education for terminology in Swedish and English with regard to the grading system, <http://www.skolverket.se/sb/d/357/a/1271> (retrieved May 3, 2009).

6.2.8 Interviews

Eight students (one boy and one girl from each class) were interviewed in the present study.³⁷ The interviews took place at the end of the school year. By then, I knew the participants and wanted to conduct interviews with a number of students who together covered the whole range of EE (very low/medium/very high amount of EE) as well as the whole range of OP (very low/intermediate/very high level of OP). Thus, the interviewees were not chosen randomly. In fact, I approached the students according to a predetermined list of names. All students who were approached agreed to be interviewed.

The questions asked in the interview were the same open-ended questions as those used in the pilot study (section 6.1.1) except for the BALLI questionnaire, which was excluded. To the best of my ability I followed the guidelines for conducting interviews suggested in Ejvegård (2003). The interviews were conducted in Swedish, and each interview took 20-30 minutes. Audio recordings were used and subsequently transcribed. To increase validity/reliability, transcripts were sent to the interviewees for feedback and approval. The main purpose of the interviews was to collect individual students' comments on EE and views on English in general, as well as on their strategies for learning words. I considered such comments important because, as was previously mentioned (see section 6.1.8), subtle meanings associated with individuals are lost when only quantitative methods are used. I will use some student comments to illustrate the results obtained from the statistical analyses when the main findings are discussed in chapter 10. Interview data are too limited, however, to merit a separate section in chapters 7-9.

6.3 Methods of analysis

In this section, the methods of analysis used in the study are described. First, there is a discussion of various ways of clustering the data (section 6.3.1). This is followed by four sections that refer to my research topics: EE (6.3.2), OP (6.3.3), vocabulary (6.3.4), and background variables, motivation, and students' views on English (6.3.5). The final section (6.3.6) presents the statistical methods that were used. An overview of the mixed methods research design is provided in Table 6.7, with references to relevant chapters/sections.

³⁷ The three teachers were also interviewed in order to retrieve background information, e.g. about their educational background and teaching career.

Table 6.7. Overview of the mixed methods research design.

See chapter/section	Data/Tool	Aims to measure	Assessor/Observer	Method of analysis
6.3.2	Language diary	• EE	Student	Quantitative
6.3.2	Questionnaire	• EE • Background variables • Motivation • Views on English	Student	Quantitative
6.3.3	Speaking tests/ Assessment data	• Oral proficiency	Rater	Qualitative ↓ Quantitative
6.3.3	Speaking data/ Intra-utterance pauses > 1 s	• Oral fluency	Doctoral candidate	Quantitative
6.3.4	Written vocabulary tests	• Vocabulary size	Doctoral candidate	Quantitative
6.3.4	Cloze test	• Vocabulary	Teacher	Quantitative
6.3.3	Speaking data/ Use of polysyllabic words in speech	• Advanced vocabulary	Doctoral candidate	Quantitative + Qualitative
10	Interviews	• EE • Strategies for vocabulary acquisition • Views on English	Doctoral candidate	Qualitative

6.3.1 Group design

There are several ways of clustering data depending on what one is interested in. For instance, school class and gender are examples of naturally given variables on which group design may be based. Finding the “right” way of clustering data may constitute a problem since even for the same data set, there can be several ways of clustering data depending on one’s focus and how fine-grained an analysis one wishes to make. However, the main aim of the clustering must be to find a group design that optimally answers to the research question(s) and, at the same time, withstands scrutiny. In addition, the group design should facilitate the communication of results and preferably also maintain a certain amount of granularity, i.e. level of detail in a set of data. In other words, the group design per se should not lose information that might be of importance.

In the present study, the group design is a matter of clustering eighty students in grade 9. Generally, previous empirical studies within a certain field of research would provide support in the choice of method of group design, but that is not the case here. Empirical studies on extramural English and its potential effects on learning outcomes are too few and/or not informative enough regarding group design (cf. chapter 3). However, statistics books are helpful regarding what would constitute a suitable number of subsets depending on the size of the sample. One rule of thumb sets the number of subsets to $g \sim (n/2)^{1/2}$, where g is the number of subsets and n is the number of objects (in this case students) (Mardia, Kent, & Bibby, 1979:365). According to this formula, the number of subsets in the present study should be around six. Even though this piece of information is helpful regarding the number of subsets, the problem still remains of how to cluster the students.

Logically there are two alternative group designs here, both of which are based on extramural English, i.e. the main focus of the thesis. In the first alternative, the number of subsets *and* the number of students within each subset are first determined and then the distribution of the sample based on extramural English data is taken into consideration to set the time intervals for each subset. For example, five subsets of equal size (i.e. approximately 15 students per subset) would be a reasonable group design, bearing in mind the rule of thumb from statistics (around six subsets) and considering the number of students in the sample (80). In such a group design, subset 1 would include the 15 students with the lowest amounts of EE and the time interval for subset 1 would start at 0 hours/week and end with the 15th student's reported hours/week. Subset 2 would begin with the 16th student and end with the 30th student, and the time interval for subset 2 would correspond to the reported EE times for these students, etc. In contrast, in the second alternative, the group design is directly connected to the actual distribution of the students' values of EE. That is, the clustering of the students into various subsets would aim at mirroring the results obtained for EE. Thus, the purpose of the second alternative group design would be to catch students where they are in terms of their total amount of EE. The EE variable is expressed in hours per week so each subset would, then, span over a specific time interval that would be "given" by the data. The number of subsets would also be more or less given in view of the rule of thumb and the actual distribution of the sample. The basic difference between the two alternative group designs lies in the clustering process. In the first alternative, the number of subsets and the number of students within each subset is first decided upon, and then the EE variable is

used to set time intervals for each subset. In the second alternative, it works the other way around. The EE variable is first studied and then the number of subsets is decided upon so that the subsets mirror the distribution of the students' amounts of EE. The number of students within each subset and the time intervals for each subset will be derived automatically from the subsets, so to speak.

I arrived at the two alternative group designs described above after having considered other options, which I will briefly relate in the following. For instance, it would be possible to cluster students into two subsets of equal size, i.e. 50% of the students in one subset and 50% in another, but in such a design each subset would have very large time intervals in terms of EE and two subsets must be considered too few in the present study. It would also be possible to cluster students into three subsets, where the subsets could be of either equal or unequal size, e.g. 25% of the students in subset 1, 50% in subset 2, and 25% in subset 3. However, three subsets is probably also too rough a clustering of the current sample because only three subsets would be detrimental to the granularity of the data. Therefore, four, five, or perhaps six subsets (of equal or unequal size) would be the preferred number, but not more. It would be difficult to digest the results from more than six subsets.

Having considered the two alternatives at hand and with a need to choose between them, because there is not room to present parallel results for both group designs in this thesis, my first-hand choice was to adopt the second alternative of group design in the statistical analyses. That is, I chose to study the EE variable first and then to decide upon the number of subsets. This means that the subsets should mirror the distribution of the students' amounts of EE and that the number of students within each subset and the time intervals for each subset are derived automatically from the subsets. The most important argument for this choice is that the subsets that will emerge from this specific group design will be the closest representations of what is indeed found on EE in this study, based on available language diary data. Thus, in the present project, data are analyzed for so-called EE subsets, in addition to the naturally given group designs of school classes and gender.

6.3.2 Extramural English

Extramural English was measured with the language diary and certain questions in the questionnaire. Data from the language diaries are viewed as primary EE data and form the basis of my whole investigation, which is mainly statistical. The statistical software SPSS was used in the analyses. The virgin SPSS file

(which included all data that was quantified in the present study) comprised a total of 521 variables for 80 students, i.e. approximately 41,500 reference points.³⁸

The results on EE, both from the questionnaire and the language diaries, are presented in chapter 7. EE results were analyzed for the whole sample and the participating classes, as well as for boys and girls. Furthermore, subsets based on the EE variable (i.e. the total amount of EE) emerged (see section 7.1.2). I will refer to these subsets as “EE subsets”. The EE subsets are in focus in all subsequent analyses together with analyses on sample level, even though other group designs (such as school classes and gender) will also be used, but not as frequently. In addition to investigating the total amount of EE, the seven EE activities and the eighth open category listed in the diary were also examined.

EE was correlated with results on oral proficiency (chapter 7) and vocabulary (chapter 8) to see whether there were any significant correlations and, if so, how they could be described and perhaps explained. EE was also investigated with regard to certain background variables and motivation (chapter 9). In chapter 10, all results are summarized and there I will also use EE data from the interviews in my interpretation of the results.

6.3.3 Oral proficiency

As can be seen in Table 6.6, each student was assessed by three raters in each of the five speaking tests. The raters awarded the students factorial grades (one for fluency, one for vocabulary) and an overall grade. All grades were normally distributed and interrater reliability high. Based on the collected assessment data, SPSS calculated a mean oral proficiency grade for each student which was based on their overall grades from three raters (3×5 tests = 15 in total). The mean oral proficiency grade is used as a measure of students’ oral proficiency in the study and will henceforth be referred to as *the OP grade*, ranging from 1 (the lowest grade) to 6 (the highest). Its reliability is satisfactory or high (cf. section 6.2.4). I also consider its validity high. However, it should be noted that the OP grade was affected by a number of factors beyond the control of the students, such as the interlocutors they were paired up with, the tasks at hand (excluding the topic of Test 4), and the fact that they were recorded. Such factors might have been detrimental to the performance of some students. However, the

³⁸ A separate SPSS file was created for the investigations of oral fluency and use of polysyllabic words in speech. Likewise, a separate SPSS file was created with the questionnaire items used in the comparisons of my sample with that of NU-03.

inclusion of as many as five speaking tests should level out at least some possible negative effects of the design.

Oral fluency, which is one aspect of oral proficiency, is analyzed for ten of the students in the study, namely those with the five highest and five lowest OP grades. Data from Test 5 were used. The measure used for oral fluency was mean intra-utterance pause length beyond one second. An *intra-utterance pause beyond one second* is defined as a pause, filled or unfilled, of more than one second in an internal position in an utterance. A *filled pause* includes sounds such as “uh” and “uhm”, whereas an *unfilled pause* is equivalent to silence. An *utterance* is made up of one or several words that provide referential or pragmatic meaning (see Brock, 1986:52-53). For example, the single word “yes” counts as an utterance in answer to the question “Do you like your neighbors?” since it provides referential meaning.³⁹ An utterance may be equivalent to a speaker’s turn, but a speaker’s turn may just as well consist of several utterances. Example (1) below illustrates the latter; it is a turn that is made up of four utterances (indicated by raised figures 1-4).

- (1) 1162 ¹ok good # uhm ²well i think that's pretty all (.) and
 oh ³i have a neighbor that is a masseus ma (.) *massör*
 (.) ⁴what is it called [512021162, line 13]

In its context (not provided here but available upon request) the first utterance provides pragmatic meaning (confirms previous speaker’s answer). The second utterance sums up the current topic of conversation (referential meaning). However, the speaker suddenly recalls that she wants to add information to what she has talked about earlier, hence the utterance about the neighbor (referential meaning). Finally, there is a fourth utterance in the form of a direct question (pragmatic meaning).

Sometimes it is difficult to decide on the actual boundaries of an utterance, since spoken utterances are often incomplete or elliptic and include pauses, both short and long ones. In the analysis, it was therefore necessary to decide whether the speech after a pause belonged to the on-going utterance or if it should rather be interpreted as a new utterance. The question whether the pause is intra-utterance is due to this decision/interpretation. If the speech that follows is interpreted as a new utterance, the pause is an utterance initial pause rather than an intra-utterance one. In the present study, only intra-utterance pauses were counted. Examples (2) and (3) illustrate the problem. In example

³⁹ Example from transcript 510411072, student 1041, line 27.

(2), the 1.8 second long pause was interpreted as an intra-utterance pause (bold face). In example (3), the 2.0 second pause was interpreted as an utterance initial pause (underlined). Note that the pause of 3.1 seconds at the end of example (2) is utterance final and thus excluded from the current analysis.

(2) 1041 yes (.) and at home (**1.8**) it is it can be very noisy
 (3.1) [510411072, line 43]

(3) 1162 hello # my name is [name] and uh (.) my house is red
 and uh (.) we live on the country so it's (.) we have
 a lot of (.) uh big area and uh (2.0) uhm we also
 have we have to have a big area because we have
 sheeps (.) so they have to be somewhere # [512021162,
 line 1]

The one second threshold of the measure – mean intra-utterance pause length *beyond one second* – is interesting in the light of research on native speaker data. In her huge corpus, Jefferson (1989:170) found that there was what she calls a *standard maximum tolerance* for silence at approximately one second. Thus, a threshold of one second is not necessarily an arbitrary choice, even though Jefferson's findings, based on native speaker data, cannot automatically be presumed to be applicable also in an analysis of non-native speaker data. Also, the threshold is culture specific. Nevertheless, the one second threshold has been applied also in research on L2 speech (see e.g. Iwashita et al., 2008:34; Mehnert, 1998:90).

It is often essential in measures of oral fluency (and also of oral vocabulary, see section 6.3.4) to have an idea of individual speakers' speech length. The simplest way is to use time, for example the total number of seconds of speech per speaker. However, in my dyadic set-up, using time as a measure of individual students' speech length was impractical due to, for instance, the turn-taking processes that occurred. Instead of measuring speech length in time, I counted the total number of words uttered by each student. All English words were included in the word count; any instances of first language vocabulary were excluded, as were pause fillers (e.g. *uh, ah*) and false starts that did not make up any existing words (e.g. *ar always > always*). In this way, pruned transcripts were created for each student in which what counted as a word equaled an English token. Word counts were made using the tool for counting

words offered by Microsoft Word.⁴⁰ This means that I could use the same word count as a measure of speech length both in my investigation of pauses and fluency and in my investigation of polysyllabic vocabulary (see section 6.3.4).

6.3.4 Vocabulary

Students' vocabulary was measured with my shortened versions of the PLT (see App. 12) and the VLT (see App. 14). The rational cloze test "Working for Change", which was part of the Swedish national test of English in 2007, was also used for this purpose. The PLT scores were calculated by counting the number of correct answers (1 point/correct answer), ignoring minor spelling and grammatical mistakes (see section 5.2). The learners' total scores were keyed into SPSS and double-checked. The frequency levels of the target words are shown in Table 6.8.

Table 6.8. Frequency levels of the test items in the Productive Levels Test.

Test items	Word family frequency level	Points
1-17	2,000	17
18-32	3,000	15
33-40	5,000	8
41-45	University Word List	5
Total		45

In scoring the VLT, each correct combination of target word and definition was awarded 1 point. The learners' scores on each level and their totals (see Table 6.9) were inputted into SPSS. To ensure the reliability of the process of keying in the values, I double-checked the scores for each student. With regard to "Working for Change", each student's score on the test was obtained from the copy of their national test scoring profile, entered into SPSS, and finally double-checked.

Table 6.9. Frequency levels of the test items in the Vocabulary Levels Test.

Test items	Word family frequency level	Points
1-10	2,000	30
11-20	3,000	30
21-30	5,000	30
Total		90

⁴⁰ However, since that tool counts contractions as one word, I had to make the final calculations manually in order to arrive at the final number of total words per student.

I served as the test administrator for the PLT in all four classes. The English teachers administered tests to those who happened to be absent. The PLT was taken in the fall semester. In the spring, the teachers administered the VLT with the help of written instructions (see App. 13). For obvious reasons (it was part of the national test), “Working for Change” was also taken in the spring (see Table 6.1).

Students’ use of peripheral vocabulary in speech was also analyzed for a subset of students in the present study. As was mentioned in chapter 5, core and peripheral vocabulary are related to the syllable structure of words and I examined words with *three syllables or more*. I labeled such words *polysyllabic words*, even though, strictly speaking, words with two syllables are also polysyllabic. By drawing the line at three syllables, the probability of investigating clearly peripheral vocabulary increased due to the fact that “the notions of core and periphery are correlated not only with frequency of usage, but also with the parameters of grammatical type, meaning, etymology, and *syllable structure*” (Minkova & Stockwell, 2006:465; my italics). I should add that although polysyllabic words tend to be peripheral, not all of them are. For instance, a word such as *family* is polysyllabic but not peripheral because, according to frequency rankings, *family* belongs to core vocabulary.⁴¹ This means that among the polysyllabic words identified in my study, there might be items which are core vocabulary. However, the great majority of the polysyllabic words should be peripheral.

What counts as a word in the investigation of polysyllabicity of words? First of all, as I said, a polysyllabic word is defined as a word that consists of three or more syllables. I counted the number of syllables according to CV-tier theory presented in chapter 5.⁴² Such words were suitable for analysis because the use of polysyllabic words in one’s speech (or writing) is an indication of the size of a speaker’s (or writer’s) vocabulary. Basically, the more polysyllabic tokens and types a speaker (or writer) produces, the higher the probability that the same speaker (or writer) has also acquired advanced vocabulary.

My study on polysyllabic words in learner speech is based on speech data from Test 5, which was the most important test in the study from the students’

⁴¹ The word *family* is included in the first 1,000 most frequent English word families as well as the first 1,000 most frequent English head words, according to Paul Nation’s frequency lists available at Tom Cobb’s website “Compleat Lexical Tutor”, <http://www.lextutor.ca/> (accessed May 12, 2009).

⁴² Should there have been any doubtful cases of whether a word should be counted as having two syllables or three, I had decided beforehand to count such words as having three syllables (i.e. as a polysyllabic word). Compare, for example, the two ways of pronouncing the word *mackerel* given in the *Macmillan English Dictionary for Advanced Learners of American English* (2002): *macke-rel* or *mack-er-el*. However, no such problematic words turned out to be part of the relevant data.

point of view. It was also the last speaking test in grade 9 which, I think, made it particularly interesting to study. The instructions/tasks in Test 5 contained a number of polysyllabic words, which were subsequently used by the students in the dialogues. There was a need to distinguish between such polysyllabic words in the speech data and other polysyllabic words which were the students' "own", i.e., polysyllabic words that were not included in and thus elicited by the test material. It was not within the scope of the present study to investigate polysyllabic words for all eighty students in my sample. Instead, ten students were selected for the investigation, namely the students with the five highest and five lowest OP grades (i.e. the same students whose oral fluency was also examined, see section 6.3.3). Transcripts were analyzed and the ten students' own polysyllabic words identified.⁴³

Based on speech data from these students, I calculated the total number of *tokens* (i.e., all English words), *core tokens* (i.e., all tokens minus polysyllabic tokens), *polysyllabic tokens*, *polysyllabic types*, and *the students' "own" polysyllabic types* (i.e., the polysyllabic types that were not included in and elicited by the test material). Four types of word ratios were also calculated: (1) *polysyllabic tokens per all tokens*; (2) *polysyllabic types per polysyllabic tokens*; (3) *the students' own polysyllabic types per polysyllabic types*, and (4) *polysyllabic tokens per core tokens*. The measure of the students' own polysyllabic types per polysyllabic types is an example of an intrinsic rarity measure (see section 5.3). By examining students' vocabulary both in absolute numbers of words and in word ratios, I was hoping to discriminate between the students in terms of how advanced the vocabulary they had used was. The method is similar to the two measurements of lexical sophistication used by Afitskaya (2002): (a) advanced tokens per all tokens and (b) advanced tokens per basic tokens.

I used the same pruned transcripts and word counts per student in the present analysis of use of polysyllabic words in speech as I did in my analysis of oral fluency. That is, what counted as a word equaled an English token. As a consequence, the use of for example the singular and plural forms of a word for a speaker was recorded as two words/tokens for that speaker in the total word count (i.e. speech length). As was mentioned in chapter 5, tokens are all the words in a language sample. A type, on the other hand, is each individual word. This means that when a word is repeated, there will be two (or more) tokens of one type (Malvern et al., 2004:19). Furthermore, compounds were counted as one word. Contractions were counted as separate words: *there's* = *there is* (i.e.

⁴³ The method of excluding lexical items that were part of task instructions has been used by others in measures of lexical richness and lexical sophistication (see e.g. Wesche & Paribakht, 1996:25).

two words). The *Macmillan English Dictionary for Advanced Learners of American English* (2002) was used as a reference book for determining syllable structure and compounds.

After all polysyllabic words had been extracted from the transcripts and the students' own polysyllabic tokens and types had been identified by a manual count, I ran the words through the Lexical Frequency Profile to confirm the total numbers of tokens and types.⁴⁴ Then, I examined the etymology of the polysyllabic words with the help of the *Oxford English Dictionary Online*. Using a method of analysis in which both the words' polysyllabicity and etymology are considered is here hypothesized to yield reliable results on how advanced a student's vocabulary is. As has been mentioned previously, the notion of core and peripheral vocabulary correlates with etymology so that core vocabulary is mainly Germanic. In contrast, peripheral vocabulary has a large proportion of non-Germanic words while the proportion of Germanic words drops in the peripheral layers of the lexicon (see e.g. Minkova & Stockwell, 2006).

The *Oxford English Dictionary Online* was used to classify each polysyllabic word etymologically. The words were classified as "Germanic", "Non-Germanic", or "Hybrid". An example of a polysyllabic Germanic word is *anything*, where both *any-* and *-thing* are of Germanic origin. A polysyllabic word of non-Germanic origin is *environment*, which is from Old French (*environnement*). Hybrids are words which combine parts derived from a Germanic and a non-Germanic language. One example of a hybrid word is *individually*, where the stem, *individual*, is of Latin origin (i.e. a non-Germanic language) and the suffix *-ly* of Germanic origin (Old English *-lice*, probably influenced by Old Norse *-liga*). Hybrids may also be typical noun plus noun compound sequences, such as in present-day English *paper bag*: *paper* (non-Germanic; Anglo-Norman *papir/paper/papere/papire* and Middle French *papierre*) and *bag* (Germanic; Old Norse *baggi*). As was discussed in chapter 5 and above, the proportion of Germanic words decreases and the proportion of non-Germanic ones increases in peripheral vocabulary, i.e. beyond the 1,000 word level. Thus, studying the etymology of students' polysyllabic words adds a qualitative dimension to the investigation of students' word stock.

6.3.5 Background variables, motivation, and views on English

Background variables, motivational factors, and students' views on English were analyzed with the help of a number of questions in the questionnaire (see

⁴⁴ The Lexical Frequency Profile is available online at <http://www.lextutor.ca/vp/eng/> (accessed May 15, 2009).

Table 6.10). The results are presented in chapter 9. Where applicable, comparisons were made with the results of the *National evaluation of 2003 (NU-03)* (Oscarson & Apelgren, 2005). The SPSS file with data from the *NU-03* is public and I could thus use it to compare my sample with the statistical population of the *NU-03*, i.e. all ninth graders in Sweden.⁴⁵ With regard to the background variables and motivation, these were examined in relation to the EE variable and the OP grade. Several of the variables were also discussed in the interviews. Relevant results from those are integrated in the discussion of the role of EE in language learning in chapter 10.

Table 6.10. Schematic description of the analysis of students' background, motivation, and views on English.

Target of investigation	Name of variable	Analyzed with question(s)
Background	Travels abroad	2:a-e
	Parents' educational background	13
	Number of books in the home	30
	Rural versus urban residency	4
Motivation	Self-efficacy	12; 21:a-b; 26:a-b, e-f
	Anxiety related to speaking	5:a-c; 22; 28
Views on English	Beliefs about where English is learned	17
	Self-assessment of work in the school subject English	12; 21:a-c
	Opinions about the school subject English	26:a-j

6.3.6 Statistical methods

What statistical analyses are carried out depends on the research question and the type of data collected. The main part of my study consists of a quantitative analysis of my data. In quantitative research, the numerically coded data are always of three types: *nominal* (or *categorical*) *data* (variables that have no numerical values, e.g. gender; the assigned values are arbitrary), *ordinal data* (variables that involve ranked numbers, e.g. a multiple-choice item whose responses can be placed on a “frequency” continuum), and *interval data* (variables that can be seen as ordinal data but in which the various values are at an equal distance or intervals from each other on a continuum, e.g. test scores) (Dörnyei, 2007:207-208). My study includes data of all three types. For instance, certain questions in the questionnaire yield nominal data, e.g. items 6

⁴⁵ My assistant supervisor Hugo Wikström was responsible for preparing the *NU-03* data for analysis and made the SPSS file available for the present study.

(informants' type of housing) and 10:b (with whom the informant lives, when his or her parents do not live together), whereas others provide ordinal data, e.g. items 3 (how frequently informants receive help from family members with homework) and 5:c (how frequently informants speak English). Some questions can be claimed to yield interval data, e.g. item 12 (where each informant's belief about his or her final grade in English is the topic). Interval data provide the most precise information, provided that there is a reasonable normal distribution of the data, whereas nominal and ordinal data are less precise (Dörnyei, 2007:227). It is common in the Humanities to treat ordinal data as interval data which is a standard that I will adopt in the present analyses when standard deviations do not differ considerably or the distribution is normal or close to normal (cf. Dörnyei, 2007:227-228).

Statistics can be divided into two principal areas: descriptive and inferential. *Descriptive statistics* are basically used "to summarize sets of numerical data in order to conserve time and space" (Dörnyei, 2007:209). One example would be when the mean and range (minimum and maximum values) of a variable, e.g. the scores on a vocabulary test, are presented for a group of learners, such as a school class. The *null-hypothesis* states that there is no difference between boys' and girls' scores.⁴⁶ However, if there happens to be a difference between boys' and girls' scores in the class, for example the girls outperform the boys, based on descriptive statistics it is only possible to say that in this particular class, the girls scored higher than the boys. However, with the help of *inferential statistics*, it would be possible to examine if the difference is significant in the statistical sense. By doing so, generalizations beyond the particular sample are possible and the null-hypothesis may be rejected. Thus, if a result is statistically significant in a sample, that result is generalizable to the statistical population and everybody can "utter a sigh of relief", as Dörnyei (2007:210) puts it. This means that if a result is *not* statistically significant, it might occur in the sample only due to chance.

Probability is very important in science and it is particularly important in inferential statistics. *Probability* is defined as 'the expected relative frequency of a particular outcome', where an *outcome* is the result of an experiment or any other situation in which the result is not known in advance (Aron et al., 2005:117-118). *Statistical significance* is measured by a probability coefficient (p) which can range from 0 to 1. Something that has no chance of happening has a probability of 0, whereas something that is certain to happen has a probability

⁴⁶ Basically, the *null-hypothesis* is a statement that in the population, there is no difference (or a difference opposite to that predicted) between populations (Aron, Aron, & Coups, 2005:446).

of 1 (Aron et al., 2005:119). In applied linguistics, results are typically considered significant if $p < .05$, i.e. the probability that the results are due to chance is less than five per cent (Dörnyei, 2007:210). Significance is generally indicated in tables with asterisks, where one asterisk (*) indicates $p < .05$ and two (**) indicates $p < .01$ (i.e. less than one per cent). I will use such indicators when presenting my results.

To sum up, if a result is statistically significant, it means that it is not random or due to chance. It can, therefore, be generalized to the statistical population. A significant result indicates that the null-hypothesis can be rejected because the actual behavior of the data examined contradicts the assumption of the null-hypothesis.⁴⁷ An account of the most frequently used statistical tests or procedures in the present study now follows (sections 6.3.6.1–6.3.6.7).

6.3.6.1 *Independent samples t test*

Independent samples *t* tests are used for comparing the results of two groups that are independent of each other, for example the scores for boys and girls on the vocabulary tests in the present study. By using the *t* test, the difference between two sets of scores is examined to determine whether it is big enough to reach statistical significance. The general assumption of the test is that the examined variable has a normal distribution for each of the investigated populations (i.e. “boys” and “girls” in my example) (see e.g. Aron et al., 2005:272; Dörnyei, 2007:215).

6.3.6.2 *Paired samples t test*

Paired samples *t* tests are used when two sets of scores (i.e. two variables) for the same group are to be compared, for example learners’ grades in English and Swedish. The paired samples *t* test is also used when the same informants are measured more than once (Aron et al., 2005:229; Dörnyei, 2007:215). I use the paired samples *t* test when examining the results of the four school classes over time on speaking tests 1, 3, and 5.

6.3.6.3 *ANOVA*

When the means of three or more groups are to be compared, a one-way analysis of variance, ANOVA, is used. Two variables are needed in ANOVA: a

⁴⁷ In SPSS, *two-tailed tests of significance* is the default: “Hypothesis-testing procedure for a nondirectional hypothesis; the situation in which the region of the comparison distribution in which the null hypothesis would be rejected is divided between the two sides (tails) of the distribution” (Aron et al., 2005:448).

dependent variable, which is the target variable to be compared (e.g. EE or OP), and an independent variable, which is the grouping variable (e.g. the four school classes or the EE subsets) (see e.g. Dörnyei, 2007:218-219). Whenever differences between three or more groups are investigated in the present work, ANOVA is the statistical analysis used.

6.3.6.4 *Chi-square*

Chi-square (also Chi-2) tests are used for nominal data or in situations when it cannot be assumed that the examined population distributions are even roughly normal (Aron et al., 2005:333). In the present study, chi-square tests are used to compare the responses of my sample to certain questions in the questionnaire with those of the *NU-03*. It is worth pointing out that chi-square cannot be used when the expected value of one cell is less than five. As an example, if fewer than five informants tick option X on a particular item in a questionnaire, chi-square cannot be used.

6.3.6.5 *Correlation analysis*

Correlation analysis is used to examine the relationship between two variables. A correlation coefficient (r) is computed between the variables and it can range between -1 and 1. A high value of r is an indication of a strong relationship (Dörnyei, 2007:223). This means that if an individual scores high on one variable, s/he is also likely to score high on the other. In contrast, a correlation coefficient which is close to zero implies that the correlation is negligible and if r indeed equals zero, there is no correlation at all. Negative correlations suggest inverse relationships, e.g. the higher the value of variable X, the lower the value of variable Y (see e.g. Byström, 1973:138-144; Dörnyei, 2007:223-224).

In contrast to Dörnyei (2007:223), I will refrain from discussing the results of my correlations in chapters 7–9 in terms of strength, i.e. as “strong” or “weak”, because what counts as “strong” or “weak” is always relative. The same value of r may be referred to as “strong” in one context but not in another (Aron et al., 2005:105-106; Byström, 1973:138). Thus, I will instead focus on describing the correlations in statistical (neutral) terms: “positive”, “negative”, “negligible”, or “non-existent”. When correlations for different groups (e.g. boys and girls) are compared, the results are discussed relative to each other. In chapter 10, however, when I sum up all results and interpret them, correlations will be discussed also in terms of strength (cf. above).

Correlation analysis is a frequently used method in the present study. The EE variable is commonly one of the variables in my correlations, and it will be

correlated with variables such as the OP grade and self-efficacy. I may add that in applied linguistics research on motivation and achievement, r between .3 and .5 was considered meaningful (Dörnyei, 2007:223), which is something my values of r can be compared with. Furthermore, two tests that correlated with each other in the order of .6 were considered to basically measure the same thing (Dörnyei, 2007:223).

It is worth pointing out that a statistical correlation between two variables indicates that there is a real relationship between the variables, but it does not reveal anything with regard to a direction of causality between the variables. In fact, in any correlations between variables X and Y, there are at least three possible directions of causality: X could be causing Y, or vice versa, or some third factor could be causing both X and Y (Aron et al., 2005:75). It is also worth pointing out that the correlation coefficient describes linear relationships (Dörnyei, 2007:224). To understand what the coefficient stands for – to make r tangible – it is common to square its value “because the result thus obtained represents the proportion of the variance shared by the two variables (i.e. the variation in the score on one variable which is accounted for by the other)” (Dörnyei, 2007:224). For example, for the tests mentioned above ($r=.6$), 36% (i.e. $.6^2$) of the variance in the scores is explained by the correlation between the variables rather than by some other factor(s) or by chance (cf. Aron et al., 2005:67-75; Dörnyei, 2007:223-227).

There are several types of correlation coefficients and I use two in my study: Pearson (r) and Spearman (r_s). The Pearson correlation coefficient (r) shows the degree of linear correlation between two variables based on actual values. There are two requirements for using the Pearson correlation coefficient: (1) data must be of interval type and (2) the distributions of each variable in the population must be assumed to be normal or close to normal (Byström, 1973:142).⁴⁸ The Spearman rank correlation coefficient (r_s) gives a measure of the degree of concordance between two rankings (Byström, 1973:142). Thus, as the name suggests, the correlation is computed from the ranks of the data in a rank order, rather than from the actual values (Dörnyei, 2007:230). Spearman (r_s) can be used more easily than Pearson (r) because the only condition that has to be satisfied is that each variable is at least based on ordinal data (Byström, 1973:142). Generally, scattergrams are helpful in examining correlations, to make sure that the pattern is linear and not

⁴⁸ In practice, Pearson (r) is also commonly used for ordinal data (Byström, 1973:142).

curvilinear (Aron et al., 2005:74). Therefore, some of my figures are in the form of scattergrams.

6.3.6.6 Backward linear regression analysis

With backward linear regression analysis, a function provided by SPSS, it is possible to examine the relative importance of various variables in relation to a dependent variable. In the present study, backward linear regression analysis was used in order to investigate which EE activities are more important than others for oral proficiency and vocabulary respectively.

6.3.6.7 Cronbach's alpha and index variables

Cronbach's alpha is a widely used measure of reliability which reveals the extent to which the items of a measure assess a common characteristic; it should have a value of at least .6 or .7 and preferably closer to .9 (Aron et al., 2005:383). According to Dörnyei (2007:207), Cronbach's alpha should be at least .7 in L2 research. If items are found to assess a common characteristic – the same feature – they can be used to create an *index variable*. Thus, an index variable is based on (and created from) a minimum of two variables.

In order to identify questionnaire items whose responses cluster and thus make up a potential index variable, hierarchical cluster analysis (a tool in SPSS) was used on all questions that I considered assessed a common characteristic. The identified items were then chosen for reliability analysis. This type of analysis includes an option which computes what Cronbach's alpha would be if a particular question were deleted. This option is a useful tool because it makes it possible to see which questionnaire items in fact make up the best index variable. I used this procedure when identifying and creating the index variables for questionnaire data. To this end, Cronbach's alpha was very helpful in creating reliable index variables. For example, in my study, the questions listed in Table 6.8 aiming to measure “self-efficacy” could be used to create an index variable because their responses were sufficiently similar and reliability was high. Another index variable I used was created from the two vocabulary tests.

I have now accounted for the material and methods that were used in my study. In the following three chapters, I present and discuss my results, beginning with the results on EE and OP in chapter 7.

7 Extramural English and oral proficiency: Analysis and results

In this chapter I present analyses and results on students' extramural English and oral proficiency. The chapter begins with the results on extramural English based on data from the language diaries (section 7.1), followed by similar data from the questionnaire (7.2). In section 7.3, results are presented regarding oral proficiency; that section is based on assessment data from the five speaking tests. A study on pauses and oral fluency based on data from ten of the students in the sample is also accounted for (7.3.5). In section 7.4, results on EE and OP are correlated. The chapter ends with a summary and conclusion (7.5).

7.1 Extramural English based on language diary data

The results presented below are based on language diary data. I first treat the overall amount of EE that students reported, i.e. the time/amount of EE is in focus. This will be followed by data for the eight different EE activities that the language diary listed. The amount of reported time per EE activity will be presented for the whole sample as well as for the four classes and gender. After that, I present results on the amount and type of EE for the so-called EE subsets 1-5 (see below).

The EE mean for the whole sample (N=80) was 18.4 hours/week and the median 15.9 hours/week. The lowest value was 0 hours/week and the highest 57.1. Individual variation was large (STD 12.9) (see Fig. 7.1). Figure 7.1 shows that the distribution of the EE variable is skewed to the right. This means that the right side of the peak, which is a visual description of the number of students with high amounts of EE, is long and spread out like a tail. In a perfect normal distribution, the bars would approximate the precise mathematical distribution called the normal distribution or the normal curve, and the normal curve would be bell-shaped (Aron et al., 2005:107-108). The EE variable is, however, not perfectly bell-shaped. Previous studies have shown that there is huge individual variation in spare time activities (see e.g. Medierådet, 2008:26-27), something which explains the skewed distribution of the variable.

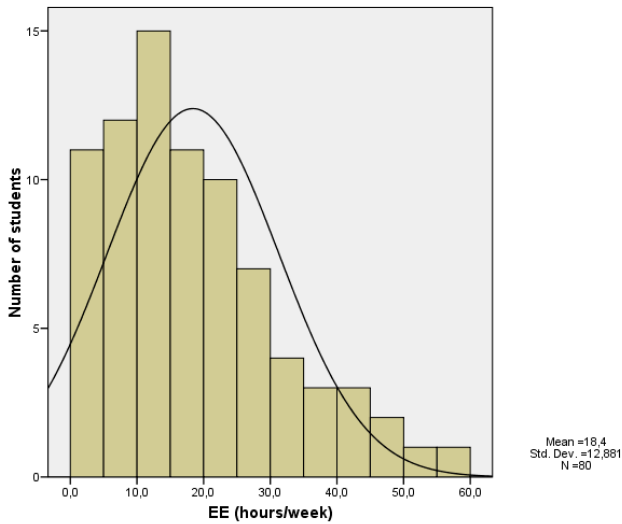


Figure 7.1. Students' total amount of EE (the EE variable).

As mentioned in chapter 6, students were encouraged to report the time spent on each of the following eight EE activities as truthfully as possible: “reading books”, “reading newspapers/magazines”, “watching TV”, “watching films”, “surfing the Internet”, “playing video games”, “listening to music”, and “other activity”. Overall results show that the means (hrs/w) for the time spent on these activities varied a great deal. As can be seen in Table 7.1, most time was spent on music, which was followed by video games, TV, films, the Internet, other activities, and at the end, the two reading categories, books followed by newspapers/magazines. The amount of time spent on reading newspapers or magazines appears almost infinitesimal in comparison to the more popular activities. Also, the time spent on books was very limited in comparison to the other spare time activities. The order of popularity of EE activities in my study is almost identical to the one found in Forsman (2004; cf. section 3.2).

Table 7.1. Amount of time (hrs/w) spent on the EE activities.

N=80	Music	Video games	TV	Films	Internet	Other	Books	Newsp. / Mag.
Mean	6.58	3.95	3.71	2.85	.70	.33	.20	.02
Median	3.75	.50	3.11	2.60	.08	.00	.00	.00
Min.	.00	.00	.00	.00	.00	.00	.00	.00
Max.	41.00	40.05	25.33	13.50	6.00	12.04	6.00	.67

There was huge individual variation in the data as can be seen in, for example, the different minimum and maximum values for the different activities. Another way of illustrating this individual variation is to study the percentage of students who in fact had no EE at all, or at least they did not report having any contact with EE (cf. “Minimum”, Table 7.1). Percentages for students with zero hours per week for the various activities are given in Table 7.2 along with the actual number of students per activity.

Table 7.2. Percentage and number of students with no reported time per extramural English activity.

	Music	Video games	TV	Films	Internet	Other	Books	Newsp. / Mag.
%	5	44	8	6	49	78	76	91
N	4	35	6	5	39	62	61	73

Since music was the EE activity where students reported the highest number of hours overall, it was no surprise that almost all students also listened to music; only five per cent stated that they never listen to music. The figure was similar for films, another category which had a high total amount of time in the sample. Table 7.1 reveals that there were students who spent the equivalent of a fulltime working week listening to music or playing video games. We all recognize typical “extreme music students”; they tend to wear their earphones almost all the time. The possibility of students having reported times for concurrent EE activities should be pointed out here. For example, they might have listened to music while using the Internet (cf. Forsman, 2004:85).

One reason why the extreme video game-playing student reached high values of EE is that he (and it is almost always a he, see section 7.1.1) took part in so-called daily raids with his guild in the popular online game *World of Warcraft* (*WoW*), a massively multiplayer online role-playing game (MMORPG, or simply MMO). The typical *WoW*-player is well described in Linderoth and Bennerstedt (2007:9). Normally a raid begins in the late afternoon (i.e. after school) and it lasts approximately three hours. The playing time is often increased at the weekend. The second most popular game with the boys in my study was *Counter-Strike* (an online, multi-player, strategic first-person shooter (FPS) video game), whereas girls most frequently reported that they play *The Sims* (an off-line, single player, strategic life-simulation video game) and *Burnout 3* (an off-line, single- and multi-player, high-speed racing video game). The results on what games boys and girls in my study played resemble findings on video game playing for Swedish youth in general (cf. Medierådet, 2008:33).

I have some additional comments to make regarding use of the Internet, TV, and the open “other” category. First of all, almost half the students in the sample did not surf the Internet in English (see Fig. 7.2), something which I found rather surprising. However, a comparison with data on spare time activities in Swedish, also obtained from the language diaries, revealed that as many as 88% reported time spent on the Internet there. That is, a majority of the students spent time on the Internet regularly, but many surfed more on Swedish sites than on English ones. Moreover, it should be mentioned that “surfing the Internet” is an extremely broad EE activity which necessarily involves some amount of reading in English. In addition to reading, surfing the Internet may also entail activities such as listening to English and writing in English. Second, when reporting time for watching TV, it was possible for students to specify the programs they watched. Among the most popular TV-shows were *The Simpsons* and *Desperate Housewives* (see also section 7.2.1). Finally, the open category in the diary was particularly interesting to study since the students in fact had to list/write down what they did. The following were some examples of students’ “other” EE activities: chatting, both text and voice; talking on the phone to friends in other countries; talking via Skype to co-players while playing video games online; writing comments on a video game discussion forum; and writing essays: “first I drafted a 4-page essay on the computer and then I wrote the final version by hand” [my translation].

7.1.1 Results for the four participating classes and gender

There were differences between the overall amounts of EE in the four participating classes (see Table 7.3). The class with most time spent on EE was class 3, followed by classes 4, 1, and 2. A comparison of EE for the four classes showed that the differences between the classes were statistically significant ($p=.035$). Independent samples *t*-tests were used to check if there were any significant differences between boys and girls. For the whole sample, boys spent more time on EE than girls, but the difference was not significant. However, for class 1, the difference was significant and for class 3, it was very close to being so.

Table 7.3. EE and the four participating classes; significance values for gender differences.

Class	Gender	Mean EE (hrs/w)	N	Sig. (2-tailed)
1	Boys	22.9	10	.031*
	Girls	8.9	10	
	Total	15.9	20	
2	Boys	13.1	10	.914
	Girls	13.8	8	
	Total	13.4	18	
3	Boys	30.2	9	.062
	Girls	20.6	13	
	Total	24.5	22	
4	Boys	16.6	7	.537
	Girls	19.7	13	
	Total	18.6	20	
Total	Boys	20.8	36	.136
	Girls	16.4	44	
	Total	18.4	80	

*Sig. at .05 level

Results for the eight EE activities are illustrated in Figures 7.2–7.9. Note that the scale of the y-axis varies with the variable. The purpose of these figures is to illustrate visually where the major differences between the classes and genders were found. An ANOVA revealed that for music, the differences between the four classes were in fact significant ($p=.048$). Class 3 had the largest amount of time spent on music (10.2 hrs/w), followed by class 4 (6.5 hrs/w), class 1 (5.2 hrs/w), and class 2 (3.7 hrs/w). Regarding gender, the differences were significant for two of the EE activities, namely video games ($p=.000$) and the Internet ($p=.016$). Boys spent more time than girls on both of these (video games 7.9 hrs/w vs. .7 hrs/w; the Internet 1.1 hrs/w vs. .4 hrs/w), which is a gender difference in line with previous findings (cf. Medierådet, 2008; Sylvén, 2004; see also chapter 3). No other significant differences were found for gender. In addition to bar graphs, it is possible to present the EE activities for boys and girls in separate pie charts, where each EE activity makes up a proportion of the total amount of EE. The gender differences become salient when boys' and girls' pie charts are compared (see Figures 7.10 and 7.11).

7.1.1.1 *A comparison with Swedish and other languages*

Since the language diary data also included data on extramural activities in Swedish and other languages, it was possible to analyze those and make comparisons with results for EE. Here I will report on gender differences which were found to be significant for certain extramural activities in Swedish and other languages.

Regarding activities in Swedish, a significant difference was found for gender and reading books ($p=.024$). On average, the girls read 1.5 hours per week, which was about twice the time boys spent on reading. In English, however, the boys in my sample in fact read more than girls, but the time reported for reading books in English was very limited (cf. Table 7.1) and no significant difference was found for gender. In fact, among those who read, only one student reported reading extensively (six hours per week) in comparison with the others. The second highest reported value was 1.6 hours. In total, five students read one hour or more per week in English. Furthermore, in the data for Swedish, a gender difference was also found for video games ($p=.012$), where boys reported playing more than girls. However, in comparison with playing video games in English, boys' playing time was very low (.9 vs. 7.9 hrs/w).

Extramural activities in other languages were limited but interesting nevertheless. For instance, students were actually exposed to more languages than we would perhaps expect. Some students commented on extramural activities in the modern languages that they study (i.e. French, German, Spanish), but there were also comments on as many as twelve other languages: Albanian, Arabic, Danish, Estonian, Finnish, Italian, Japanese, Kurdish, Norwegian, Serbo-Croatian, Swahili, and Thai. Some of these reported extramural activities had to do with contacts with relatives but there were also several examples of extramural activities in foreign languages out of what seemed to be pure joy or keen interest. For instance, there were students who tried to pick up foreign language vocabulary from their immigrant friends. Others listened to songs in different foreign languages. One boy reported watching Japanese films. Several students had extramural German when they watched an Austrian TV broadcast of a high profile interview.¹ And one girl noted that she found Arabic hard to learn on her own: "Tried reading in Arabic – it wouldn't work" [my translation]. The only activity where a significant

¹ The first interview with Natascha Kampusch, who escaped from her kidnapper after eight years of captivity.

difference was found between boys' and girls' times was TV ($p=.036$), where girls spent almost half an hour per week watching TV in languages other than English and Swedish whereas boys reported doing so very little, less than .1 hour/week.

7.1.2 Results for subsets based on the EE variable

As was discussed in chapter 6, I will present results not only for the participating classes and gender (see above), but also for a group design based on the EE variable, i.e. a clustering of the students based on language diary data. From the raw data in the EE variable (Fig. 7.1), five subsets emerged (see Figure 7.12). The purpose of clustering the students in this way was, as I explained in chapter 6, to create subsets which approximate the original distribution of the students based on their EE. Due to the nature of the EE data, the five subsets are slightly skewed to the right (cf. Fig. 7.1). Figure 7.12 shows that boys and girls are represented in each of the subsets. There is an even distribution of boys and girls in subsets 1 and 3, but in subset 2 there are more girls than boys and in each of subsets 4 and 5 there are more boys than girls. In the whole sample, there were more girls (44) than boys (36). Thus, the gender distribution in the five subsets can be considered even, but with boys perhaps being slightly overrepresented in subsets 4 and 5. Moreover, the distribution of students showed that subsets 4 and 5 are small in comparison with the other subsets. The explanation lies in the EE data. That is, the number of students who spent a great deal of time on EE was low and, therefore, subsets at the higher end of the scale out of necessity become small.

Descriptive statistics for the five subsets are given in Table 7.4. As has already been mentioned, subsets 4 and 5 were small. Nevertheless, the differences between the subsets were statistically significant ($p=.006$), i.e., it is very likely that the differences are not due to chance.²

² Initially an alternative group design was also considered, where subsets 4 and 5 were combined into one subset of ten students. The idea of adopting that design was abandoned, however, as the current design of five subsets was found to be a closer and better representation of the EE variable.

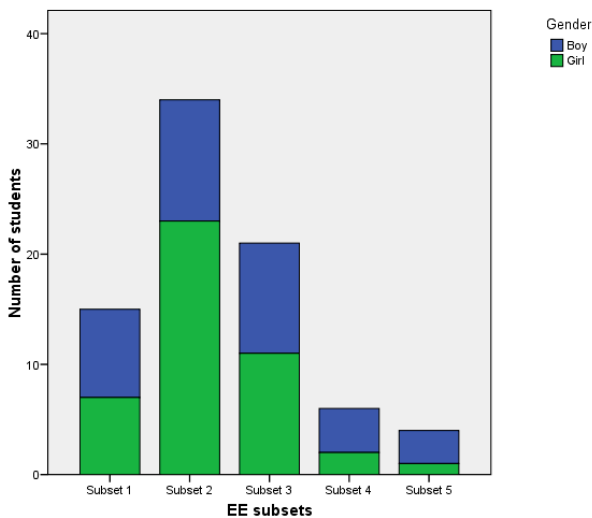


Figure 7.12. The five EE subsets and gender distribution.

Table 7.4. Descriptive statistics for the five EE subsets.

EE subset	Time interval (hours/week)		Mean EE (hours/week)	N	Std. Deviation
	From	To			
1	≥ 0	≤ 8	3.5	15	2.5
2	> 8	≤ 20	13.1	34	3.4
3	> 20	≤ 34	25.3	21	4.1
4	> 34	≤ 44	40.2	6	3.3
5	> 44	≤ 58	50.0	4	5.2
Total	≥ 0	≤ 58	18.4	80	12.9

The distribution of the four classes in the five subsets was uneven, as is shown in Figure 7.13. That is, not all classes were represented in each subset. For example, in subset 1 with the lowest time interval of EE there was not a single student from class 3 even though subset 1 was fairly large, comprising 15 students. Moreover, in subset 5 where the students had the highest time interval of EE, there was no student from class 2. The fact that not all classes were represented in subset 5 is not very surprising since it was the smallest subset with only four students. Class 1 was the only class with students in all five subsets which might indicate that, at least in terms of EE, this was the most heterogeneous class. Here it should be noticed, however, that also class 4 spanned over all five subsets but with no student in subset 4.

Results for the EE subsets on time spent on the eight EE activities are shown in Table 7.5. Significant differences (01 level) between the five subsets were found for all activities except two: “reading books” and “reading newspapers/magazines”.

Table 7.5. Descriptive statistics for the EE subsets and the eight EE activities.

EE subset	EE activity (hours/week)							
	Books	News. /mag.	TV	Films	Internet	Video games	Music	Other
1	.1	.0	1.0	.8	.0	.5	.9	.0
2	.1	.0	3.8	3.1	.5	1.8	3.7	.1
3	.5	.1	5.5	3.3	.8	5.1	9.5	.5
4	.0	.0	4.0	5.1	2.4	11.1	17.3	.3
5	.0	.0	2.1	3.2	1.4	18.3	21.1	3.0
Total	.2	.0	3.7	2.8	.7	3.9	6.6	.3

7.2 Extramural English based on questionnaire data

Eleven items in the questionnaire addressed extramural English. They were linked to, for example, travels abroad and specific EE activities, such as watching TV, playing video games, watching films, listening to music, reading, and participation in role-playing games. In this section, I present the results on EE based on questionnaire data that relate to three areas: media (TV, films, music), reading, and online role-playing games. The results will be presented in that order, and for each area the results are first presented globally (the whole sample) and after that, when applicable, separately for the four classes, gender, and the EE subsets.

7.2.1 Media

According to the responses to question 11, item a, all students in my sample had TVs. When asked if they have a favorite English-speaking program (question 14, item a), 82 per cent responded in the affirmative. Among the programs listed the most frequently were *Lost*, *The Simpsons*, *Prison Break*, and *The O.C.* Of the 80 students in the sample, 66% said that they watch English-speaking TV-shows with Swedish subtitles on a daily basis and 30% said that they did so once or a few times a week (question 16, item a). Only four per cent reported watching such programs to a lesser extent, i.e., once or a few times per month. Question 18 specifically asked how often the students watch English-

speaking TV-programs *without* Swedish subtitles. This was less common in the sample. Two students said that they watch such programs on a daily basis. Eleven students, or 14% of the sample, did it once or a few times per week. A majority (52%), however, stated that they never (or very rarely) watch English-speaking TV-programs without subtitling support, whereas 29% reported that they may do so once or a few times per month.

The results for films (question 20, item a) were slightly different from those for TV, even though there were similarities. Like TV, film was a medium that the students favored.³ The majority (41%) watched films on a weekly basis (once or a few times per week) whereas those who did it on a daily basis made up 34% of the sample. The rest watched films once or a few times per month (24%) or never/almost never (1%).

My study shows that music was a very popular category of EE. In fact, as many as 70 students (88%) said that they listen to music with English lyrics on a daily basis. Eight students (10%) listened to music with English lyrics on a weekly basis and only two students replied that they listen less often. It is difficult, however, to determine how active students are when they listen to music. On the one hand, listening to music may be quite “inactive”, i.e. more of a receptive activity than a productive one (e.g. having music in the background while studying or playing video games; cf. Forsman, 2004:85). On the other hand, listening to music may also be “active”, i.e. more of a productive activity than a receptive one. For instance, some students commented in the diaries that they had written down the lyrics of a song in English or translated the English lyrics into Swedish.⁴ In spite of the difficulty of determining the level of activity involved in listening to music, music was undoubtedly a frequent and important EE activity for students in general. Data from the questionnaire support data from the language diaries in this respect: question 23 correlated positively with the EE category for music ($r=.410^{**}$). Another question that was related to music was question 24, item c, which asked if the students play an instrument. Only eight students replied “yes” to this question and these eight in fact reported listening less often to music in their spare time than those who did not play an instrument (5.2 hrs/w vs. 6.7 hrs/w). Apparently, there was no

³ It is possible that some students also considered films in question 18 (item a), when they were asked about English-speaking TV-programs without Swedish subtitles. However, not a single student listed a film as an example of such a program (18-b); instead they listed specific channels, CNN and MTV, and in addition to those channels these TV-series: *Scrubs*, *Jackass*, *Viva la Bam*, *Wild Boys*, *South Park*, and *Family Guy*. One student listed a program from Swedish educational TV (Swe. *Utbildningsradion*).

⁴ In one of the student interviews, a girl also mentioned singing karaoke in English, which is another type of active/productive music activity in English.

connection between playing an instrument and the amount of time spent on listening to music.

Results for the four classes on question 18, item a (How often do you watch English-speaking TV-programs without Swedish subtitles?) indicate that students in class 2 spent less time watching English-speaking TV-programs without Swedish subtitles than the other classes. When question 20, item a (How often do you watch English-speaking films?) was examined, it was found that differences in film habits between the classes were small.

Results for the four classes on question 23 (How often do you listen to music with English lyrics?) confirm that music was a very frequent activity of extramural English as it was popular with all four classes. However, relatively speaking, in comparison with the other classes, class 2 had a higher ratio of students who did not listen to music on a daily basis. In the whole sample, there was only one student (a boy, class 1) who reported that he never (or almost never) listens to music.

7.2.1.1 *Media and gender*

Overall there were few differences between boys and girls and their responses to media-related questions and I will briefly summarize the results here. It should be mentioned that it was possible to use chi-2 tests to examine the results of some of the questions. For example, a chi-2 test for question 14 (which was about whether or not the students have a favorite English-speaking TV-program) revealed no significant difference between boys' and girls' responses, but it was more common that girls responded "yes" (86%) to this question than boys (78%). Among the listed favorite programs, some were TV-series that both boys and girls appreciated, e.g. *Lost* and *Prison Break*. However, there were also TV-shows that only girls listed (e.g. *Desperate Housewives* and *The O.C.*) and others that only boys listed (e.g. *Married with Children*, *Pimp My Ride*, and *Buffy, the Vampire Slayer*), something which might indicate a genre-related gender difference. Responses to question 18 showed that it was neither very common nor gender-related to watch English-speaking TV-programs without Swedish subtitles. Similar to the responses to question 14, no differences for gender were found for films (question 20, item a). Question 20, item b, asked which language students most commonly choose for subtitling. Here, 77 students responded and 90% stated that they preferred Swedish subtitles. Two students chose to watch the film without subtitles, and only one student opted

for English subtitles.⁵ That is, when the students watched films, they commonly did so with L1 support.⁶ Regarding listening to music with English lyrics (question 23), no major differences between boys and girls were found. Finally, as has already been mentioned, in my study there were eight students who played an instrument; seven of those were girls and one was a boy. Here there appears to be a difference between genders, but the overall number of students who played an instrument was too low to draw any conclusions.

7.2.1.2 *Media and the EE subsets*

In this section, five media related questions from the questionnaire (14-a, 16-a, 18-a, 20-a, 23) are examined in relation to the five EE subsets. I summarize the results and comment briefly on some of the findings.

Question 14, item a, asked about having a favorite English-speaking TV-program. A majority ($\approx 60\%$) of those who did *not* have a favorite English-speaking TV-program were in subset 1, something which can be considered logical since subset 1 comprised the students with the least amount of EE overall. Responses to question 16, item a (watching English-speaking TV-programs with Swedish subtitles), showed that for subsets 1 and 2, such programs seemed to be an important source of extramural English because approximately half of the students in subset 1 and more than half in subset 2 responded that they watch English-speaking TV-programs with Swedish subtitles on a daily basis. The four students in subset 5 all said that they watch such programs on a daily basis. As for watching English-speaking TV-programs *without* Swedish subtitles (i.e. question 18, item a), two students (one in subset 2 and one in subset 3) said that they watch such programs on a daily basis. Students in all five subsets were found in the remaining three response categories (“once or a few times per week”; “once or a few times per month”; “never or almost never”). Regarding the frequency of watching English-speaking films (question 20, item a), students from all five subsets were found in three of the response categories (“daily”; “once or a few times per week”; “once or a few times per month”). There was only one student who ticked the box “never or almost never” to this question.

As has already been shown, music was a very popular EE activity. Responses to question 23 confirmed this fact since almost all students said that

⁵ Of the remaining students who answered the question, 2% stated that they had never had the chance to make any choices regarding subtitles and 3% were missing values.

⁶ As was previously mentioned (section 6.1.5), some students had other first languages than Swedish. For these students, the subtitling support was in L2 Swedish.

they listen to music with English lyrics on a daily basis and, consequently, all subsets were found in that response category. Of those who listened less often than daily, none was in subsets 4 and 5, which makes sense since these two subsets included the students with most hours of EE and we know that within the EE variable, music contributed proportionally with the most time.

7.2.2 Reading

In the questionnaire, question 7 addressed reading habits and four types of text were in focus: (1) books, (2) newspapers, (3) comics, and (4) other texts (manuals, lyrics etc.). The question addressed reading habits in both Swedish and English. As in section 7.2.1, I summarize the results and comment briefly on what was found. I begin with the responses that relate to reading habits in English.

Very few students in my sample read newspapers or comics in English. Only a handful of students mentioned that they read such texts, and it was not done often. The most common reading category of the four was the last one, i.e. “reading other texts”, e.g. manuals and lyrics. Here, seven students ticked “daily” and another eleven “once or a few times per week”. Reading books in English was more common than reading newspapers or comics, a result which rhymes with what was found for the two EE reading activities presented in section 7.1. Four students said that they read English books once or a few times per week and nine students mentioned that they did so once or a few times per month. On the whole, results on reading based on questionnaire data confirm the results found for reading based on language diary data.

Not surprisingly, students read more in Swedish than in English. It was common that students read newspapers (13 responded “daily”; 33 “once or a few times per month”) and books (10 “daily”; 33 “once or a few times per month”). There were 25 students who never (or almost never) read newspapers and 24 students who never (or almost never) read books in Swedish. In other words, about one third of the sample rarely read anything at all in Swedish in their spare time. Five students read comics daily and another fourteen stated that they did it on a weekly basis. Reading other types of texts was not common at all. More than half of the students responded that they never or almost never read such texts; only two said that they did so daily, and another ten that they did so weekly. Thus, reading other types of text turned out to be a more common activity in English than in Swedish.

7.2.2.1 *Reading and gender*

A chi-2 test showed that no significant gender difference was found for question 7, item a (Do you read in your spare time?). However, when responses to the four reading categories in question 7 were compared with reading in English, some interesting results appeared. For instance, responses from the questionnaire again confirmed language diary data regarding reading books in English. In my sample, it was more common that boys read books than that girls did so. Four boys (11%) stated that they read English books once or a few times per week; no girls checked this box. Six girls instead said that they read English books once or a few times per month; three boys also reported that they read on a monthly basis (see Fig. 7.14). But, just as the language diary data showed, generally speaking, reading books in English was not common: 81% of the boys and 86% of the girls responded “never or almost never” to this question.⁷ Reading books in Swedish was more common among girls than boys in my study, something which also confirms results based on language diary data.

Very few reported reading newspapers in English. Only one student (a boy) replied that he reads newspapers on a daily basis and another student (a girl) said that she reads them weekly. Another two boys along with a girl read English newspapers on a monthly basis whereas the rest said that they never (or almost never) do so. It was, generally, more common to read newspapers in Swedish. Again, more boys ($\approx 23\%$) than girls ($\approx 11\%$) said that they read newspapers on a daily basis. Similarly, among those who never (or almost never) read the paper in Swedish, there were more boys than girls. As for the other two reading categories in Swedish, the results were very similar for both boys and girls and neither reading comics nor reading other texts was very common.

7.2.2.2 *Reading in English and the EE subsets*

As has already been shown, overall there were few students who spent time on reading. Nevertheless, it might be interesting to see in which EE subsets the readers were found. Similar to what was discussed in sections 7.2.2 and 7.2.2.1, four categories of text are in focus: (1) books, (2) newspapers, (3) comics, and (4) other texts (manuals, lyrics etc.).

⁷ Due to the small number of students who indeed read, no Chi-2 analyses could be made to see whether the differences between boys and girls were significant or not.

Not a single student responded that s/he reads English books on a daily basis. Instead, those who read books in English read on a weekly or monthly basis and they were found in subsets 2, 3, and 4. Reading newspapers in English was even rarer, but those who read newspapers were found in subsets 2 and 3 (“daily-readers” only in subset 3). Reading comics in English was extremely rare and was only reported to occur monthly (scattered students were found in subsets 2-5). However, reading texts of the fourth category (“other texts”) was more common and here “daily readers” could be found in subsets 2-5. Those who responded “weekly” in this response category were found in subsets 2, 3, and 5, whereas those who responded “monthly” were found in all subsets except subset 4. Needless to add, the rest of the students, which was the majority, responded “never or almost never” and they were found in all five subsets.

7.2.3 *Online role-playing games*

Almost all students (78 out of 80) had a computer at home. Among those who owned a computer, only one reported lacking access to the Internet (question 9). In the questionnaire, question 15 (item b: 6) specifically asked about habits regarding playing games on the computer.⁸ About one fourth of the sample stated that they play on a daily basis. Those who reported playing on a weekly basis together with those who reported monthly made up 16% each, whereas 35% of the students said that they never (or almost never) play computer games.⁹

Results for the four classes for the same question (15, b: 6) are shown in Figure 7.15. The figure shows that two of the classes (1 and 4) had very similar game playing habits. Class 3 differed from classes 1 and 4 by having a larger proportion of students who never (or almost never) played, but class 2 differed even more in this respect. Figure 7.15 reveals that among the four classes in my sample, class 2 was the class which was least involved in playing games on the computer. In addition to having the largest proportion of students who responded that they never (or almost never) played computer games, class 2 also had the smallest proportion of students who responded that they play daily (approximately 5%); something which can be compared with about 25% in class 3, and more than 30% in classes 1 and 4.

⁸ Note that the question asked about playing habits in general and did not involve any reference to what language the student used. That is, the figures given might very well also include playing games in Swedish, even though this is generally less common than playing games in English. Note also that the question refers to playing games on the computer (Swe. *dator*). In the language diary, the game playing context is much broader (Swe. *TV-/Dataspel*).

⁹ Non-responses (i.e. missing values) make up the remaining percentage ($\approx 9\%$).

There was an additional question in the questionnaire (question 24) that concerned role-playing games in general in English (item g) and in Swedish (item e). Eight students (all boys) answered in the affirmative that they were involved in role-playing games in English. All of the examples of games that the students listed were so-called online RPGs (role-playing games), i.e. no one mentioned other forms of RPGs (e.g. board games or card games). The dispersion of the eight boys who said that they play role-playing games in English were as follows: three from class 3, two each from classes 2 and 4, and one from class 1. Finally, for role-playing games in Swedish, only one student (a boy) responded “yes”.

7.2.3.1 Online role-playing games and gender

The results for boys and girls were almost a mirror-image of each other when the general question on playing computer games (Q15-b: 6) was split for gender (see Fig. 7.16). Boys played much more often than girls. The results echo those of other sources (cf. e.g. Medierådet, 2008:33; Sylvén, 2004:226). With regard to role-playing games and gender (Q24-e, Q24-g), only boys responded in the affirmative. They were shown to be involved in online RPGs, discussed above (section 7.2.3).

7.2.3.2 Online role-playing games and the EE subsets

From students’ responses to question 15 (item b: 6) it was possible to learn in which EE subsets frequent and infrequent players of computer games were found, and the results were interesting (see Fig. 7.17). Subset 1 (low amounts of EE) did not have any student among those who replied that they play on a daily basis, a result that could perhaps be expected since game playing habits tend to lead to several hours of EE per week. At the other end of the scale, among the four students in subset 5, i.e. those with very high amounts of EE, only one was not a game player. Three of the four students in subset 5 responded that they play on a daily basis, which could be expected, and they were the three boys in subset 5. The student who did not play on a daily basis but instead responded “never or almost never” was the girl. That is, the different game playing habits that are related to gender also showed up in the extreme subset 5.¹⁰

¹⁰ Note that no significant difference for gender was found for question 15, item b: 2, which had to do with surfing the Internet. This result does not agree with the significant difference between boys and girls that was found for the EE activity “surfing the Internet” based on language diary data.

7.3 Oral proficiency based on assessment data from the speaking tests

In the present section I give an account of the results on the students' OP. All results are based on assessment data from the five speaking tests. The section is made up of six subsections. The first subsection (7.3.1) discusses the results for the four participating school classes and gender. It is followed by subsection 7.3.2 which is about the progression over time for the classes. Subsection 7.3.3 presents descriptive statistics and analyses of the results for the five EE subsets. In subsection 7.3.4, I account for findings on the factorial grade for fluency. Subsection 7.3.5 also has a focus on fluency, but here I present the results of a quantitative study on pauses and oral fluency based on data from a selection of ten students in the sample. In the last subsection (7.3.6), students' results are viewed from a different perspective: the clustering of the students is based on the students' level of oral proficiency rather than the amount of time spent on extramural English.

7.3.1 Results for the four participating classes and gender

In contrast to the EE variable (Fig. 7.1), the OP grade variable was normally distributed, as is shown in Figure 7.18. The number of students who completed all five speaking tests was 74 (cf. section 6.1.5). Girls had a slightly higher mean OP grade (3.5) than boys (3.2), but the difference was not statistically significant ($p=.196$).

In Table 7.6, mean scores for all four classes are shown, for the five tests as well as overall. Class 3 had the highest mean score on all five tests and also overall, and class 2 had the lowest ones, with classes 1 and 4 in between. Class 4, however, had higher mean scores than class 1 on four out of five tests. It was possible to compare the results on Tests 1, 3, and 5 since these tests were similar in design. A discussion of the results on these particular tests follows in a separate section (7.3.2). Results on Tests 2 and 4, however, are discussed below, one at a time, since the test designs for these two tests differ both from each other and from Tests 1, 3, and 5. The values for standard deviation indicate broad variation among learners, which is similar to findings in Iwashita et al. (2008; cf. section 4.3).

Table 7.6. OP grades for the four classes on Tests 1-5.

Class		Test 1	Test 2	Test 3	Test 4	Test 5	Tests 1-5
1	Mean	2.95	3.35	2.93	3.23	3.53	3.22
	N	20	19	19	19	19	19
	Std. Dev.	1.19	1.10	.75	1.12	1.17	.91
2	Mean	2.22	2.57	2.63	3.07	3.26	2.75
	N	18	18	18	18	18	18
	Std. Dev.	.76	1.18	1.03	1.11	.62	.85
3	Mean	3.62	4.19	3.62	3.90	4.59	4.00
	N	22	21	21	21	21	21
	Std. Dev.	.94	.90	.68	.93	.75	.63
4	Mean	3.04	3.43	3.30	3.04	4.08	3.31
	N	19	18	18	16	16	16
	Std. Dev.	.85	.95	.66	1.13	.65	.59
Total	Mean	2.99	3.41	3.14	3.34	3.88	3.35
	N	79	76	76	74	74	74
	Std. Dev.	1.06	1.17	.86	1.11	.98	.88

Test 2 was a long and highly structured test (see Hasselgren, 1996b). Since it was approximately twice as long as the other tests (see section 6.2.3), it was challenging for the students in terms of concentration. On the other hand, the various test tasks and material provided a great deal of support which helped the students in their production. As the test instructor, I felt that less proficient students in particular were aided by the actual design of the test. That is, all students managed to produce speech, something which at times was very challenging in Tests 1, 3, and 5, where some students easily fell silent. These students' moments of silence might be due to the test format adopted in the three tests, all national tests of English used in Sweden. One challenging task in these specific tests was, for example, to discuss matters that were introduced without much time for planning and which did not necessarily catch the students' interests (cf. chapter 6). In Test 2, even though parts 1 and 2 were demanding (picture-cued story-telling and picture-cued conversation) and revealed lacking vocabulary as well as lacking strategies, less proficient students were still able to speak in English. In some cases, students would ask their interlocutor or me in Swedish for certain words or phrases in English, and subsequently used them in the conversation. More proficient students were able to excel in the sense that they demonstrated knowledge of infrequent vocabulary elicited by the pictures (e.g. *stamp collector*, *lawn mower*, *hose*, *compost*, *rake*, *earphones*, and *window sill*) or they were able to avoid problems with the help of various strategies, such as circumlocutions, vague referencing, and asking for help but doing so in English.

In part 3 of Test 2, one of the tasks was to read a passage aloud; a task that proved very difficult for some students, as did another one where they were supposed to spell names in English, something which revealed that some ninth graders do not know the names of the letters in English. Some students solved this specific problem by writing down the name that they were supposed to say, and then they could pronounce the letters aloud.¹¹ By doing so, they avoided relying on their working memory only when spelling the name out loud. It is possible that some of the students who failed in spelling would have succeeded if visual support had been available.

Test 4 was different from the other four tests in that one week's planning time was allowed. The test was a prepared speech followed by a question and answer session. A great majority of the students came well prepared to the test session. For example, many brought photos or other memorabilia as props to aid their speech, and they also brought key words or drafts of their speech. Among those who had failed to prepare, most gave impromptu speeches to the best of their ability. At the individual level, some students showed a much better flow of speaking in comparison to previous performances. This was particularly true for normally less proficient students, some of whom had spent a great deal of time rehearsing their speech, at times from written manuscripts. My impression that less proficient students took the opportunity to prepare their speech was confirmed both in one of the interviews (student 4171) and also in several of the discussions that followed upon each speech. That is, many less proficient learners had taken the chance to rehearse at home, something which helped them to produce speech at a faster rate than they had previously managed to do, and they also used more advanced structures. However, a few of these students were all but unintelligible due to, for example, a very fast rate of delivery. It should also be mentioned that several of the learners used more advanced vocabulary in Test 4 (cf. section 8.7).

Some speech topics in Test 4 were, for instance, "My textbook in English – *Wings*", "Manchester United", "My grandfather", "When I went to Venice with my orchestra", "My dog", "When the tree in our garden fell down", "My trip to Greece", "Vampires", "The red army", "Sweden", "German tanks", and "My idol – Ronaldinho".

¹¹ In some cases, I wrote the name down for them.

7.3.2 *Progression over time for the four classes*

In Figure 7.19, results for the four classes on Tests 1, 3, and 5 are shown. Since these tests were very similar in design, a comparison of the results over time was possible. The time span from Test 1 to Test 5 was between eight and nine months and there was a clear progression for all classes during that period. Paired samples *t* tests confirmed that the improvement was significant for all classes from Test 1 to Test 5, whereas only class 2 showed a significant improvement between Test 1 and Test 3 (see Table 7.7). Furthermore, ANOVAs of the results for each class and test showed that the differences between the classes were statistically significant on all three tests (T1: $p=.000$; T3: $p=.001$; T5: $p=.000$). That is, of the participating classes in my sample, class 2 began the study at the lowest level and class 3 at the highest, whereas classes 1 and 4 were at almost identical levels at this point in time. At the end of the study, all classes had improved their results. Class 3 ended the study the way it started it, that is, at the highest level in relation to the other classes. Class 2 still had the lowest score but relatively speaking, class 2 improved more than for example class 1 did. The gap between classes 1 and 2 was smaller at the end of the study than at the beginning. Moreover, if we compare classes 2 and 4, which started out at a very similar level, Figure 7.19 shows that class 4 improved more than class 2. Of the four participating classes, class 1 showed the least progression over time.

The differences in progression over time for classes 1 and 2 can be explained. Both classes were from the same school and were taught by the same teacher but she described the classes as very different. Class 1 she described as an average class academically but not very hard-working, whereas class 2 was described as a below average class but with many highly motivated students and good work ethics. The results reported for class 3 can also be explained. In the sample, class 3 entered the study at the highest level of the classes in the sample. The school's overall grade year after year is well above average in comparison to the overall grade nationwide. This, of course, does not automatically mean that class 3 were bound to score well, but the probability of this class doing so was greater than for the other three classes. The teacher in class 3 described it as a typical class at that school: a rather homogeneous class where the students, in general, had good support from their parents. In comparison to the other classes, class 4 was the most heterogeneous class (cf. sections 6.1.4–6.1.5) in the sample. The teacher described it as an average class academically but with many interesting personalities in the class, which made it rather special. More girls than boys from this class took part in the study, which

might be a factor which explains the progress during the school year, since in Sweden, girls generally outperform boys, both overall and in the subject English (Björnsson, 2005:14-15).

Table 7.7. Improvement over time (paired samples t tests).

Class	Tests 1 and 3			Tests 1 and 5		
	<i>t</i>	df	Sig. (2-tailed)	<i>t</i>	df	Sig. (2-tailed)
1	.627	18	.538	-2.354	18	.030*
2	-2.106	17	.050*	-8.044	17	.000**
3	.761	20	.456	-5.460	20	.000**
4	-1.224	17	.238	-4.626	15	.000**
Tot	-.929	75	.356	-9.127	73	.000**

*Sig. at the .05 level, **Sig. at the .01 level

7.3.3 Results for the EE subsets

One of the main aims of my study was to map out students' EE and in meeting that aim, results for the five EE subsets on various measures are very important. In the present section, focus is on the mean oral proficiency grades, i.e. the OP grades, for these EE subsets.

In Table 7.8, the OP grades are shown for the subsets, with details for the sample as a whole and for the two genders. Note that the total number of students was 74 and not 80 (which was the total number of students who began the study), since six students did not complete all five speaking tests (see section 6.1.5). Of the six students who dropped out of the study, two were from EE subset 1, another two from EE subset 2, and one each from EE subsets 3 and 4. Using ANOVA, I found that the differences between the five EE subsets were statistically significant ($p=.003$).

The subset with the least EE, i.e. subset 1, clearly had the lowest mean OP grade. The other four subsets did not seem to differ much in terms of their oral proficiency, but the differences that did exist were statistically significant. This means that those with the highest amount of EE were also the ones that in fact had the highest mean OP grade (subset 5). It is noticeable, though, that subset 4 had a lower OP grade than subsets 2 and 3. The result is an indication that there is not a straightforward relationship between EE and OP in the sense that more time spent on EE activities automatically implies better scores on OP. The relationship is more complex than that, which means that analyses of pure amount (i.e. time) of EE and potential effects on oral proficiency need to be accompanied by further analyses regarding for example the types of activity students are involved in (see section 7.4.4).

Table 7.8. EE subsets and OP grades.

EE subset	Sample			Boys			Girls		
	OP grade	N	Std. Dev.	OP grade	N	Std. Dev.	OP grade	N	Std. Dev.
1	2.5	13	.62	2.4	8	.58	2.7	5	.69
2	3.5	32	.88	3.4	9	.93	3.6	23	.87
3	3.5	20	.84	3.6	10	.49	3.5	10	1.12
4	3.3	5	.69	3.2	4	.77	3.5	1	.
5	3.6	4	.45	3.8	3	.41	3.2	1	.
Total	3.3	74	.88	3.2	34	.81	3.5	40	.93

7.3.4 The factorial grade for fluency

As was described in section 6.3.3, the raters not only awarded an overall OP grade to each student on the speaking tests, they also awarded two factorial grades, one for *message and fluency* (henceforth *fluency*) and one for *language structures and vocabulary*. In this section, the focus is on the first of these two, i.e. on fluency.

The factorial grade for fluency had a normal distribution in the sample. The participating classes had different mean grades for fluency and their results were similar to the ones for oral proficiency. This was to be expected since the overall OP grade was based on the two factorial grades and, according to the test instructions, the grade for fluency should in doubtful cases be weighted more heavily than the grade for language structures and vocabulary. Class 3 had the highest mean grade for fluency (4.1), followed by class 4 (3.4), class 1 (3.2), and class 2 (2.8). The differences between the classes were significant ($p=.000$). Girls had a higher mean grade for fluency (3.5) than boys (3.3) but this difference was not statistically significant ($p=.248$).

7.3.5 Pauses and fluency

As was mentioned in chapter 6, two aspects of oral proficiency (fluency and use of advanced vocabulary) are investigated in more detail in the present work. In the factorial grade for fluency, a qualitative assessment was quantified into a grade (1-6), which is one method of measuring fluency. As was discussed in chapter 4, fluency is an elusive concept, notoriously difficult to define and problematic to measure objectively. Several methods of measuring fluency quantitatively were discussed in chapter 4 and in chapter 6 one of them, mean intra-utterance pause length beyond one second, was presented as the preferred method in the present study. A combination of both qualitative and quantitative

measures of fluency is one attempt at methodological triangulation of data in the present work, to ascertain reliable results on fluency.

Thus, the measure adopted here is pause length, more specifically mean intra-utterance pause length beyond one second. Based on the OP grade on all five tests, ten students were selected, namely the five students with the five lowest mean OP grades in the study (Low OP group) and the five students with the highest mean OP grades (High OP group). In the Low OP group there were three boys and two girls, four from class 2 and one from class 1. In the High OP group all were girls, three from class 3 and one each from classes 1 and 4. Since transcription and pause measures are very time-consuming, the analysis was limited to speech produced in one test only. I chose Test 5 since it was the mandatory national test for the students.

As is shown in Table 7.9, there were great differences between the two groups with respect to the number of pauses made and words used.¹² However, since the length of the tests varied in terms of time, it was necessary to consider the means in terms of pauses per uttered words. Students in the High OP group apparently had fewer pauses per uttered words than students in the Low OP group. At the same time, students in the High OP group produced many words and scored well on fluency. Fillmore (1979:93) says that we find such fluency with people who are rarely tongue-tied and who seem at ease irrespective of the setting.

Table 7.9. Pauses and words for the Low and High OP group.

Student ID	Low OP group			High OP group			Student ID
	Pauses	Words	Pauses/ Words	Pauses	Words	Pauses/ Words	
Low 1	38	329	.116	0	503	.000	High 1
Low 2	20	180	.111	6	702	.009	High 2
Low 3	24	210	.114	8	665	.012	High 3
Low 4	25	303	.083	9	852	.011	High 4
Low 5	8	172	.047	6	928	.006	High 5
Sum	115	1,194	.096	29	3,650	.008	Sum

A scattergram of the mean OP grade versus mean intra-utterance pause length further illustrates the difference between the two groups (see Fig. 7.20). The difference in mean pause length between them, 2.97 s for the Low OP group versus 1.27 s for the High OP group, was statistically significant ($p=.001$) and there was a negative correlation ($r=-.902^{**}$) between pause length and OP

¹² Words equal English tokens (see section 6.3.4).

grade. That is, the longer the pause, the lower the grade, and vice versa. The results for fluency confirm previous studies (see e.g. Iwashita et al., 2008:41).

The total amount of extramural English for the students in the Low and High OP group respectively is shown in Table 7.10. When a comparison was made between the pause length for the groups and their EE, an interesting pattern appeared (see Fig. 7.21). The Low OP group had a negative correlation between pause length and EE. That is, with increased time spent on EE, this group seemed to have improved fluency because the mean pause length became shorter. In contrast, the mean intra-utterance pause length increased with more time spent on EE for the High OP group which means that, at least when using this measure of fluency, students' fluency in the High OP group in fact became worse with increased EE. These findings are both bewildering and interesting. However, since the number of students is so small in each group, no conclusions or generalizations of the correlations described above between fluency and EE can be drawn or made. Nevertheless, the findings here suggest that extramural activities in a second language might very well matter also in analyses of language at a micro level. Moreover, the positive correlation between extramural English and oral fluency in the Low OP group neatly reflects another finding in the present study, namely a positive correlation between EE and OP among less proficient learners (cf. section 7.4.2).

Table 7.10. Time spent on EE for the Low and High OP group.

Student-ID	EE (hours/week)		Student-ID
	Low OP group	High OP group	
Low 1	.0	5.2	High 1
Low 2	4.9	13.0	High 2
Low 3	10.8	20.9	High 3
Low 4	11.9	25.8	High 4
Low 5	21.2	13.9	High 5
Sum	48.8	78.8	Sum
Mean	9.8	15.8	Mean

7.3.6 Understanding the data from a different perspective

As was mentioned in the introduction to section 7.3, this last subsection aims to view the data from a different perspective. So far, different variables have been seen through lenses defined by students' reported times spent on extramural English. It is, of course, also possible to study the data in other ways. An obvious way of examining my data is to do it from the perspective of students'

oral proficiency, i.e., to cluster the students depending on their level of oral proficiency. What results will turn up when such an approach is adopted?

In this section, the sample is again divided into five subsets, but this time the subsets are based on the students' mean OP grades with the aim of having subsets that mirror the distribution of the OP grade for the whole sample (cf. Fig.7.18). That is, the rationale behind the clustering is similar to the one used in creating the EE subsets, except for the fact that here the subsets are based on students' OP grades instead of on their time spent on extramural English. The OP grade was normally distributed and the five subsets are, therefore, more symmetrical than the ones created based on the EE variable (see Fig. 7.22).

Descriptive statistics for the five subsets based on oral proficiency (henceforth OP subsets) are found in Table 7.11. Note that each of the two extreme OP subsets (subsets 1 and 5) consists of five students and they are, thus, equivalent to the two groups in the analysis in section 7.3.5. The intermediate subset is the biggest one, as is always the case when data are normally distributed. The differences in OP grades between the five subsets were statistically significant ($p=.000$). The gender distribution in the five OP subsets is illustrated in Figure 7.22, which confirms what has already been mentioned, namely that girls in my sample were graded higher on OP than boys.

Table 7.11. Descriptive statistics for the five OP subsets.

OP subset	Level of OP	OP grade interval		Mean OP grade	N	STD
		From	To			
1	Very low	≥ 1.6	≤ 2.0	1.8	5	.15
2	Low	> 2.0	≤ 2.7	2.4	16	.24
3	Intermediate	> 2.7	≤ 3.6	3.3	27	.27
4	High	> 3.6	≤ 4.6	4.1	21	.34
5	Very high	> 4.6	≤ 5.3	5.0	5	.26
Total		≥ 1.6	≤ 5.3	3.3	74	.88

7.4 A comparison of students' extramural English and oral proficiency

In order to answer the main research question, whether extramural English has an impact on students' oral proficiency, it is necessary to correlate results on students' EE and OP. This section focuses on such correlations and, as was discussed in chapter 6, with the help of statistics it is possible to point to relationships between variables, but not really to make any claims regarding the

cause and effect of matters. Such claims belong to the interpretation phase, which, in the present work, is in the final chapter.

The section begins with correlation analyses of the variable for extramural English (EE) on the one hand and the variable for oral proficiency (OP grade) on the other, on sample and class level (7.4.1). Correlations are also made for gender. In section 7.4.2, EE is the independent variable and the OP grade is the dependent one and as before, results are presented for the four classes as well as gender, and also for the two methods of clustering the sample: five EE subsets and five OP subsets. In section 7.4.3, results are presented in the opposite way, that is, with the OP grade as the independent variable and EE as the dependent one. Results on the rank of importance for different activities of extramural English on oral proficiency are presented in section 7.4.4.

7.4.1 Correlations between extramural English and oral proficiency

Based on data for the whole sample, a correlation analysis of students' total amount of EE and their mean OP grade on all five speaking tests shows that there was a positive and statistically significant correlation between extramural English and oral proficiency ($r_s=.307^{**}$). This means that almost 10% (i.e. $.307^2$) of the variance in OP grades is explained by the relationship between EE and oral proficiency rather than by chance or some other factor(s) (see section 6.3.6.5). To illustrate the correlation between EE and OP, Figure 7.23 is provided, where the fit line for the total is an illustration of the positive correlation that was found. The fact that the correlation was significant shows that the relationship between EE and OP was not due to chance. As was previously mentioned (see section 6.3.6.5), this quantitative method of analysis cannot make any claims with regard to causality or the direction of causality between two variables, but it can point to relationships between variables and the two variables currently investigated were found to correlate positively.

Interestingly, additional correlation analyses at class level revealed great differences. Figure 7.23 provides a visual description of the dispersion of the students and the differences between the four school classes, as shown by the added fitlines for each class. Note that there were positive correlations between EE and OP for classes 1 and 2 but negative correlations for classes 3 and 4 and with varying slopes. At first sight all this is bewildering but, in a sample such as the one in the present study, the result is not surprising. There might be, for example, socioeconomic factors of importance which might explain the differences, or there might be a teacher effect. It is also possible that there are gender differences which affect the whole sample. Therefore, I also correlated

EE and OP for boys and girls, both on sample level (Fig. 7.24) and on class level, to see whether there were any similarities or differences between the genders, and to see if the explanation for the differences between the classes might lie there.

For boys, there was a positive and statistically significant correlation between EE and OP ($r_s=.515^{**}$). In short, there was a relationship between the amount of EE boys had and their assessed level of OP, which can be described in the following way: the more time the boys had reported for EE, the higher their OP grade was, and vice versa. For girls there was also a positive correlation between EE and OP, but it was less positive than for boys and not statistically significant ($r_s=.118$). Basically, it seems as if it did not matter whether girls had a low or high amount of EE because the correlation with their assessed level of OP was negligible. I deliberately write “seems”, because the fact that the result for girls lacked statistical significance makes it more difficult to interpret than the result for boys. Nevertheless, the conclusion is that boys seem to be much more sensitive to EE than girls. Gender is, thus, one factor that explains why the results for the classes varied the way they did.

7.4.2 Oral proficiency viewed through the EE subsets

Viewed through the EE subsets, it is possible to study how the five OP subsets disperse differently (see Figure 7.25). In Figure 7.25, the stacks of each EE bar are defined by the different OP subsets. The fact that OP is the dependent variable does not mean that there is a cause and effect relationship at play in real life; it is only a matter of expressing or illustrating relationships. One thing that Figure 7.25 reveals is that the students who had a very high level of OP in this sample (i.e. OP subset 5) had intermediate amounts of EE. That is, they were found in EE subsets 2 and 3. In addition, Figure 7.25 reveals that students with a very low level of OP belonged to EE subsets 1, 2, and 3. Generally speaking, the results show that there were no students who had the combination of low amounts of EE and a high level of OP. Neither were there students who had a very large amount of EE (EE subset 5) and a very high level of OP (OP subset 5). Finally, few students had large amounts of EE and low levels of OP.

It was not reasonable to assume that a significant correlation could be found between EE and OP within a particular EE subset, because informants within a specific subset had a very similar amount of EE. However, for those with very little extramural English, which was the case for students in EE subset 1, any differences in time spent on EE activities, however small they

might be, could be relevant because, relatively speaking, there is a greater difference between zero and eight hours of EE activities (i.e. the limits for subset 1) than between, for example, 45 and 53 hours (eight hours well within the limits of EE subset 5). Thus, a correlation analysis of the EE and OP grade variables was carried out for EE subset 1. It revealed that there was a positive and significant correlation ($r_s=.590^*$) between time spent on EE and results on OP within this particular subset (see Fig. 7.26). In other words, for those who had very low amounts of extramural English, the EE they did come in contact with seems to have been of great importance. Even though perhaps five or eight hours of EE per week was very little in comparison to what others might have had, these hours appear to have been crucial for individual students in EE subset 1.

7.4.3 EE viewed through the subsets based on oral proficiency

In this section, extramural English will be viewed through the five subsets that are based on oral proficiency (the OP subsets), i.e. the opposite of what was done in the previous section. In Figure 7.27, the stacks of each OP bar are defined by the different EE subsets. Figure 7.27 confirms conclusions drawn in section 7.4.2 and, moreover, is another way of understanding the present data. A close look at Figure 7.27 reveals that EE subset 5, i.e. the students with the highest amount of EE, had intermediate or high levels of oral proficiency (i.e. OP subsets 3 and 4). EE subset 4 was dispersed into OP subsets 2, 3, and 4. EE subset 3 was dispersed into all five OP subsets, and so was EE subset 2. Finally, the majority of the students in EE subset 1, i.e., the students with very little EE, were found in OP subsets 1 and 2. However, note that three students from EE subset 1 were found at intermediate OP level (i.e. in OP subset 3) and one in OP subset 4. This way of viewing the present data confirms the previous finding, namely that students with low amounts of EE and a high level of OP were unlikely to be found, and students with high amounts of EE and a low level of OP were also unlikely to be found. Furthermore, there were no students who belonged to both EE subset 5 and OP subset 5, i.e. the highest end of the EE and OP scale, respectively. However, at the lowest end of these scales, students who belonged both to EE subset 1 and OP subset 1 could be found.

7.4.4 Rank of importance: EE activities in relation to oral proficiency

One of my research questions has to do with relative importance. That is, are some EE activities more important than others for oral proficiency? With the

help of backward linear regression analysis (see section 6.3.6.6), it is possible to examine the relative importance of various measured EE activities. Hence, it is possible to calculate the rank of importance for EE activities in relation to the OP grade. In my regression analysis, I decided to include seven of the EE activities. I excluded the last open category that was listed in the language diary, because the information given by the students there was too disparate to be considered a single EE activity suitable for inclusion in the regression analysis together with the other seven EE activities. Results on the regression analysis are shown in Table 7.12.

As can be seen in Table 7.12, the two EE reading activities were ranked as having the highest relative importance for oral proficiency. Ironically, these two activities were at the same time the ones that students on average spent the least time on (cf. Table 7.1). Listening to music, which was the EE activity that students spent most time on, was ranked in third place. The least important EE activity was watching films.

Table 7.12. Rank of importance for EE activities in relation to the OP grade.

Rank	EE activity
1	Reading newspapers/magazines
2	Reading books
3	Listening to music
4	Watching TV
5	Surfing the Internet
6	Playing video games
7	Watching films

7.5 Extramural English and oral proficiency: Summary and conclusion

This chapter presented results on students' EE, based on language diary data and data from the questionnaire. Overall, "listening to music" was found to be the EE activity that students spent the most time on. Music was followed by "playing video games", "watching TV", "watching films", "surfing the Internet", "other activity", "reading books", and finally, "reading newspapers/magazines". Significant differences were found between the amount of time boys and girls spent on playing video games and surfing the Internet. Boys spent on average more time than girls on both of these EE activities. For video games, boys reported 7.9 hours per week, whereas girls reported much less (.7 hrs/w). With regard to the Internet, boys reported 1.1 hours per week and girls .4 hours per week. Another way of illustrating the gender difference is to study the proportion of the total EE time the two EE

activities accounted for, for each gender. For the boys, video games and the Internet accounted for approximately 44% of their total EE time (20.8 hrs/w), which can be compared with 6% of the girls' total EE time (16.4 hrs/w) (see Figures 7.10 and 7.11).

On the whole, students' responses to EE-related questions in the questionnaire agreed very well with the results found on EE based on language diary data. This is a finding which strengthens the reliability of my results.

Results on students' oral proficiency, based on assessment data from five speaking tests, were also presented. The mean OP grade (on a 1-6 scale) for the whole sample was 3.3. Girls had a higher mean OP grade (3.5) than boys (3.2), but the difference was not statistically significant. The time span from Test 1 to Test 5 was 8-9 months. Paired samples *t* tests confirmed that all four participating classes improved over time.

A study on pauses and oral fluency based on data from ten of the students in the sample was also accounted for. These were the five students with the lowest OP grades in the sample (Low OP group) and the five students with the highest OP grades (High OP group). Mean intra-utterance pause length was measured for both groups and a significant difference was found ($p=.001$) between the groups. In addition, there was a negative and statistically significant correlation ($r=-.902^{**}$) between pause length and OP grade, which means that the longer the pause length, the lower was the OP grade.

In order to answer my main research question, results on EE and OP were correlated. A positive and significant correlation was found between students' total time spent on EE and their OP on sample level ($r_s=.307^{**}$). Similar analyses on class level revealed very varying and bewildering results. An explanation for these results on class level was found when gender was examined in relation to EE and OP. Correlation analysis showed that there was a positive and statistically significant correlation between EE and OP for boys ($r_s=.515^{**}$). For girls there was also a positive correlation, but less so than for boys and it was not statistically significant ($r_s=.118$). Thus, boys appeared to be more sensitive to EE than girls.

The sample was clustered into five so-called EE subsets, where EE subset 1 had the lowest amount of time spent on EE and EE subset 5 the highest. ANOVAs revealed that the differences in EE between the five subsets were significant ($p=.003$). Moreover, EE subset 1 clearly had the lowest mean OP grade (2.5). The other four subsets did not differ much in terms of their oral proficiency, but the differences that were found were statistically significant: EE subset 2 (3.5), EE subset 3 (3.5), EE subset 4 (3.3), and EE subset 5 (3.6). The

results indicate that there was no straightforward relationship between EE and OP in the sense that the OP grade improved with each EE subset. It is noteworthy that for students in EE subset 1, a positive and statistically significant correlation was found between EE and OP ($r=.590^*$). This finding implies that for such learners, even slight increases in the amount of time spent on EE activities might be important.

Finally, with the help of backward linear regression analysis, a ranking list with the relative importance of different EE activities for OP was established. The two EE activities that involved reading ended up on top, followed by listening to music, watching TV, surfing the Internet, playing video games, and watching films.

In the next chapter, results for students' vocabulary are presented. Among other things, results on correlation analyses between EE and vocabulary are included, as well as an investigation of the relative importance of different EE activities for vocabulary.

8 Extramural English and vocabulary: Analysis and results

In this chapter I present results on students' extramural English and vocabulary. The chapter begins with the results on two written vocabulary tests that are based on word frequency rankings: the Productive Levels Test (8.1), which measures productive vocabulary, and the Vocabulary Levels Test (8.2), which measures receptive vocabulary. For each of these tests, results are presented for the four participating classes and gender, followed by results for the five EE subsets. In section 8.3, results on a cloze test (Working for Change) are presented. The cloze test is very short, and hence does not carry the same weight as the other two vocabulary tests in the present study. I will, therefore, not compare the results for classes and gender on the cloze test; instead I will focus on results that relate to extramural English, that is, how the results on the cloze test relate to the five EE subsets. In section 8.4, results are presented regarding the factorial grade for vocabulary, based on assessment data from the five speaking tests. In the following section (8.5), results on correlations between extramural English and vocabulary are discussed. The rank of importance for extramural English activities in relation to vocabulary is treated in section 8.6. A study on polysyllabic vocabulary for ten of the students in the sample, based on speech data from Test 5, is accounted for in section 8.7. There is also a brief discussion of correlations between polysyllabic vocabulary and extramural English at the end of that section. The chapter ends with a summary and conclusion (8.8).

8.1 Productive Levels Test

The results presented in this section are based on data from the Productive Levels Test, which was the written vocabulary test that the students took during the fall semester. As was described in chapter 6, the Productive Levels Test measures productive vocabulary, and the target words at the beginning of the test were easier (more frequent) than those at the end (less frequent).

8.1.1 Results for the four participating classes and gender

Mean scores on the Productive Levels Test for the four participating classes and gender are shown in Table 8.1. The scores were normally distributed. Class 3 had the highest mean score in the sample whereas class 4 had the lowest. Note, however, that the differences between the classes were not significant on this test ($p=.091$). A comparison of scores was also made between boys and girls. Boys scored better than girls, but the difference was not statistically

significant. The boys in classes 1 and 3 did particularly well on the Productive Levels Test. In contrast, the girls in class 1 clearly lagged behind all other groups. Even though class 1 only had twenty students in total, the discrepancy in the scores between boys and girls in that class was almost statistically significant ($p=.052$). It was previously shown in chapter 7 (Table 7.3) that there was a significant difference in EE between boys and girls in class 1. If EE is connected with vocabulary, it might be one possible explanation why the PLT scores for boys and girls in class 1 differ.

Table 8.1. Mean scores for the Productive Levels Test for the four participating classes; significance values for gender differences.

Class	Gender	Productive Levels Test (max 45)	STD	N	Sig. (2-tailed)
1	Boys	19.3	9.7	10	.052
	Girls	11.9	5.7	10	
	Total	15.6	8.6	20	
2	Boys	14.0	9.1	10	.829
	Girls	14.9	7.3	8	
	Total	14.4	8.2	18	
3	Boys	21.9	6.7	9	.141
	Girls	17.8	5.7	13	
	Total	19.5	6.3	22	
4	Boys	14.9	8.3	7	.799
	Girls	14.0	6.4	13	
	Total	14.3	6.9	20	
Total	Boys	17.6	8.9	36	.106
	Girls	14.8	6.4	44	
	Total	16.1	7.7	80	

8.1.2 Results for the five EE subsets

Mean scores for the Productive Levels Test for the five EE subsets are shown in Table 8.2. The highest mean scores were found in EE subsets 4 and 5. It needs to be pointed out, though, that there was a rather large individual variation between the students in these subsets despite the small size of them. Nevertheless, the differences between all five subsets were statistically significant ($p=.016$). The results on oral proficiency for the five EE subsets accounted for in chapter 7 (see Table 7.8) showed that the OP grade did not improve with each EE subset. When it comes to vocabulary, on the other hand, the situation was different: the mean scores on the Productive Levels Test were

either the same or improved with each subset.¹ This finding could be an indication that there is a more straightforward relationship between EE and vocabulary than between EE and oral proficiency.

Table 8.2. Mean scores for the EE subsets on the Productive Levels Test.

EE subset	Productive Levels Test (max 45)	N	STD
1	10.3	15	4.2
2	17.0	34	7.3
3	17.0	21	7.4
4	18.3	6	10.4
5	21.5	4	9.9
Total	16.1	80	7.7

8.2 Vocabulary Levels Test

The results presented below are based on data from the Vocabulary Levels Test, the written vocabulary test that the students took in the spring semester. The Vocabulary Levels Test measures receptive vocabulary with an increase in difficulty level; i.e., test items become more peripheral with each section of the test. I first give an account of the results for the four participating classes and gender. After that, I present mean scores for the five EE subsets. An analysis follows where extramural habits among students with very high and very low scores are investigated.

8.2.1 Results for the four participating classes and gender

Mean scores for the Vocabulary Levels Test for the four participating classes and gender are shown in Table 8.3. As was the case for the Productive Levels Test (see section 8.1.1), the results were normally distributed. Contrary to what was found for the Productive Levels Test, there were significant differences between the scores for the four school classes on this test ($p=.049$), where class 3 had the highest mean score and class 2 the lowest. The overall result is similar to what was found for the EE subsets on oral proficiency with one exception. On the Vocabulary Levels Test, class 1 performed better than class 4, whereas in oral proficiency, class 4 performed better than class 1 (cf. Table 7.6).

¹ EE subsets 2 and 3 had the same mean score.

Table 8.3. Mean scores for the Vocabulary Levels Test for the four participating classes; significance values for gender differences.

Class	Gender	Vocabulary Levels Test (max 90)	STD	N	Sig. (2-tailed)
1	Boys	69.0	13.3	10	.019*
	Girls	53.6	12.7	9	
	Total	61.7	14.9	19	
2	Boys	56.9	14.3	10	.637
	Girls	53.5	15.6	8	
	Total	55.4	14.6	18	
3	Boys	71.7	8.0	9	.049
	Girls	61.8	12.5	13	
	Total	65.8	11.7	22	
4	Boys	57.3	11.6	7	.774
	Girls	55.6	12.7	12	
	Total	56.2	12.0	19	
Total	Boys	64.3	13.5	36	.017*
	Girls	56.7	13.2	42	
	Total	60.1	13.7	78	

*Sig. at the .05 level

In contrast to what was the case for the Productive Levels Test, which measured productive vocabulary, there was indeed a significant difference between boys and girls on the Vocabulary Levels Test ($p=.017$). Again, boys scored higher than girls (see Table 8.3). The test results thus indicate that boys had acquired a larger passive vocabulary than girls. Similar to what was found for the Productive Levels Test, boys in classes 1 and 3 outperformed the rest of the students in the sample.

8.2.2 Results for the five EE subsets

As was mentioned in chapter 1, one of my aims was to map out students' EE. In doing so, results for the five EE subsets on various measures, such as on productive vocabulary (reported above, section 8.1.2) and on receptive vocabulary (reported in the present section) are important. The mean scores for the Vocabulary Levels Test for the five EE subsets are provided in Table 8.4. The differences between the EE subsets were statistically significant for the total scores ($p=.004$) and for two parts of the test: Level 2,000 ($p=.000$) and Level 3,000 ($p=.001$). On the third and most difficult part of the test, Level 5,000, the differences between the EE subsets were not statistically significant ($p=.091$), which is probably due to the fact that the words tested at the 5,000 Level were too advanced for a majority of the students and, therefore, the

probability of finding significant differences between all the groups decreased. It is still noteworthy that the mean score for students in EE subset 5 on Level 5,000 was much higher than the scores for the other subsets. The expected declining scores on sample level with each part of the test are also noteworthy, from 23.5 on the first (and easiest) part of the test, to 20.5 on the second part, ending up at 16.2 on the third (and most difficult) part of the test.

Table 8.4. Mean scores for the EE subsets on the Vocabulary Levels Test.

EE subset	N	Vocabulary Levels Test				STD
		Level 2,000 (max 30)	Level 3,000 (max 30)	Level 5,000 (max 30)	Total (max 90)	
1	14	18.6	16.2	14.0	48.8	9.7
2	34	24.4	21.0	16.4	61.7	12.9
3	20	24.8	21.6	15.8	61.6	13.3
4	6	23.5	22.0	17.5	63.0	16.4
5	4	26.5	24.0	23.0	73.5	11.9
Total	78	23.5	20.5	16.2	60.1	13.7

As was pointed out in section 8.1.2, the mean score on the Productive Levels Test was either the same or improved with each subset, something which was interpreted as a potentially straightforward relationship between EE and vocabulary. The results on the Vocabulary Levels Test confirm such an interpretation: the difference between EE subsets 2 and 3 is minimal and together with the rest of the scores, the results on the Vocabulary Levels Test also suggest a fairly straightforward connection between EE and vocabulary.

8.2.3 EE habits among students with extreme scores

I examined the extramural English habits among students with extreme scores on the Vocabulary Levels Test, i.e., those with the highest and lowest scores. Eight students were included in each of the groups. The maximum score on the test was 90 points.

The top eight scores on the Vocabulary Levels Test ranged from 78 to 88 points; six of those were achieved by boys and two by girls. Three of the students were from class 1 and they were all boys. This explains why the boys in class 1 stood out as a group with the highest mean score in both vocabulary tests. There was no student from class 4 among those who had the eight highest scores. But what type of EE activities were the eight high-scoring students involved in?

The top eight students spent an average of almost 29 hours per week on EE activities, but the individual variation was large (STD 16), ranging from 13

to 57 hours. None of them reported any time on reading books or newspapers/magazines in English.² Most time was instead reported for playing video games (12 hrs/w). They also listened to music quite extensively (6 hrs/w) and watched TV almost as much. In comparison to the times reported for the various EE activities on sample level (see Table 7.1), these eight students spent much more time on the last one of those, namely “other activity”. The top eight students reported almost two hours per week for this specific category of EE, which can be compared with approximately 20 minutes per week on average for the whole sample. Other EE activities that these eight students had listed were, for instance, speaking to foreigners in English. One student had talked to Brits via Skype while playing video games on several occasions during the measured period of two weeks. Two other students had apparently met Germans in their spare time, because both had listed “speaking English to a German” in their respective language diaries. In the diaries, there were also comments about chatting in English and sending questions or comments to various support sites on the Internet. One student had been to a musical (*Jesus Christ Superstar*) which was in English. The fact that these eight students in comparison with the sample mean reported spending much more time on other activities, may be part of the explanation as to why they performed well on vocabulary, because doing other EE activities than the seven that were listed in the language diary most likely demanded a genuine interest in English and probably a certain level of general proficiency as well. In any case, no other factor could be identified which might explain why these students scored so well.

In comparison with the top eight students, the eight students at the other extreme end reported much less time spent on EE: 13 hours per week on average. Again, the individual variation was large (STD 13) with a range from zero reported hours to 41. For the eight students in the group with a low amount of EE, most time was reported for the EE activity “listening to music” (6 hrs/w), followed by “watching films” (3 hrs/w) and “watching TV” (3 hrs/w). For the remaining EE activities, figures below one hour per week were reported.³ Of the eight students with low scores on the vocabulary test, five were girls and three boys. They came from all four classes: four from class 2,

² However, in the reported activities in Swedish, which were also available in the language diaries, the eight students who had the highest scores had above average figures for reading, both for reading books and reading newspapers/magazines.

³ For activities in Swedish reported in the language diaries, the eight students in the low extreme group had below average figures for reading, both for reading books and for reading newspapers/magazines.

two from class 4, and one each from classes 1 and 3. Their scores on the Vocabulary Levels Test ranged from 33 to 38.

8.3 Cloze test: “Working for Change”

As was mentioned in chapter 6, the cloze test “Working for Change” was part of the Swedish national test of English in 2007 and I chose to include it in my investigation of student vocabulary, even though cloze tests are generally considered to measure more than just vocabulary (D. H. Brown, 2004:9). The present cloze test was taken during the spring semester, before the Vocabulary Levels Test (see Table 6.1). As I mentioned in the introduction to the present chapter, because the cloze test was short and not considered as important as the Productive Levels Test and the Vocabulary Levels Test, in this section I focus on the results that relate directly to the main topic of the thesis, i.e., the five EE subsets.

The mean scores for the EE subsets on the cloze test are shown in Table 8.5. Similar to what was found for the Vocabulary Levels Test, the differences between the five EE subsets were statistically significant ($p=.040$) also for “Working for Change”. Again, EE subset 5 had the highest mean score on the test, which was the case also for the Productive Levels Test and the Vocabulary Levels Test. However, this time, there was no straightforward relationship between the scores on the test and the EE subsets, i.e., the mean score on the cloze test was not the same, nor did it improve, with each EE subset. For instance, EE subset 2 had a higher mean score than both subsets 3 and 4. Instead, the picture that emerged was very similar to the one that emerged for EE with regard to oral proficiency (cf. section 7.3.3). The reason for this finding is probably that a cloze test such as “Working for Change” indeed measures more than just vocabulary. Therefore, the results in the present section partly deviate from other results on EE and vocabulary in this study and instead resemble results on EE and oral proficiency.

Table 8.5. Mean scores for the EE subsets on the Working for Change Test.

EE subset	Working for Change (max 12)	N	STD
1	7.4	14	1.9
2	9.6	34	2.5
3	9.1	20	2.0
4	8.2	6	2.5
5	10.0	4	3.4
Total	9.0	78	2.4

8.4 The factorial grade for vocabulary

In this section, focus is on the factorial grade for *language structures and vocabulary* (henceforth vocabulary) that the external raters awarded the students (cf. section 6.3.3). Similar to what was found for the factorial grade for fluency (section 7.3.4), the factorial grade for vocabulary also had a normal distribution in the sample. Since the OP grade was based on two factorial grades, the results for the factorial grade for vocabulary for the four classes resembled those found for the OP grade (see Table 7.6). Class 3 had the highest mean grade for vocabulary (3.9), followed by class 4 (3.2), class 1 (3.1), and class 2 (2.6). The order of achievement for the classes is identical to the one found also for the factorial grade for fluency. There seems to be a pattern, because with regard to vocabulary as well, the differences between the classes were significant ($p=.000$). Likewise, again girls had a higher mean grade than boys: 3.3 as opposed to 3.1. It should be noted that this difference was not statistically significant ($p=.285$), i.e. it might be due to chance.

I examined the factorial grade for fluency for the five EE subsets. There were significant differences between the subsets ($p=.001$). As can be seen in Table 8.6, the pattern is similar to the one that was found between the EE subsets and the OP grade (cf. Table 7.8). That is, EE subset 1 had the lowest mean grade for vocabulary and EE subset 5 the highest. Subsets 2 and 3 had the same mean grade for vocabulary and subset 4 a slightly lower result.

Table 8.6. EE subsets and the factorial grade for vocabulary.

EE subset	Factorial grade for vocabulary	N	STD
1	2.4	13	.6
2	3.4	32	.8
3	3.4	20	.8
4	3.2	5	.7
5	3.7	4	.5
Total	3.2	74	.9

8.5 Correlations between extramural English and vocabulary

As part of my main aim was to see whether EE has an impact on vocabulary, it was essential to carry out correlation analyses between EE and vocabulary. I used data from the Productive Levels Test and the Vocabulary Levels Test to

correlate each test with the total amount of EE, i.e., the EE variable. Interestingly, there was a positive and statistically significant correlation between the Productive Levels Test and the EE variable ($r_s=.352^{**}$; $N=80$). The result was very similar when EE was correlated with the Vocabulary Levels Test ($r_s=.354^{**}$; $N=78$). The results thus show that time spent on EE is clearly connected with students' vocabulary scores, which is an important finding.

Correlation analyses at class level revealed results similar to those for EE and OP (cf. Fig. 7.24). That is, positive correlations between EE and vocabulary for classes 1 and 2, but negative ones for classes 3 and 4. However, the bewildering results on class level could yet again be explained by taking gender into account. The results presented in this section agree to a large degree with those in the previous chapter on EE and oral proficiency. For boys, the correlation between the EE variable and the Productive Levels Test was positive and statistically significant ($r_s=.575^{**}$, $N=36$). The result was very similar in correlations regarding boys' EE and the Vocabulary Levels Test ($r_s=.594^{**}$, $N=36$). In contrast, for girls there was no apparent correlation between their EE and the results on vocabulary, neither for the Productive Levels Test nor for the Vocabulary Levels Test (Productive Levels Test: $r_s=.037$, $N=44$; Vocabulary Levels Test: $r_s=.060$, $N=42$). Note that the correlations between EE and vocabulary for girls were non-significant, something which makes the girls' results more difficult to interpret than those of the boys. In any case, what I found here on gender in relation to EE and vocabulary to a large extent confirmed what I previously found on EE and oral proficiency: It seems as if boys are more sensitive to EE than girls, but it might also have to do with the fact that boys and girls are engaged in different EE activities. Why the correlations for boys and girls differ the way they do is further discussed in chapter 10.

8.5.1 An index variable for vocabulary

It was possible to create an index variable for vocabulary, the Vocabulary Index Variable (see section 6.3.6.7), based on data from the Productive Levels Test and the Vocabulary Levels Test. Results from the Productive Levels Test and the Vocabulary Levels Test correlated strongly ($r=.843^{**}$) and reliability was within the optimal range (Cronbach's alpha .832).

The relationship between vocabulary and EE was established already in the previous section, and a correlation analysis of the Vocabulary Index Variable and the EE variable confirmed the conclusions drawn above. The Vocabulary Index Variable correlated positively with the EE variable

($r_s=.357^{**}$) and the difference between genders was also confirmed (Boys: $r_s=.590^{**}$; Girls: $r_s=.011$). This means that 35% of the variation in the boys' vocabulary scores was accounted for by EE, rather than by chance or some other factor(s). In contrast, no such relationship between EE and vocabulary could be seen for girls.

The Vocabulary Index Variable is suitable to use when investigating the rank of importance for extramural English activities in relation to vocabulary in general. Such an investigation is important and highly relevant in the present study, since it answers directly to one of my research questions: “Does the amount of time of EE seem to be important? *Or is it rather the type of EE that seems to matter?* Or both?” (cf. section 1.2). I account for the results in the following section.

8.6 Rank of importance: EE activities in relation to vocabulary

In chapter 7, with the help of backward linear regression analysis, the two reading activities (“reading books”; “reading newspapers and/or magazines”) were found to be of highest relative importance for the results on oral proficiency. They were followed by “listening to music” (see Table 7.12). In the present section, the same type of regression analysis was used to examine the rank of importance for extramural English activities in relation to vocabulary (the Vocabulary Index Variable). The outcome of the relative importance of the EE activities was *not* the same this time. For vocabulary, “playing video games” and “surfing the Internet” were instead the most important EE activities, followed by “watching TV” and the two reading activities (see Table 8.7).⁴ It is also worth pointing out that the EE activity “watching films” appeared to be relatively unimportant both for OP and for vocabulary.

Table 8.7. Rank of importance for EE activities in relation to the Vocabulary Index Variable.

Rank	EE activity
1	Playing video games
	Surfing the Internet
3	Watching TV
4	Reading newspapers/magazines
5	Reading books
6	Listening to music
7	Watching films

⁴ Due to its disparate and heterogeneous nature, the EE activity “other activity” was excluded from this type of analysis (cf. section 7.4.4). This does not mean that other EE activities are unimportant for second language acquisition (see section 8.2.3).

It was shown in section 7.1.1 that boys spent significantly more time than girls both on playing video games and surfing the Internet in English, i.e., they spent more time than girls on the two most important EE activities with regard to vocabulary. In sections 8.1 and 8.2, results revealed that the boys' scores were higher than the girls' on both the Productive Levels Test and on the Vocabulary Levels Test (significantly higher on the latter of the two tests). Moreover, in the section above (8.5.1), the correlation between vocabulary and EE was positive and significant for boys but negligible and not significant for girls. This means that the boys' habit of spending more time on playing video games and surfing the Internet is a possible explanation as to why boys outperformed girls on vocabulary, at least in the present study. In hindsight, boys' and girls' varying results on EE, oral proficiency, and vocabulary are not very surprising.

8.7 Polysyllabic vocabulary in the High and Low OP groups in Test 5

The proportion of non-Germanic vocabulary increases in peripheral vocabulary, as does the proportion of polysyllabic words (see chapter 5). Thus, actual production of polysyllabic words may indicate that a learner has acquired an advanced vocabulary. In order to study students' use of advanced vocabulary, polysyllabic vocabulary was investigated for ten students in the sample, namely the students with the highest and lowest OP grades, i.e. the same students that were used when studying pauses and fluency (see section 7.3.5; the High OP group and the Low OP group). I decided to limit the current investigation of vocabulary to one speaking test only (Test 5, cf. section 6.3.3). As I said above, the actual production of polysyllabic words may indicate that a speaker has acquired an advanced vocabulary. It should be noted that spoken language per se contains fewer advanced lexical items than written language (see section 5.1), so instances of students' use of polysyllabic vocabulary in speech are interesting from the point of view of second language vocabulary acquisition.

The total number of *tokens* (here defined as 'all English words used by a speaker'), *core tokens* (all tokens minus polysyllabic tokens), *polysyllabic tokens*, *polysyllabic types*, and the students' "*own*" *polysyllabic types* (the polysyllabic types that were not included in and elicited by the test material) are shown in Table 8.8 for the Low OP group and in Table 8.9 for the High OP group. Note that *core tokens*, equivalent to 'all tokens minus polysyllabic tokens' according to how I defined the term in my investigation, may include lexical items beyond the

1,000 level, i.e. beyond the core as described by Minkova and Stockwell (2006). It is also possible that some of the *polysyllabic tokens* (tokens with three syllables or more) in my investigation may belong to the most frequent one thousand words in English (i.e. the core) rather than to peripheral vocabulary beyond the core. My division of the lexicon at polysyllabicity – ‘three syllables or more’ – was hypothesized to yield reliable results on students’ vocabulary (see section 6.3.4), despite the fact that there may indeed be some lexical items in the investigation that from a strict word frequency perspective were not signs of advanced vocabulary use.

In Tables 8.8 and 8.9, four types of word ratios are also provided: (1) polysyllabic tokens per all tokens; (2) polysyllabic types per polysyllabic tokens; (3) the students’ own polysyllabic types per polysyllabic types, and (4) polysyllabic tokens per core tokens. One reason why these ratios were used is linked to what was mentioned in section 4.3, namely the fact that examiners are sensitive to the number of types in learner speech, as well as to the rareness of vocabulary used (Lorenzo-Dus, 2007).

Table 8.8. Tokens, types, and ratios for the Low OP group in Test 5.

Stud ID	Tokens and types					Ratios			
	All tokens	Core tokens	Poly# tokens	Poly types	Own poly types	Poly tokens per All tokens	Poly types per Poly tokens	Own poly types per Poly types	Poly tokens per Core tokens
Low 1	329	319	10	7	4	.03	.73	.57	.03
Low 2	180	174	6	6	4	.03	1.00	.67	.03
Low 3	210	205	5	3	1	.02	.60	.33	.02
Low 4	303	295	8	6	3	.03	.75	.50	.03
Low 5	172	164	8	6	3	.05	.75	.50	.05
Sum	1,194	1,157	37	28	15	.16	3.80	2.57	.17
Mean	239	231	7	6	3.0	.03	.76	.51	.03

Polysyllabic

Table 8.9. Tokens, types, and ratios for the High OP group in Test 5.

Stud ID	Tokens and types					Ratios			
	All tokens	Core tokens	Poly# tokens	Poly types	Own poly types	Poly tokens per All tokens	Poly types per Poly tokens	Own poly types per Poly types	Poly tokens per Core tokens
High 1	503	477	26	20	10	.05	.77	.50	.05
High 2	702	678	24	16	10	.03	.67	.63	.04
High 3	665	620	45	29	27	.07	.64	.93	.07
High 4	852	819	33	23	19	.04	.70	.83	.04
High 5	928	891	37	25	20	.04	.68	.80	.04
Sum	3,650	3,485	165	113	86	.23	3.45	3.68	.24
Mean	730	697	33	23	17.2	.04	.69	.74	.05
# Polysyllabic									

Independent samples *t* tests revealed that the difference between the Low and High OP group was statistically significant at the .01 level for all five categories of word counts of tokens and types, with higher figures for the students in the High OP group in each of the counts. The results on polysyllabic tokens, polysyllabic types, and the students' own polysyllabic types (i.e. use of advanced/infrequent vocabulary) indicate that students in the High OP group produced more advanced words than the students in the Low OP group. I also carried out independent samples *t* tests on the four measures of word ratios. There was a statistically significant difference between the two groups for one of those measures, namely for the students' own polysyllabic types in relation to their total number of polysyllabic types ($p=.046$). The students in the High OP group used 86 polysyllabic types that were their own out of a total of 113 polysyllabic types (74%), whereas the students in the Low OP group used 15 that were their own out of a total of 28 (51%) (see Tables 8.8 and 8.9). Thus, students in the High OP group were more successful than those in the Low OP group in incorporating polysyllabic vocabulary into speech, a finding which, most likely, is linked to students in the High OP group having had access to more polysyllabic words in their mental lexicon.

The specific result on the students' own polysyllabic words is interesting because it can be viewed in light of Afitskaya's (2002:5) results on lexical sophistication. Afitskaya's examination of advanced tokens per all tokens yielded statistically significant differences between her two groups (advanced and intermediate level English students). A similar calculation in my study (polysyllabic tokens per all tokens) did *not* yield any significant difference between the High and Low OP group. However, when I studied the students'

own polysyllabic types (cf. Afitskaya's "advanced tokens") in relation their total polysyllabic types (cf. Afitskaya's "all tokens"), there was indeed a statistically significant difference also in my study, despite the small scale of it. That is, students in the High OP group not only used more advanced/peripheral words than the students in the Low OP group, they also used a higher ratio of sophisticated vocabulary, using Afitskaya's terminology.

As was mentioned in chapter 6, the etymology of the different polysyllabic types was also examined. English core vocabulary is to a large extent Germanic and often mono- or disyllabic (cf. Minkova & Stockwell, 2006:467). A list of the various polysyllabic types used in speech in Test 5 by the ten students in the High and Low OP group is provided in Appendix 16. With the help of the *Oxford English Dictionary Online*, the etymology of each polysyllabic type was identified and labeled "Germanic", "Non-Germanic", or "Hybrid" (see App. 16).

It has already been shown that the students in the High OP group produced more polysyllabic tokens and types than the students in the Low OP group. When the etymology of all polysyllabic types of each group was examined, I found that the Low OP group used more Non-Germanic vocabulary (76%) than the High OP group (63%). On the other hand, the High OP group used more Hybrids (15%) than the Low OP group (7%). However, because the total number of polysyllabic types used by the Low OP group was significantly lower than that of the High OP group, this is a halting comparison. Moreover, at times polysyllabic words were part of the actual test instructions which the students read aloud. Sometimes it was their interlocutor who first read the polysyllabic word, which was then used in speech by students in both the High and the Low OP group. In order to distinguish between the students' own polysyllabic words (cf. Tables 8.8 and 8.9) and those included in and elicited by the test materials, the latter were marked with an asterisk in Appendix 16. The remaining types (without an asterisk) are the students' own, so to speak. The students' own polysyllabic types are provided in a condensed format in Table 8.10.

A mere reading of the students' various own polysyllabic types listed for each group in Table 8.10 reveals that the words produced by the High OP group were not only more numerous than those produced in the Low OP group; a great majority of the words produced by the High OP group were clearly peripheral and sophisticated, especially for second language learners in the ninth grade.

Table 8.10. The students' own polysyllabic types.

Student ID	The Low OP group	The High OP group	Student ID
Low 1	<i>animals, anymore, family, restaurants</i> (4)	<i>animals, anything, area, discussion, everybody, everything, families, salary, trampoline, unfortunately</i> (10)	High 1
Low 2	<i>accepted, computer, everywhere, family</i> (4)	<i>actually, anything, disturbing, everyone, families, horrible, interesting, society, violence, whatever</i> (10)	High 2
Low 3	<i>animal</i> (1)	<i>addicted, another, anyone's, bicycle, dangerous, different, educated, education, environment, everyone, everything, family, fireplace, glass bottles, liquor store, living-room, mud-houses, neighborhood, opportunities, otherwise, paper bags, privacy, recycle, recycled, tinnitus, together</i> (27)	High 3
Low 4	<i>grandparents, understand, unusual</i> (3)	<i>actually, anyone, area, civilized, connection, countryside, easier, everything, expensive, fashion clothes, forever, Germanish, Germany, grandchildren, massage therapist, organization, otherwise, situation, teenagers</i> (19)	
Low 5	<i>anything, difficult, everyone</i> (3)	<i>accepted, another, apartment, cigarettes, difference, different, emergency, everybody, everyone, excellent, execute, family, favorite, neighborhood, newspaper, reminding, represents, separated, solution, usually</i> (20)	High 5
Total	15	86	Total

In another analysis of students' use of polysyllabic words, the number of types (used in 'Test 5) was correlated with the students' OP grades (based on Tests 1-5). The correlation was positive and statistically significant ($r=.905^{**}$) and the difference between the Low and High OP groups apparent (see Figure 8.1). The result was similar for polysyllabic tokens and the OP grades ($r=.888^{**}$). These results indicate that students who used more peripheral vocabulary probably gained from doing so. Therefore, it seems reasonable to conclude that students in the High OP group had acquired a more advanced vocabulary than students in the Low OP group. Students in the High OP group used a higher number of advanced lexical items in speech than students in the

Low OP group did. In addition, students in the High OP group used a significantly higher ratio of their own polysyllabic words in speech.

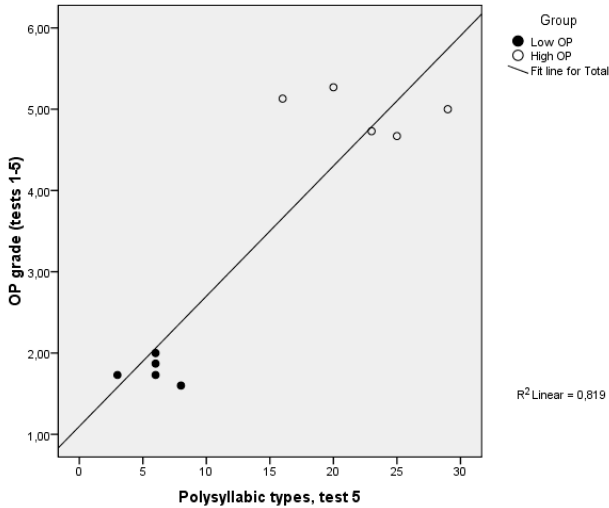


Figure 8.1. The relationship between polysyllabic types and oral proficiency for the High and Low OP group.

A final comment in this section is on the results on polysyllabic vocabulary as they relate to results on students’ extramural English. Based on data from all ten students, I found a positive correlation between “polysyllabic tokens/all tokens” and the EE variable ($r_s=0.673^*$). That is, the more EE, the higher the ratio of polysyllabic tokens produced per uttered tokens. The two groups had differing totals for EE, approximately 17 hours/week for the High OP group and 10 hours/week for the Low OP group, but this difference was not statistically significant. However, for one of the EE activities (“watching TV”), there was a significant difference ($p=0.025$): the High OP group watched close to seven hours on a weekly basis, whereas the Low OP group watched less, about two and a half hours per week. In sum, there was a fairly straightforward relationship also between polysyllabic vocabulary and EE, similar to what was found between the two written vocabulary tests and EE.

8.8 Summary and conclusion

The present chapter presented results on students’ vocabulary based on data from two written vocabulary tests, one cloze test, and one speaking test, namely

Test 5. Mean scores for the five EE subsets on two of the vocabulary tests, the Productive Levels Test and the Vocabulary Levels Test, revealed a fairly straightforward relationship between EE and vocabulary, a relationship which differed from the more complex one that was identified between EE and oral proficiency in chapter 7. Students in EE subset 1, i.e., those with the least amount of EE, had the lowest vocabulary scores. The scores were either the same or improved with each EE subset and ended with EE subset 5 having the highest mean scores on both vocabulary tests.

A Vocabulary Index Variable was created (see section 6.3.6.7) and it correlated positively with the EE variable. The correlation was statistically significant ($r_s=.357^{**}$). A gender difference was identified; for boys the correlation between the EE variable and the Vocabulary Index Variable was positive and statistically significant ($r_s=.590^{**}$). In contrast, the correlation between EE and vocabulary was negligible and not statistically significant for girls ($r_s=.011$), something which makes the result for girls more difficult to interpret than the result for boys. Similar to what was found for correlations between EE and oral proficiency, correlations between EE and vocabulary point in one direction, namely that boys in particular seem to be sensitive to EE. However, regression analysis revealed that for vocabulary, “playing video games” and “surfing the Internet” were the most important EE activities, followed by “watching TV”, the two reading activities, “listening to music”, and “watching films”. This means that the boys’ habit of spending significantly more time than the girls on playing video games and surfing the Internet is a more possible explanation why they outperformed the girls on vocabulary, than a higher degree of “sensitivity” towards EE (see also chapter 10).

Cloze tests are considered to measure general language proficiency, i.e., not only vocabulary (D. H. Brown, 2004:9), but I still decided to include a cloze test in my study on vocabulary. Results on the cloze test for the five EE subsets resembled results for the EE subsets and oral proficiency rather than results for the EE subsets and vocabulary. Because of that resemblance, it is likely that the cloze test indeed measured proficiency more than it measured vocabulary. Thus, results on the cloze test confirmed that there is no straightforward relationship between EE and general proficiency in English. That particular relationship is more complex and the relationship between EE and vocabulary is more straightforward in comparison, at least as measured in the present study.

Using polysyllabicity of words as a tool for a qualitative investigation of learners’ vocabulary worked well. Statistically significant differences were found

between the High and the Low OP groups on five word counts, where the High OP group outperformed the Low OP group on each one: the total number of tokens, core tokens, polysyllabic tokens, polysyllabic types, and the students' own polysyllabic types (i.e. the polysyllabic types that were not included in and elicited by the test material). The inclusion of an etymological analysis of the polysyllabic types added an interesting dimension but, since the Low OP group produced relatively few polysyllabic words, it was not possible to draw any major conclusions from that part of the vocabulary investigation. However, from an analytical point of view, a study of students' own polysyllabic vocabulary proved important. In the four different word ratios that were calculated, a significant difference was identified for "students' own polysyllabic types per polysyllabic types". In the High OP group, 74% of the polysyllabic types used were the students' own, whereas the corresponding percentage for the Low OP group was 51. Based on the results in the study on polysyllabic vocabulary, I conclude that the students in the High OP group had acquired a more advanced vocabulary than the students in the Low OP group. First, students in the High OP group used a higher number of advanced lexical items in speech than students in the Low OP group did. Second, students in the High OP group used a significantly higher ratio of their own polysyllabic words in speech. Finally, correlations between EE and various measures of polysyllabic vocabulary confirmed the fairly straightforward relationship between time spent on EE and vocabulary results that was previously mentioned.

In the following chapter, I present results for various background variables in relation to students' extramural English and their English results in school. I also address students' motivation and views on English. The purpose is to see whether background variables, motivation, and students' views on English may shed more light on the potential importance of extramural English in second language acquisition.

9 Background variables, motivation, and views on English: Analysis and results

The main part of the present chapter is based on data from the questionnaire and includes analyses and results on background variables, students' motivation, and students' views on English. First, four background variables are presented in section 9.1: experience of travels abroad (9.1.1), parents' educational background (9.1.2), the number of books in the home (9.1.3), and rural versus urban residency (9.1.4). Results are presented on both sample and class level, and correlations are made with EE and OP. In section 9.2, two motivational factors are in focus, self-efficacy and anxiety related to speaking, and each is examined in relation to EE and OP. A slightly different approach is taken in section 9.3, which includes an account of students' views on English. All results in section 9.3 are based on data from my questionnaire. However, it was also possible to compare my results with those based on data from another data set, namely the one of the *National evaluation of 2003 (NU-03)* (Oscarson & Apelgren, 2005), because a number of questions were identical in the two data sets. Three subsections are included in the account. The first one is about where English is learned (9.3.1). The second reports on students' self-assessment (9.3.2), and the third subsection presents the results on students' opinions about the school subject English. In 9.4, results are presented regarding the students' overall grade¹ and their subject final grade in English based on data from the students' leaving certificates from ninth grade. For the overall grade and subject final grade in English, comparisons are made between the sample and ninth graders nationwide with the help of official statistics from the Swedish National Agency for Education. In the same section, correlations are also made between EE/OP and overall grades. Finally, section 9.5 provides a summary and conclusion.

9.1 Background variables

There were a number of questions in the questionnaire which addressed background variables. I will discuss four of them in the present section, namely question 2 (students' experience of travels abroad), question 4 (residency), question 13 (mother's and father's educational background), and question 30 (number of books in the home). I chose to include these specific variables because each of them has previously been shown to be of importance for

¹ The overall grade (Swe. *meritvärde*) is described in section 6.2.7.

language learning or results in school (see chapters 2 and 3). The purpose of investigating background variables is to contribute to the general picture of the informants in a study, something which I believe these four variables will do. An alternative label could have been “socioeconomic variables”, because travels, residency, educational background, and the number of books in the home are all also connected to social and/or economic factors. I present the results for the four background variables in the order in which the corresponding questions appeared in my questionnaire. This means that I start with question 2, which asked about students’ experiences of travels abroad.

9.1.1 Travels abroad

As was discussed in chapter 3, traveling might play a role in language learning and attitudes to language learning. Question 2 in the questionnaire addressed students’ experience of travels. A majority of the students in my sample had spent time abroad, but it varied how many countries they had visited and how far away from Sweden they had traveled, for example. Based on their responses to question 2 (items a-e), students were clustered into one of four groups (A-D) depending on where they had been (see Table 9.1). Group A included students who had never been abroad. In my sample, seven students (9%) had never been outside Sweden. In group B, 18 students (22%) reported that they had only visited one or several of the Nordic countries. Students in group C had traveled farther; they listed that they had visited at least one European country outside the Nordic countries. This is the largest group: 37 students, or 46% of the sample. Finally, the last 18 students (22%) in group D reported that they had traveled even farther, i.e., outside Europe. In other words, 55 students out of 80 (69%) had traveled in Europe and/or the rest of the world. Table 9.1 provides the results for EE and OP for these four groups.

Table 9.1. Students’ experiences of traveling abroad: EE and OP grade.

Question 2 – Experience of travels – Four groups	EE (hrs/w)	N	Mean OP grade	N
A. Had never traveled abroad	22.1	7	2.6	6
B. Had traveled to one or several of the Nordic countries	15.8	18	3.1	18
C. Had traveled in Europe, outside the Nordic countries	19.7	37	3.5	34
D. Had traveled abroad, outside Europe	16.9	18	3.6	16
Total	18.4	80	3.3	74

A chi-2 test was used to compare the four groups based on the experience of travels and the groups were found to be different ($p=.000$). Values (hours/week) for the EE variable as a whole and the eight EE activities taken separately were examined for the four groups by use of ANOVA, but no differences were seen between the different groups. This means that going abroad appears to have no relationship whatsoever with students' EE. However, when the same four groups were compared with regard to the OP grade, a relationship was established between the four groups and OP. As can be seen in Table 9.1, the farther away the students had traveled, the higher their mean OP grade was. The mean OP grade was 2.6 for students who had never traveled abroad, 3.1 for those who had been to one or several of the Nordic countries, 3.5 for those who had traveled in Europe (group C), and 3.6 for the fourth group, which included the students who had been outside Europe. The differences between the four groups in reference to OP were close to being statistically significant ($p=.056$).

During time spent abroad, people who do not have English as their first language often find themselves in situations where they have to use English as a means of communication, situations which, perhaps, might be beneficial to their oral proficiency in English. However, at least in the case of the present study, it is probably more likely that travels abroad are only indirectly linked to students' oral proficiency as such experience might motivate students to study English and increase students' understanding of the importance of English as a global language, for example. Moreover, "travels abroad" can certainly be viewed as a socioeconomic variable because families might refrain from traveling due to costs, or they might not be interested in traveling for other reasons. A dispersion of groups A-D presented in Table 9.1 into the five OP subsets revealed that those who had never been abroad were in OP subsets 1-3 (very low, low, and intermediate level of OP). Moreover, OP subset 5 (i.e. the students with the highest OP grades) included the highest ratio of students who had traveled abroad outside Europe, and OP subset 4 the second highest ratio.

When the four school classes (see section 6.1.4) were compared, class 3 turned out to be the most widely traveled class with the largest proportion of students who had been outside Europe ($\approx 30\%$) and the smallest proportion of students who had never gone abroad ($\approx 4\%$). Class 4 had the second largest proportion of students who had been outside Europe ($\approx 22\%$). However, at the same time, class 4 comprised the highest ratio of students who had never been abroad ($\approx 20\%$), which was in great contrast to the other three classes (approximately 4% in class 3, see above, and approximately 5% in classes 1 and

2). Classes 1 and 2 had similar backgrounds with regard to traveling. For example, among the students who had traveled in Europe outside the Nordic countries, approximately 20% were from class 1 and another 20% from class 2.

9.1.2 Parents' educational background

Question 13 asked specifically about the educational background of the mother and father and students ticked one out of four options for each parent. The responses were coded into either “1” or “2”. The former included three of the options, namely “nine-year compulsory school” (Swe. *grundskola*), “senior high school” (Swe. *gymnasieskola*), and “I don’t know”, and the latter included responses that said that the parent had a university level education. I found it reasonable to assume that ninth graders of highly educated parents would actually know about their parents’ educational background and tick that specific box. Based on this assumption, those who responded “I don’t know” could be grouped together with those who responded “nine-year compulsory school” or “senior high school”.

In total, 20 students had parents where both the mother and the father had a university level education, here called group A. The others, 53 students, made up group B (the rest were missing values). Students in group A had more EE (19.6 hrs/w) than the students in group B (17.6 hrs/w), but the difference was not statistically significant ($p=.553$). The values for the eight EE categories were fairly even when both groups were compared. However, despite the fact that the reported values overall were very low for the EE activity called “reading newspapers and/or magazines”, a significant difference was found here ($p=.004$). The students in group A read more newspapers and magazines in English than those in group B. Furthermore, students in group A had a higher OP grade (3.5) than the students in group B (3.3), but statistical significance was not found ($p=.390$). In sum, there was a relationship between students who reported reading in English in their spare time, the educational level of their parents, and the students’ results on oral proficiency in school.

I studied the dispersion of groups A and B in the five EE subsets. Students in group A (both parents had a university level education) were evenly dispersed through all the EE subsets, something which was to be expected since students’ EE was not connected to the parents’ educational background, at least not in my study. In contrast, when the groups were examined with regard to the five OP subsets, there were no students from group A in OP subset 1 (students with a very low level of oral proficiency). This finding was also to be expected considering what was found above on the relationship

between parents' educational background and students' results on OP in school.

As was described in chapter 6, the four school classes in my study were from different schools and communities. A striking difference was identified with regard to the parents' educational background in the classes. In class 4, 40% of the students had both a mother and a father with a university level education. In class 3, the corresponding ratio was 35% and in class 1 20%. In class 2, however, it was much lower (5%). It was shown in chapter 7 that class 2 had the lowest OP grade compared to the other three classes. The parents' educational background might help explain that particular finding regarding class 2.

With regard to gender, about one third of the girls and one fifth of the boys had two parents with university level education. In other words, in my sample, there happened to be more girls than boys who had both a mother and a father who were highly educated. This is a finding which should be kept in mind when results for gender are discussed in the present thesis.

Finally, in my investigation of the present background variable, I also examined if having at least one parent with a university level background (as opposed to both) seemed to be connected to students' EE as a whole or to OP. The results revealed that students with at least one university-educated parent (N=39) had less EE per week than students who did not have any parent with a university education/students who did not know about the parent's educational background (N=34), 17.7 hours per week on average as opposed to 18.7. Regarding oral proficiency, students with at least one university-educated parent had a slightly higher mean OP grade (3.4; N=36) than the rest (3.3; N=31). However, the differences in EE and OP between the groups were very small and in neither case statistically significant.² With regard to parents' educational background, in the present study it thus seems that two university-educated parents were "required" in order to permeate the data and be statistically significant, and it did so in the results related to reading (newspapers/magazines), but not in the results related to EE as a whole. Reading was shown to be highly important for OP (cf. section 7.4.4); hence, there is evidently a connection between students who reported reading in English in their spare time, the educational level of their parents, and the students' results on OP.

² There were no significant differences between the two groups when the eight EE activities were examined either.

9.1.3 *Number of books in the home*

I measured the number of books in the students' homes with the help of question 30 in my questionnaire. As I said in chapter 6, measuring the number of books is an attempt to get an indication of what Bourdieu termed cultural capital, even though I am fully aware that one single question cannot possibly capture the whole Bourdieuan concept of cultural capital (see chapter 2). Nevertheless, this type of question has been used by others and, in fact, been labeled "cultural capital" then (see e.g. Öquist & Wikström, 2006:18). Öquist and Wikström (2006) found that the number of books in the home correlated positively with students' grades. That is, the more books there were in the students' homes, the higher their grades in school. In my questionnaire, the students were asked to tick one box out of five for question 30, the one that best described how many books they thought they had at home. I investigated my sample by dividing it at 50 and 200 books respectively, similar to what was done in Öquist and Wikström (2006:34).

In the first investigation of my sample, with the dividing line at 50 books, group A reported having 0-50 books at home and group B more than 50. Group B had higher values than group A on the following points of measure: the OP grade, the final English grade, the overall grade, the Productive Levels Test, and the Vocabulary Levels Test (see Table 9.2). Independent samples *t* tests revealed that the difference between the two groups was significant for one of the points of measure, namely the overall grade ($p=.021$). With regard to EE as a whole, group A had a lower value (15.7 hrs/w; N=13) than group B (18.9 hrs/w; N=67), but the difference was not significant. Likewise, there were no significant differences between the groups for the eight EE activities.

In the second investigation, the cut off point was instead set at 200 books; i.e., students in group A reported having 0-200 books at home and group B more than 200. This time, group B had higher values than group A for the same points of measure as above and, in addition, independent samples *t* tests revealed that the difference between the groups was significant for the OP grade ($p=.035$), the final English grade ($p=.014$), the overall grade ($p=.000$), the Productive Levels Test ($p=.044$), and the Vocabulary Levels Test ($p=.006$) (see Table 9.2). For the EE variable, group B also had a higher value (19.6 hrs/w; N=32) than group A (17.6 hrs/w; N=48), but the difference was not statistically significant. No significant differences were found between the groups for the eight EE activities.

Table 9.2. Results for students on five points of measure with cut off points between the sample at 50 and 200 books (case = student).

Number of books	Group			OP grade	English grade	Overall grade	Productive Levels Test	Vocabulary Levels Test
0-50	A	N	Valid cases	10	13	13	13	12
			Missing cases	3	0	0	0	1
			Mean	3.1	2.4	186*	14.2	57.0
>50	B	N	Valid cases	64	66	66	67	66
			Missing cases	3	1	1	0	1
			Mean	3.4	2.8	226*	16.4	60.6
0-200	A	N	Valid cases	43	48	48	48	47
			Missing cases	5	0	0	0	1
			Mean	3.2*	2.6*	201**	14.7*	56.6**
>200	B	N	Valid cases	31	31	31	32	31
			Missing cases	1	1	1	0	1
			Mean	3.6*	3.0*	249**	18.2*	65.3**

*Sig. at the .05 level, **Sig. at the .01 level

The results of my sample regarding the number of books in the home (Table 9.2) were very similar to those in Öquist and Wikström (2006), which in turn were based on data from the *National evaluation of 2003 (NU-03)* (Oscarson & Apelgren, 2005), where the statistical population was all ninth graders in Sweden. My results show that ninth graders from homes with a certain number of books (“homes with strong cultural capital”) apparently do better in English than ninth graders from homes with few or hardly any books (“homes with weak cultural capital”). Moreover, the more books there were in the homes, the more the students’ results improved: when the dividing line was moved from 50 to 200 books, the differences in the results grew when comparisons were made between students from homes with a strong cultural capital and students from homes with a weak cultural capital.

Analyses of the responses to question 30 established a connection between the number of books in the home and various measures for English in my sample. Moreover, the similarity of the results in my study and those in Öquist and Wikström (2006) strengthens the reliability of the results found in the present study and adds face validity to my sample. As pointed out in the two previous sections, no correlations were found between EE and the investigated background variables (travels abroad, parents’ educational background). Similarly, there was no correlation between the background variable investigated here (the number of books in the home) and EE. That is,

the number of books in the home (“cultural capital”) does not seem to correlate with students’ EE, neither for students’ total EE nor for any of the EE activities. In other words, the present variable does not explain why students’ EE varies.

9.1.4 Rural versus urban residency

Question 4 in the questionnaire asked whether the students needed transport to go to school due to the distance between their home and the school. It is reasonable to assume that students who responded “yes” to this question live in a more rural setting rather than in an urban one. I would like to stress that “urban” in my study refers to small or medium-sized towns and not major cities, which would be easy to assume. Students who have to go by bus (or other means of transportation) to school have longer school days than other students, but it was not within the scope of the present study to examine whether longer school days per se may have an impact on school results. What has been shown in other studies, however, is that students who live in the countryside, i.e. in rural areas, generally achieve lower grades in school than students who live in urban areas (Öquist & Wikström, 2006:54-55). Moreover, with regard to grades, there is a gender difference: the boys lag behind the girls. Boys’ grades are lower than the girls’ regardless of where they live in Sweden, but it is noted that boys’ lower grades in some rural settings makes the gender difference very obvious (Björnsson, 2005:57). My sample includes both students who live in rural settings and students who live in urban settings, hence the current investigation.

In the present analysis, students were divided into two groups depending on their residency, i.e., whether they lived in a rural setting or an urban one. In Table 9.3, results for these two groups are given for a number of points of measure together with significance values derived from independent samples *t* tests. Table 9.3 shows that the students who lived in rural settings had lower scores/grades than the students in urban settings on four points of scholastic measure, namely the OP grade, the English grade, and the two vocabulary levels tests. The difference between the groups was significant at the .01 level for the OP grade. Conversely, with regard to the overall grade, the students from rural areas instead had a higher value than the students from urban areas, but this difference was not statistically significant.

Table 9.3. Results for six points of measure and residency: rural versus urban.

Point of measure	Residency				Sig. (2-tailed)
	Rural		Urban		
	N	Mean	N	Mean	
EE (hrs/w)	49	15.9	31	22.4	.027*
OP grade (1-6)	45	3.1	29	3.7	.002**
Productive Levels Test (max 45)	49	14.9	31	18.0	.079
Vocabulary Levels Test (max 90)	47	58.5	31	62.4	.215
Overall grade (max 320)	48	224	31	213	.431
English grade (1-4)	48	2.7	31	2.8	.794

*Sig. at the .05 level, **Sig. at the .01 level

In terms of EE, students who reported living in urban settings had a higher amount of EE than the others (see Table 9.3). The difference was statistically significant at the .05 level. In comparison with students who lived in rural settings, the students who lived in urban settings reported more time on seven out of eight EE activities. The exception was video games (rural residency 4.4 hrs/w; urban residency 3.2 hrs/w). The difference between the groups was statistically significant at the .05 level for “listening to music” and “reading books”. Students living in urban areas reported listening to music more than nine hours per week on average, which can be compared with a little bit less than five hours per week for those in rural settings. With regard to reading books in English, neither group read much but the students in urban areas reported reading almost half an hour a week on average, whereas the others barely reported anything (on average approximately six minutes per week). It has already been established that “listening to music” and “reading books” were two of the three most important activities for oral proficiency (see section 7.4.4). Therefore, it was not surprising that the students in rural areas had lower results on OP than the students in urban areas.

I split the data for gender to create four groups which I will here refer to as “rural boys”, “rural girls”, “urban boys”, and “urban girls”. Next I compared the mean EE for the four groups and found that rural girls were clearly the ones with the least EE (see Table 9.4). Rural boys had more EE than rural girls but less than both urban boys and girls. Urban girls had less EE than urban boys but more than rural boys and girls. Consequently, urban boys had most EE of all four groups. When the OP grades were examined for the same four groups I found that urban girls clearly had the highest mean OP grade (see Table 9.5). They were followed by the group of urban boys. Rural boys and girls had the same mean OP grade (3.1).

Table 9.4. EE (hours/week) for boys and girls with regard to residency.

Gender	Residency			
	Rural		Urban	
	N	EE	N	EE
Boys	23	34.5	13	42.0
Girls	26	23.9	18	40.6

Table 9.5. Mean OP grades for boys and girls with regard to residency.

Gender	Residency			
	Rural		Urban	
	N	OP	N	OP
Boys	22	3.1	12	3.4
Girls	23	3.1	17	4.0

It was shown in section 7.4.1 that EE and OP correlated positively and significantly for boys, whereas the correlation between EE and OP was different for girls. It was positive, but less so than for boys and non-significant. When I used the present background variable, residency, in a similar correlation analysis between EE and OP, I found that EE and OP correlated positively and significantly for the forty-five students in rural settings ($r_s=.402^{**}$), but there was no correlation between EE and OP for the remaining twenty-nine students in urban settings ($r_s= -.074$). To shed more light on the potential importance of residency as a background variable to understand EE, I examined it in relation to the five EE subsets.

The analysis of the residency variable and the five EE subsets revealed that the proportion of students from urban areas within each subset grew steadily with each subset (see Fig.9.1). In other words, the more time spent on extramural English per week (from EE subset 1 with the least time spent on EE, to EE subset 5 with the most time spent on EE), the larger the proportion of students within each subset from an urban setting. A similar examination of the residency variable in relation to the five OP subsets yielded a slightly different picture, even though the general tendency was the same (see Fig. 9.2). There was a higher ratio of students from rural settings towards the left hand side of the scale, i.e. in OP subsets 1 and 2 (very low and low level of OP), and a higher ratio of students from urban settings towards the right hand side of the scale in OP subset 5 (very high level of OP).

In sum, residency clearly has a connection with students' oral proficiency (see Table 9.5 and Figure 9.2). Whether residency may be a factor which facilitates understanding of EE is more difficult to say, but it appears to have

some relevance. For example, when I divided the sample on the basis of the students' residency, there was indeed a statistically significant difference in EE depending on where the students lived: students who lived in urban areas had more EE per week than those in rural areas ($p=.027$). Another reason why residency may play a role in explaining students' EE was the positive and statistically significant correlation between EE and OP for students in rural areas (see above). However, the groups defined by residency were more uneven in terms of the number of students per group than the groups that were defined by gender. Furthermore, the statistically significant correlation between EE and OP for boys was more positive than the correlation between EE and OP for students in rural areas. Therefore, my conclusion is that residency plays a role in predicting students' EE, but it is less important than gender.

9.2 Motivational factors

Motivation is important for all learning, and for successful L2 acquisition motivation is often considered crucial (see e.g. Dörnyei, 2001). Therefore, several questions in the questionnaire were included with the purpose of measuring students' motivation for learning English. Two motivational factors are examined in the present section: self-efficacy (9.3.1) and anxiety related to speaking (9.3.2).

9.2.1 Self-efficacy

As was discussed in chapter 2, self-efficacy refers to people's judgment of their own ability to carry out specific tasks (cf. Dörnyei, 2001:22-23). Self-efficacy has been found to be a good predictor of subsequent grades in English (Schunk et al., 2008:54). In the questionnaire, there were three questions that relate to self-efficacy: question 12, question 21 (items a-c), and question 26. With the help of hierarchical cluster analysis, it was possible to identify and create an index variable for self-efficacy made up of seven items (see Table 6.8). These items addressed students' beliefs about their future final grade in English, the effort they put into their English studies, and whether they took responsibility for their English studies. In addition, the items addressed students' interest in English and whether they considered English important. The index variable also comprised information about whether the students believed English would be important for future studies or work. That is, I interpret the term "self-efficacy" broadly in the present analysis. The index variable was reliable (Cronbach's $\alpha=.703$) and normally distributed.

The self-efficacy index variable was correlated with the EE variable and the two variables correlated positively with statistical significance ($r_s=.293^*$). With regard to the eight EE activities, self-efficacy correlated positively with all of them. The correlations were statistically significant at the .01 level for “surfing the Internet” and “listening to music” and at the .05 level for “watching TV”, “reading newspapers and/or magazines”, and “other activity”.

I also investigated self-efficacy in relation to the five subsets of the sample which were based on the EE variable (see Table 9.6). In comparison with the other subsets, EE subset 1 had the lowest value for self-efficacy and EE subset 3 the highest. EE subsets 2 and 5 had very similar values. The differences between the subsets on self-efficacy were statistically significant ($p=.037$) despite the fact that missing values reduced subset 5 to only two students. The most interesting finding in Table 9.6 is probably that, relatively speaking, the lowest value for self-efficacy was found in EE subset 1. That is, the fact that students in EE subset 1 had low self-efficacy might contribute to our understanding of why these particular students did not spend as much time on EE as students in the other subsets did. It is possible that spending time on out-of-school English activities was perhaps not very appealing for students in EE subset 1 since they considered their own ability to manage such activities low.

Table 9.6. Self-efficacy and the five EE subsets.

EE subset	Mean self-efficacy	N
1	21.7	14
2	23.1	30
3	24.2	20
4	22.2	6
5	23.0	2
Total	23.1	72

The previously identified differences between boys and girls with regard to EE on the one hand and OP or vocabulary on the other – with positive and statistically significant correlations between the variables for boys, but negligible correlations for girls – reappeared when the EE and self-efficacy variables were correlated with each other. Again, there was a positive and statistically significant correlation for boys ($r_s=.438^{**}$). For girls, there was also a positive correlation but less so than for boys and not significant ($r_s=.107$). This means that for boys, there was a relationship between how much time they spent on EE and their self-efficacy, i.e. the more EE, the higher self-efficacy (and vice versa), whereas no such relationship could be confirmed for girls. I would like

to emphasize that the relationship described between EE and self-efficacy is exactly that: a relationship, or a correlation, which is the proper statistical term to use. By no means should this correlation be interpreted as a cause and effect relationship between EE on the one hand and self-efficacy on the other. A correlation only describes in what way two variables relate to each other: positively, negatively, negligibly, or not at all. In addition, the correlation may be statistically significant, i.e. not due to chance, or not statistically significant, i.e. difficult to verify or falsify. However, as has previously been mentioned, I will return to the various correlation analyses in chapter 10, where I interpret them in terms of strength (cf. section 6.3.6.5).

Only a slight gender difference was found when self-efficacy was correlated with OP. For both boys and girls there were positive and statistically significant correlations between the two variables but, relatively speaking, the correlation was more positive for boys ($r_s=.728^{**}$) than for girls ($r_s=.588^{**}$). The conclusion is, without a doubt, that students' self-efficacy was very important for their oral proficiency. Table 9.7 shows that the values for self-efficacy increased with each OP subset, from 19.0 for the students with a very low level of oral proficiency in OP subset 1, to 26.0 for the students with a very high level of oral proficiency in OP subset 5. In addition, these observed differences between the OP subsets were statistically significant ($p=.000$).

Table 9.7. Self-efficacy and the five OP subsets.

OP subset	Mean self-efficacy	N
1	19.0	5
2	21.5	15
3	23.5	24
4	24.7	18
5	26.0	5
Total	23.3	67

9.2.2 Anxiety related to speaking

Anxiety related to speaking was the second motivational factor to be examined. There were two questions in the questionnaire that tapped into that specific area of interest, namely questions 22 (fear of making mistakes while speaking English) and 28 (state of feeling when speaking English in the classroom). These two questions were turned into an index variable aimed at measuring anxiety particularly related to speaking in English. Cronbach's alpha for the index variable was .746, an optimal value for reliability.

The anxiety index variable had a skewed distribution to the right; i.e., there were more students with low values of anxiety than there were students with high values. Thus, most students in the sample were not very worried or anxious about speaking in English, which was a positive finding.

Similar to what I did with regard to the other motivational factor, self-efficacy, the anxiety index variable was correlated with EE and OP. I will start by relating the results for EE and anxiety. First of all, there was a negative correlation between EE and anxiety ($r_s = -.136$), but the correlation was non-significant and, in addition, close to negligible. However, for one of the EE activities, “playing video games”, there was indeed a significant negative correlation with anxiety ($r_s = -.225^*$). In other words, among those who reported playing video games in English, the more time they had listed, the less anxious they reported feeling when speaking in English. Knowing this, it was not surprising to see that the lowest value for anxiety was in the small EE subset 5 (see Table 9.8), because in that subset, three out of four students had extensive video game playing habits. The highest anxiety value was found for the students in EE subset 1.

Table 9.8. Anxiety and the five EE subsets.

EE subset	Mean anxiety	N
1	4.1	15
2	3.4	33
3	3.6	20
4	3.5	6
5	3.2	4
Total	3.6	78

The anxiety index variable and OP also correlated negatively; the correlation was statistically significant ($r_s = -.380^{**}$). This means that the more anxious the students reported being, the lower their oral proficiency was assessed to be (as measured by the OP grade). This finding confirms previous studies on anxiety in L2 acquisition, which were accounted for in chapter 2 (see e.g. R. C. Gardner & MacIntyre, 1993; Matsuda & Gobel, 2004; Pappamihiel, 2002; Xiu Yan & Horwitz, 2008). A comparison of the results for the anxiety index variable revealed that, relative to others, students in OP subset 2 (i.e. low level of OP) showed the highest level of anxiety (see Table 9.9). Those with least anxiety were in OP subsets 4 and 5, 26 students in total. These 26 students were the ones with least anxiety related to speaking in the sample and, at the same time, they had the highest grades on oral proficiency, something which

was to be expected considering the findings in other studies on the same topic. The differences between the five OP subsets were significant ($p=.002$).

Table 9.9. Anxiety and the five OP subsets.

OP subset	Mean anxiety	N
1	3.4	5
2	4.2	16
3	3.8	25
4	2.7	21
5	2.8	5
Total	3.5	72

9.2.2.1 Students who speak English regularly in their spare time

Responses to question 5 (items a-c) in the questionnaire revealed information about students’ habits of speaking English regularly in their spare time. In my sample, fourteen students responded in the affirmative to question 5, item a: “Do you speak English regularly to a close relative, a friend, or someone else?” Eight of these fourteen students said that they speak English “once or a few times per week”; three said that they did so on a daily basis. One student said that s/he speaks in English “once or a few times per month” and the final two students reported doing it “once or a few times per year”. Thus, there was a qualitative difference in terms of how regularly (frequently) students spoke English, but the majority of the fourteen students indeed claimed to do it at least weekly. Ten of the students were boys and four were girls. The three students who said that they speak English on a daily basis were all boys with mean values for both EE and OP that were above average in the sample (EE 34.3 hrs/w; OP 3.5).

The sample was split into two groups (groups A and B) based on the responses to question 5 (item a), so that group A included the previously mentioned fourteen students and group B the rest, i.e. those who reported that they did not speak English regularly in their spare time. Independent samples *t* tests were then used to examine differences in mean values for EE and OP between groups A and B.

Results for EE showed that students in group A had much more EE than those in group B (28.1 h vs. 16.3 hrs/w; $p=.001$). The explanation for the difference lies in three of the EE activities where group A had significantly higher figures than group B: “playing video games” (11.7 vs. 2.3 hrs/w; $p=.000$), “surfing the Internet” (1.9 vs. .4 hrs/w; $p=.000$), and “other activity”

(1.0 vs. .2 hrs/w; $p=.029$). Students in group A (students who had responded that they speak English regularly) were evenly distributed in EE subsets 2-5, but none was in EE subset 1, which was the subset that included the students with the lowest amount of EE per week. When the same group of students (group A) was dispersed into the OP subsets, most ended up in OP subsets 3 and 4 (intermediate and high level of oral proficiency); there was one single student in OP subset 5 and another in OP subset 2. None of the students in group A was in OP subset 1, the subset that included students with a very low level of OP.

Having analyzed responses to question 5, it is possible to briefly describe students in my study who spoke English regularly in their spare time. In general, they had much more EE than students who did not speak English regularly in their spare time and they spent significantly more time playing video games, surfing the Internet, or doing other activities in English. Furthermore, students who had a habit of speaking English in their spare time had at least an intermediate level of oral proficiency. However, again it is impossible to say anything with regard to cause and effect between variables. Nevertheless, what can be said is that the students' habit of speaking English regularly in their spare time is connected with their EE and OP.

9.3 Students' views on English

Data from the *National evaluation of 2003 (NU-03)* (Oscarson & Apelgren, 2005) is public and available as an SPSS-file. Four questions in my questionnaire were identical (or almost identical) to questions in *NU-03* and they all referred to students' views on the school subject English. In this section, responses to these questions are compared and discussed for my sample and the *NU-03*. As was mentioned in chapter 1, the *NU-03* was representative of Sweden and included more than 6,700 informants in grade 9. The comparisons made in this section will show to what extent my sample, with the statistical population "ninth graders in *Western Svealand*", is similar to the statistical population of the *NU-03*; that is, "all ninth graders in Sweden". The results are presented in three subsections. The first subsection presents the results on students' beliefs about where English is learned, followed by the second subsection with an account of students' self-assessment. The third one is about students' opinions about the school subject English.

9.3.1 Beliefs about where English is learned

Question 17 in my questionnaire corresponded to question 23 in the *NU-03* and referred to students' beliefs about where they learned their English. There

were four response categories, ranging from the belief that most of what the student had learned in English had been learned in school, to the belief that most of what the student had learned had been learned outside school. I decided to dichotomize the response categories into two, drawing a line between the belief that you had learned most English *in* school as opposed to the belief that you had learned most English *outside* school. That is, response categories 1 and 2 were collapsed into one, as were response categories 3 and 4.³ Responses from the two samples (*NU-03* and my own) were compared by use of a chi-2 test, which is suitable for comparisons of variables based on nominal data.

A significant difference was found between the *NU-03* sample and mine ($p=.000$) in the way informants responded to question 17. In my sample, approximately 69% of the students said that they had learned most of their English in school and the remaining 31% that they had learned most outside school. In the *NU-03*, however, the picture was almost the opposite: approximately 18% said that they had learned most of their English in school and 82% responded outside school. This is an interesting difference between my sample and the larger *NU-03* sample. The results in my study on the relationship between extramural English and vocabulary on the one hand, and extramural English and oral proficiency on the other, showed positive and statistically significant correlations in both cases. That is, the more time spent on extramural English, the higher results on vocabulary and oral proficiency, albeit with a more straightforward relationship between EE and vocabulary than between EE and oral proficiency, and with different relative importance of individual EE activities for vocabulary and oral proficiency. The responses for the *NU-03* sample to the present question – where students believe they learn English – leaned significantly more towards a belief in out-of-school learning than my sample did. Nevertheless, for my sample, statistically significant positive correlations have been revealed between the students' EE/out-of-school activities and their results in school on both oral proficiency and vocabulary. The fact that a higher ratio of the students in the *NU-03* study (compared to mine) believed in extramural English learning makes it reasonable to conclude that extramural English definitely plays a role in second language acquisition.

³ In the *NU-03*, there were in fact five response categories. They were dichotomized in the same manner as the response categories in my questionnaire, that is, response categories 1 and 2 were collapsed into one, and categories 3, 4, and 5 into another.

9.3.2 Self-assessment of work in the school subject English

In my questionnaire one question referred specifically to students' self-assessment of their work in the school subject English: question 21 (items a, b, and c). The corresponding question in the *NU-03* was number 15 (items a, b, and d). All three items were formulated as statements and the students were to tick one box out of four, the one that best matched their opinion, from strong agreement with the statement to strong disagreement.

For question 21-a, which proposed that the students do their best in the subject, the responses in my sample and those in the *NU-03* were very similar. Both groups to a large extent claimed to do their best in the subject and the result indicated that students in general try hard to do well in English. Students in my sample were slightly more positive than those in the *NU-03*. For question 21-b, which proposed that the students take responsibility for their work in the school subject English, the responses were also very similar between the groups, but with students in my sample responding a little bit more positively than the others. On the whole, the conclusion is that students in both samples said that they take responsibility for their work in English, at least in their own opinion. For question 21-c, which proposed that the students are able to demonstrate their skills in English to their teacher, the results from the previous two questions were repeated: both samples responded in a similar fashion. Students in both studies stated that they are given the chance to show their English skills to their teacher. Again, the students in my sample were slightly more positive than the students in the *NU-03*. The fact that the students in my sample in general gave slightly more positive responses than the students in the *NU-03* probably is due to the specific circumstances at the time of filling in the questionnaire. The students filled it in at the very beginning of the school year knowing that they were part of a study which, among other things, offered the students some extra attention from me, their teacher, and the school. It is reasonable to assume that such circumstances affected their responses to become more positive than those of the *NU-03*.

The final analysis of students' self-assessment is related to their final grades in English. Question 12 in my questionnaire corresponded to question 16 in the *NU-03*. The students were to answer which final grade in English they thought they would be awarded. In my sample, the students' responses indicated a mean "believed"/expected grade of 2.93 (N=76), which means that the expected grade on sample level was just below a "pass with distinction"

(Swe. *nål godkänd*). The mean “believed”/expected grade for the *NU-03* sample was slightly higher, 3.06 (N=5,924), just above “pass with distinction”.⁴ A chi-2 test showed no significant difference between the grade expectations of students in the two samples. That is, students’ beliefs about their final grades in English were approximately the same in both groups.

In sum, no differences were identified between students’ self-assessments of their work in the school subject English when my sample was compared with the much larger student group in *NU-03*.

9.3.3 Opinions about the school subject English

Question 26 consisted of a bundle of questions (items a-j) which all related to students’ opinions about the school subject English. It corresponded to question 2 in *NU-03*. Since there were so many items in this question, I decided to use hierarchical cluster analysis on both data sets (i.e. my sample and *NU-03*) to examine whether the data sets would behave similarly. Interestingly, results showed that my data set yielded the same clustering of variables as did the other, i.e. the *NU-03* data set. This finding confirms that the sample used in the present thesis is representative because its behavior is very similar to the behavior of the much larger *NU-03* sample.

I dichotomized the four response categories (from strong agreement to strong disagreement) into two for all items in question 26, similar to the procedure that was described in section 9.3.2. That is, the two response categories for agreement were collapsed into one, and the two response categories for disagreement into another. The responses of my sample were then compared with those of the *NU-03*.

Results showed no major differences between the two groups for any of the items, with one exception (item h, “There is too little time for English [in school]”). In my sample, 28% of the students agreed with the statement, which can be compared with 38% for the students in the *NU-03*. Thus, a slightly higher proportion of the students in the *NU-03* sample would have liked more time for English in school in comparison with students in my sample. With regard to those who disagreed with the statement (i.e., they did *not* think there was too little time for English in school), 72% of the students in my sample and half of the students in the *NU-03* responded in such a manner; a result which, most likely, means that these students were more or less pleased with the

⁴ Due to missing values, the number of informants who answered the present question is lower than the total number of participating students in both my study and the *NU-03*.

number of hours they received in English. I base this interpretation of the result also on the students' responses to question 26, item i, which revealed that the vast majority of the students in both samples in fact did not think there was too much time for English in school. In sum, then, the results on whether there is too much or too little time for English in school reflect the previously drawn conclusion that students in my sample thought they had learned most of their English *in* school as opposed to *outside* school (see section 9.3.2).

There were other interesting findings in relation to the remaining items in question 26. Since there were no apparent differences between my sample and the *NU-03* with regard to these items, the discussion that follows is basically valid for all ninth graders in Sweden. For example, the great majority of Swedish ninth graders were very interested in the subject English and almost all of them found it highly important to know English. Extremely few disagreed with the statement "I think that it is important to have a good knowledge of English". Moreover, to a great extent the students said that their parents also found English important. A majority of the students ($\approx 90\%$) were rarely absent from English lessons. The students' opinions varied more in regard to whether they "work with English only to pass the tests" (26-d).⁵ Here, three out of ten agreed with the statement (that is, they mainly work with English to pass the tests) but as many as seven out of ten had other reasons for studying English. In general, students thought that they would need English for future studies and work. Finally, the results clearly showed that English was regarded as a very important and interesting subject in school by most of the students. However, students also found English challenging. Approximately four out of ten ninth graders thought that English was a difficult subject and, as I said above, this result holds for both my sample and the *NU-03*. The ratio of students who found English difficult was fairly high, something which I think Swedes in general are unaware of.

9.4 Correlations between EE/OP and overall grades

In the present section, correlations are made between EE and OP and the students' overall grade (see section 6.2.7). The overall grade is based on the sum of grade points from sixteen subjects and the maximum one can get is 320.

Results for the four school classes and totals are shown in Table 9.10. The differences in mean overall grades between the classes were not statistically significant. Table 9.10 shows that the most heterogeneous class was class 4 and

⁵ It is possible to interpret students' opinions also as a measure of their instrumental motivation (see section 2.3).

the most homogeneous one was class 3. An independent samples *t* test showed that girls had a higher mean overall grade (233; N=43) than boys (204; N=36); this difference was significant ($p=.028$). Girls had higher grades than boys in all subjects except Mathematics and PE. Of the ten highest overall grades in the sample, seven belonged to girls and three to boys. One of these ten students was in class 1; the rest were three students each in classes 2-4. The overall grade ranged from 20 (a student in class 4) to 320 (a student in class 3).

Table 9.10. Mean overall grade for the four classes and in total.

Class	Mean overall grade	N	STD
1	223	20	51
2	200	18	61
3	231	22	43
4	222	19	73
Total	220	79	58

No correlation was found between students' EE and overall grades. There was, however, a positive correlation ($r_s=.359^*$) between the overall grade and one of the EE activities, namely "reading books". This result confirms previous studies on reading and academic results (see e.g. Sylvén, 2004). In contrast to the lack of correlation between EE and overall grades, there was a positive correlation between students' OP and their overall grade ($r_s=.615^{**}$).

In section 9.3.2, it was shown that students in my sample thought they would be awarded final grades in English with a mean value of 2.93. When I analyzed the outcome – that is, the grades the students were in fact awarded at the end of grade 9 – I found that the students in my sample had slightly overestimated their own level of English. Based on the grades awarded by the teachers, the mean value was 2.75. Nevertheless, a strong correlation was found between the students' self-assessed English grades and the ones awarded by the teachers ($r_s=.753^{**}$). The result strengthens the reliability of students' responses in the questionnaire.

The results regarding the final grade in English for the five EE subsets are presented in Table 9.11. The differences between the subsets were statistically significant ($p=.010$). EE subset 5 (the students with the highest amount of EE) had the highest mean final grade in English, but as was the case with the OP grade for each EE subset (see Table 7.8), the English grade did not improve with each EE subset. The EE variable and the final grade in English correlated

($r_s = .317^{**}$), but the relationship between the variables was more complex than the relationship between for example EE and vocabulary.

Table 9.11. The five EE subsets and the final grade in English.

EE subset	Mean		
	English grade	N	STD
1	2.2	15	.4
2	2.9	34	.8
3	2.8	20	.7
4	2.5	6	.8
5	3.2	4	.5
Total	2.8	79	.7

The Swedish National Agency for Education provides official statistics on all subject grades and overall grades.⁶ I compared the final grades in English for my sample with those for Sweden. I used official data for the year 2007, which was the year in which the students in my sample finished school. Again, I found that data for my sample did not deviate much from the data for the whole of Sweden. The percentage of students with the grade “pass” was almost identical ($\approx 38\%$) in both data sets. In my sample, approximately 42% of the students had a “pass with distinction” and the corresponding percentage for Sweden was slightly lower, 37. For “pass with special distinction”, on the other hand, Sweden had a slightly higher percentage ($\approx 18\%$) than my sample ($\approx 16\%$). Viewed in combination, the minor differences of the two highest grades cancel each other out when the two samples are compared. Both in my sample and in Sweden, there were students who failed English. Relatively speaking, there were more students in Sweden who failed ($\approx 6\%$) than in my sample (only one student, i.e. $\approx 1\%$).⁷ In sum, then, based on the final grades in English, my sample agreed well with the final grades awarded to all ninth graders in Sweden.

9.5 Summary and conclusion

The main part of the present chapter comprised the results on background variables, motivation, and the students’ views on English. All these results were based on data from my questionnaire. With regard to the background variables – travels abroad, parents’ educational background, the number of books in the home, urban versus rural residency – only “residency” was found to correlate

⁶ Statistics are available from <http://sir.is/skolverket.se> (retrieved May 3, 2009).

⁷ I also checked the final grades in English for students who belonged to the four classes but declined to participate in my study. They all passed English.

with the EE variable as a whole. One important finding was that students living in urban areas had significantly higher amounts of EE per week than students living in rural areas, but on the other hand, for students in rural areas, the correlation between EE and OP was positive and statistically significant ($r_s=.402^{**}$), something which was not the case for students in urban areas. Among the remaining three background variables, “parents’ educational background” correlated with two of the EE activities, namely “reading books” and “reading newspapers and/or magazines”, but not with EE as a whole. However, the correlation between “parents’ educational background” and the two reading activities was only significant for students with two university-educated parents; i.e., one university-educated parent “was not enough” to yield a significant correlation. The other two background variables, “travels abroad” and “the number of books in the home” did not correlate at all with EE. Based on these results and what was previously found on EE and gender in section 7.4.1, I conclude that of the examined background variables, residency may be a factor in explaining students’ EE, but it is nevertheless less important than gender.

The result was different when the four background variables were correlated with OP. Without exception, all background variables correlated with OP. Students who had experiences of traveling abroad, especially if they had traveled outside Europe, had higher OP grades than students who were not as experienced travelers. The parents’ educational background was found to correlate positively and significantly with the OP grade ($r_s=.260^*$). Also the number of books in the students’ homes correlated with OP. Those with more books at home had a higher mean OP grade than those with fewer books. Also, the more books there were in the home, the higher were the students’ OP grades. For the final background variable, residency, I found that the group of students in my sample who lived in urban areas were graded significantly higher on oral proficiency than those in rural areas.

For the first motivational factor that was presented, self-efficacy, the results revealed a positive and significant correlation with EE on sample level ($r_s=.293^*$). When self-efficacy was compared for the five EE subsets, the highest values were found for EE subset 3 and the lowest in EE subset 1. There was an interesting gender difference related to self-efficacy when correlated with EE, very similar to the one found in correlations between EE and OP as well as EE and vocabulary. That is, the correlation between EE and self-efficacy was positive and statistically significant for boys ($r_s=.438^{**}$). However, for girls it was less positive and not statistically significant ($r_s=.107$).

In other words, “the more EE, the higher self-efficacy” seems to hold for boys, whereas it is more difficult to make any claims regarding EE and self-efficacy for girls. With respect to correlations between self-efficacy and OP, on the other hand, there was no major gender difference. Both for boys and girls, there were statistically significant positive correlations between the two variables: without a doubt, self-efficacy was very important for students’ oral proficiency.

The second motivational factor, anxiety related to speaking, had a negligible correlation with the EE variable as a whole. However, there was a statistically significant negative correlation with the EE activity “playing video games” ($r_s = -.225^*$), which means that among those who reported playing video games in English, the more time they had reported, the less anxious they reported feeling when speaking in English. Among the five EE subsets, EE subset 5 had the lowest anxiety value and EE subset 1 the highest. When anxiety was correlated with OP, I found that the two variables correlated negatively and the correlation was statistically significant ($r_s = -.380^{**}$), which means that the more anxious the students reported being, the lower their OP was assessed to be. This result is in line with previous studies on anxiety related to speaking.

In addition to background variables and motivation, the present chapter included results on students’ views on English. A comparison was made between questionnaire data for my sample and data from *Nationella utvärderingen 2003 (NU-03)* (Oscarson & Apelgren, 2005). Overall, the results were very similar, something which strengthens both the validity and reliability of my study. A difference between the two data sets was that, in comparison with the *NU-03*, a much larger proportion of the students in my sample thought that they had learned most of their English *in* school as opposed to *outside* school. This particular difference in beliefs was reflected also in my students’ responses to whether there was too much or too little time for English in school.

The present chapter also included results regarding correlations between EE/OP and the students’ overall grade from grade 9. EE as a whole did not correlate with the students’ overall grade, but the EE activity “reading books” did ($r_s = .359^{**}$), a finding which provides additional evidence of the relationship between reading and academic results. The correlation between the overall grade and OP was obvious ($r_s = .615^{**}$). That is, the higher the overall grade, the higher the OP grade. Finally, when the distribution of the final grades in English on sample level was compared with that of all ninth graders in Sweden,

I found that the distribution was fairly similar in both data sets, which indicates another similarity between the two samples.

In the next chapter, which is the final chapter of the thesis, results from the present chapter and results from the two previous chapters will be brought together and discussed.

10 Summary and concluding remarks

The focus of this thesis has been on Swedish ninth graders' extramural English, oral proficiency, and vocabulary. I have defined extramural English (EE) as linguistic activities in English that learners do or are involved in outside the classroom in their spare time. This final chapter gives a summary of the present study and its results. First I describe the study in section 10.1. I then give an account of the main findings in the following section (10.2), and discuss them in relation to the four theoretical background chapters, i.e., chapters 2–5. In section 10.2, I also integrate evaluations of the methods I used and comments from student interviews. In section 10.3, implications for classroom practice are discussed. In the final section (10.4), concluding remarks are given and suggestions for further research are made.

10.1 Brief description of the study

The main aim of my study was to see whether extramural English has an impact on Swedish ninth graders' oral proficiency and vocabulary. Additional aims were to identify and explicate possible correlations between EE and oral proficiency on the one hand, and EE and vocabulary on the other. The study also aimed to map out students' EE.

To achieve these aims, a longitudinal study was carried out, spanning one school year. The study was based on data collected from Swedish ninth grade English language learners in four school classes at three schools. In total, there were 80 participants: 36 boys and 44 girls. Of these, eight had a first language other than Swedish. The sample was considered representative for the part of Sweden here referred to as Western Svealand (see App. 3). Thus, my results can be generalized to that statistical population. The rate of internal attrition was low (8%); altogether, 74 students completed all parts of the study. Empirical data included a questionnaire, language diaries, recorded speech from speaking tests, raters' assessment of learners' speech, written vocabulary tests, the Swedish national test of English for grade 9, and student interviews. A mixed methods research design was adopted; i.e., both quantitative and qualitative methods were used in a separate and parallel manner and the results were integrated into the interpretation phase. However, it should be pointed out that the lion's share of my data was analyzed using inferential statistics.

Learners' EE was measured with the help of a questionnaire and two language diaries, each covering one week. In the diaries, the learners recorded how much time they had spent on seven given EE activities: "reading books",

“reading newspapers/magazines”, “watching TV”, “watching films”, “surfing the Internet”, “playing video games”, and “listening to music”. There was also an open category where students listed other EE activities they had been involved in. Speech data (\approx 46 hours in total) were collected with the help of five interactional speaking tests, in which learners had been assigned to random dyads on each test occasion. They were evaluated by four external raters, and a total of 1,140 assessment forms were collected. On each test, the student was awarded an overall grade for oral proficiency, called the OP grade, by three raters; in total, 15 OP grades per student were collected. The OP grade ranged from 1 to 6, where 1 was the lowest grade and 6 the highest. The OP grade used in subsequent analyses was thus a mean, based on assessment data from three of the raters for each of the five tests. Learners’ vocabulary was measured with an index variable based on the scores on two vocabulary tests, one which measured productive vocabulary and one which measured receptive vocabulary. For ten of the learners in the sample, additional analyses were made of oral fluency and the use of advanced vocabulary in speech.

10.2 Discussion of main findings

In the present section, the main findings of my study are discussed. I present them in the same order as my research questions were posed in chapter 1. The discussion ends with a summary of the main findings (section 10.2.9).

As was indicated above, most of my data were analyzed statistically. In chapters 7–9, I refrained from discussing the results of various correlation analyses in terms of strength. Here, however, I offer both my interpretation of the results and my conclusions. Furthermore, I will incorporate student comments from the interview data to illustrate some of the key findings (see section 6.2.8).¹

10.2.1 *Extramural English among Swedish ninth graders*

Based on the language diary data, it is possible to conclude that a majority of students spend time on EE. For the students’ total amount of EE, the sample mean was 18.4 hours per week. The standard deviation was large (12.9) and the distribution of student values for EE was skewed to the right. This means that there were fewer students who had high values for EE than there were students who had low values. Some students did not spend much time at all on EE. Thus, comments in line with those mentioned in chapter 1, that all Swedish

¹ Eight students, one boy and one girl from each class, were interviewed at the end of the study (see section 6.2.8).

children engage in English activities in their spare time, are not necessarily true. I found that, on average, boys spend more time than girls on EE activities, close to 21 hours per week as opposed to 16.4 for girls, but this difference was not statistically significant ($p=.136$). Overall, “listening to music” was found to be the EE activity that students spend the most time on, followed by “playing video games”, “watching TV”, “watching films”, “surfing the Internet”, “other activity”, “reading books”, and finally, “reading newspapers/magazines”. The order of popularity was similar to what Forsman (2004) found. Significant differences were found between the genders for two of the EE activities, “playing video games” ($p=.000$) and “surfing the Internet” ($p=.016$), where the results showed that boys spend more time than girls on both. In fact, for the boys, video games and the Internet accounted for approximately 44% of their total EE time, whereas those activities made up only 6% of the girls’ total time. With regard to the four participating school classes, class 3 spent the most time on EE (24.5 hrs/w), followed by classes 4 (18.6 hrs/w), 1 (15.9 hrs/w), and 2 (13.4 hrs/w). The differences were significant ($p=.035$). On the whole, students’ responses to EE-related questions in the questionnaire agreed very well with the language diary data, which strengthens the reliability of my results.

In my study, I used two tools for measuring EE for all students – the language diary and the questionnaire – and both functioned well in the study. The diary provided fairly detailed information about students’ EE activities and made it possible to analyze EE as a whole, as well as separate EE activities. In comparison with Forsman’s (2004) finding, the estimated time spent on EE that the students in my sample reported was much lower than the estimated time reported by the students in her study. This discrepancy is probably due to the fact that the values for EE in my study were based on diary data (supported by questionnaire data) and hers on questionnaire data only. In comparison, the informants in Forsman’s study had to “guess more” in their questionnaires than the students in my study had to do in their diaries. Thus, my interpretation is that the language diary is a more precise tool for measuring EE.

10.2.2 Extramural English and oral proficiency

Students’ oral proficiency was assessed by four raters with the help of materials (instructions and a performance profile scheme) based on Hasselgren (1996b). Interrater reliability was satisfactory or high for the OP grade (see App. 11).

The distribution of the OP grade was normal. The mean OP grade for the whole sample was 3.3. Girls had a higher mean OP grade (3.5) than boys (3.2), but this difference was not statistically significant ($p=.196$). Paired samples

t tests confirmed that all four participating classes improved over time.² Class 3 started at the highest level of oral proficiency of the four participating classes, and also ended the study at a higher level of oral proficiency than the others. Least progression over time was shown by class 1, whereas classes 2 and 4, which were at similar levels at the outset of the study, gradually diverged in the course of the study: class 4 made more improvement than class 2.

Correlation analysis showed a fairly strong positive and statistically significant correlation ($r=.307^{**}$) on the sample level between students' total time spent on EE and their assessed level of oral proficiency, the OP grade. This means that almost 10% of the variance in the OP grades is explained by the relationship between EE and oral proficiency rather than by chance or by some other factor(s). At first, "almost 10%" (9.4%, to be more exact) might not sound high or even relevant, but it definitely is. The correlation between EE and oral proficiency is strong enough not to be overlooked while discussing learners' oral proficiency. Thus, there is a relationship between the two variables. The question is, however, is this a causal relationship? Does EE have, in fact, an effect on oral proficiency? Based on my findings and what is known from previous research (see e.g. Bialystok, 1981; Pickard, 1995; Nunan, 1991), I suggest that it has. I would like to stress, however, that it is, of course, impossible to know this for a fact (cf. the causality dilemma – the chicken or the egg – section 3.1.4).

The relationship between EE and oral proficiency is complex, as indicated by the OP grades for five so-called EE subsets, i.e., the five subsets that were created from the EE variable in order to be used in my analyses. Students in EE subset 1 had the least amount of EE (from nothing to eight hours per week), whereas those in EE subset 5 had the highest amount of EE (from 44 to 58 hours per week). As it turned out, EE subset 1 had the lowest mean OP grade and EE subset 5 the highest, but EE subsets 2 and 3 had in fact higher mean OP grades than EE subset 4. My conclusion is that while the total amount of EE appears to be of importance, results for the EE subsets suggest that the specific types of EE activity that students are engaged in must also be considered when the role of EE in L2 acquisition is discussed. Finally, for students in EE subset 1, there was a strong positive correlation between EE and oral proficiency ($r=.590^{*}$). This finding indicates that for such learners, even slight increases in the amount of time spent on EE activities might have a positive impact on oral proficiency.

² For class 1, $p < .05$; for classes 2-4, $p < .01$.

I used backward linear regression analysis (see section 6.3.6.6) to find out which EE activities were more important than others for oral proficiency. It was possible to establish a ranking list.³ The two EE activities that involved reading in English ended up on top, followed by “listening to music”, “watching TV”, “surfing the Internet”, “playing video games”, and “watching films”. This means that students who read books, newspapers, or magazines in English, and perhaps also listen to music, are more likely to have a high level of oral proficiency than students who do not. This is in line with results from learner self-assessment and retrospection studies, such as Nunan (1991), Cho and Krashen (1994), Pickard (1995), and Arnold (2009). Moreover, the fact that reading was highly important supports the input hypothesis (see chapter 2).

My interviews confirmed the importance of reading and music for oral proficiency. For instance, one girl, who was awarded high OP grades throughout my study, reported that she read many books. In addition, she often listened to music and, when she did not hear or understand the lyrics of a song, she would visit YouTube or a special website for lyrics “to get everything right”.⁴ In contrast, a boy, who received very low OP grades, reported that he neither read in English nor listened to music a great deal. He was not fond of watching TV either. It is possible that the lack of EE activities per se put him at a disadvantage in Tests 1, 3, and 5, since those tests were based on the assumption that learners in general have an interest in the world around them, including for example television and music. My interpretation is that students who regularly read and listen to music are probably better prepared to discuss the variety of topics that are typically used in speaking tests, than those who do not spend that much time on these activities.

Correlation analyses between EE and oral proficiency on class level yielded bewildering results, with negative correlations for classes 3 and 4 and positive ones for classes 1 and 2. The explanation was found when correlations were examined with regard to gender. For boys, there was a strong, positive, and statistically significant correlation between EE and oral proficiency ($r_s=.515^{**}$). That is, the more time the boys had reported for EE, the higher their OP grade was, and vice versa. In fact, as much as 27% of the variance in OP grades among the boys can be explained by the relationship between EE and oral proficiency. A positive correlation was found for girls as well, but it

³ The final open category was not included in the analysis, since it covered several EE activities, which the students listed themselves.

⁴ The website was <http://www.letssingit.com/> (accessed Aug. 23, 2009). This particular student also reported reading books and newspapers in Swedish regularly.

was very weak and not statistically significant ($r=.118$, i.e. only 1% of the variance in girls' OP grades was explained by the relationship between EE and oral proficiency). This suggests that for girls, it did not seem to matter whether they had a low or high amount of EE because the correlation with their assessed level of oral proficiency was negligible. However, since the results for girls lacked statistical significance, they are more difficult to interpret than the results for boys. Nevertheless, the conclusion is that boys seem to be much more sensitive to, or affected by, EE than girls. However, there might be other explanations for the diverse correlations, for instance the fact that boys' and girls' habits are dissimilar when it comes to involvement in extramural English activities (cf. e.g. Sylvén, 2004; Linderöth & Bennerstedt, 2007; Medierådet, 2008).

Fluency is one aspect of oral proficiency which I investigated in my study. I wanted to find out whether a measure of oral fluency (pause length) discriminates between learners at different levels of oral proficiency. Mean intra-utterance pause length beyond one second in Test 5 was measured for ten students: the five students with the lowest OP grades in the sample (the Low OP group) and the five with the highest OP grades (the High OP group). A significant difference was found ($p=.001$) between the groups. In addition, there was a very strong negative correlation ($r=-.902^{**}$) between pause length and OP grade, which means that the longer the pause length, the lower the OP grade. Mean intra-utterance pause length beyond one second proved to be a very useful tool in discriminating between learners at different levels of oral proficiency. My finding confirms previous studies on pauses and fluency (see e.g. Fillmore, 1979; Wood, 2001; Foster & Skehan, 1996).

10.2.3 *Extramural English and vocabulary*

Before discussing the relationship between EE and vocabulary, I will briefly evaluate the two written vocabulary tests used and summarize the scores on each test. First of all, both tests worked very well in ninth grade because they discriminated between learners (cf. Laufer & Nation, 1999; Nation, 2001b). The sample mean on the first test, the Productive Levels Test, was 16.1 (out of 45; STD: 7.7). Boys had a higher mean score (17.6) than girls (14.8), but the difference was not significant ($p=.106$). Mean scores for the four classes ranged from 14.4 (class 2) to 19.5 (class 3); these differences were not significant either ($p=.091$). The boys in classes 1 and 3 did particularly well on the test, whereas the girls in class 1 clearly lagged behind all other groups. On the second test, the Vocabulary Levels Test, the mean score for the whole sample was 60.1 (out

of 90; STD: 13.7). Again, boys scored higher (64.3) than girls (56.7) and on this test, the difference between boys and girls was significant ($p=.017$). Also the differences in mean scores between the classes were significant on this test ($p=.049$). The mean scores ranged from 55.4 (class 2) to 65.8 (class 3). Again, the boys in classes 1 and 3 outperformed the other students in the sample. These findings indicate that the boys in my study had acquired a larger passive vocabulary than the girls. This is similar to what Sylvén (2004) found in her study.

Mean scores for the five EE subsets on the two tests suggest that there is a more straightforward relationship between EE and vocabulary than what was the case for EE and oral proficiency. Students in EE subset 1, i.e., those with the least amount of EE, had the lowest vocabulary scores. With each EE subset, the scores were either the same or better, ending with EE subset 5 having the highest mean scores on both vocabulary tests. Based on the two vocabulary tests, a reliable Vocabulary Index Variable was created. A correlation analysis between EE and this index variable revealed that, in comparison with the correlation between EE and oral proficiency, the correlation between EE and vocabulary was slightly stronger ($r_s=.357^{**}$). Interestingly, a gender difference was again identified. For boys, there was a very strong, positive, and statistically significant correlation between EE and vocabulary ($r_s=.590^{**}$), which means that 35% of the variance in the boys' vocabulary scores can be explained by the relationship between EE and vocabulary rather than by chance or by some other cause(s). In contrast, the correlation between EE and vocabulary was negligible and non-significant for girls ($r_s=.011$). Similar to the case of the correlation between EE and oral proficiency for girls, this result is also difficult to interpret since statistical significance was not established. Nevertheless, my results point in one direction: boys seem to be sensitive to and affected by EE. However, as I said in section 10.2.2, there might other explanations for the diverse correlations for boys and girls, and I will return to the observed gender-related differences in section 10.2.9.

As mentioned above, the two reading activities were found to be of the highest relative importance for the students' oral proficiency. I also examined which EE activities were more important than others for vocabulary and was able to establish a ranking list for them too. This time, "playing video games" and "surfing the Internet" turned out to be the most important EE activities, followed by "watching TV". The two most important EE activities for vocabulary are thus those where boys' and girls' habits are totally different (see section 10.2.1). Consequently, it seems highly likely that boys' high scores on

vocabulary are at least partly due to the type of EE activities they are involved in, since my study shows a strong relationship between EE and vocabulary for boys. Data from the interviews seem to corroborate this interpretation. For example, one boy, who did exceptionally well on the vocabulary tests and who was also awarded above average OP grades, was able to give a detailed description of the development of his English skills. He had gone from playing simple video games and watching English-speaking TV shows with subtitles to interacting almost daily both in speech and writing with native speakers while playing video games; he had several friends in Great Britain, for example, whom he met regularly online. In addition, he enjoyed watching films and preferred English subtitling. When asked about what he learned in school, he claimed that school did not offer much that he did not already know in English, or already was involved in (in English) at home.⁵ However, he enjoyed taking the two vocabulary tests, which he thought measured vocabulary skills very well. When I asked him whether he ever dreamt in English, he could not give an answer, but he said that he often thought in English. He further claimed that on occasion he had even started talking English to Swedish friends before realizing his mistake and switching back to Swedish. This student's level of proficiency was high considering his age and from what he said, it is apparent that he had been extensively involved in EE over the years. This student was among the students with "extreme" scores on the test of receptive vocabulary, whose EE activities were examined (see section 8.2.3). He belonged to the eight students who had the highest scores on the test of receptive vocabulary. This particular group of students reported spending a great deal of time on EE activities other than the seven specified EE activities in the language diary. For instance, they reported such things as interacting online with native speakers in speech and writing, meeting foreigners with whom they spoke English, and singing karaoke in English. Time spent on EE activities such as these may explain why these students performed so well on the test. Most likely, engaging in a wider range of EE activities than most of the other students reflects a genuine interest in English and requires a high level of general proficiency. In any case, no other factor was identified in the present study that might explain why these students scored so well, which is not to say that no other explanations are possible.

My findings on EE and vocabulary provide empirical evidence of the interaction hypothesis (see section 2.1). This is particularly evident in the case

⁵ Swe. "Det är ju inget nytt för mig [i skolan] liksom".

of one EE activity, namely playing video games. Learners who play video games have to rely heavily on their language skills in the target language. Furthermore, they need to pay attention at the level of noticing, and they need to produce target language output, often both orally and in writing. Moreover, since lexical and prosodic repetitions are integral features of video games (cf. Piirainen-Marsh & Tainio, 2009), it means that players are simultaneously involved in activities which are hypothesized to benefit L2 acquisition. In other words, video games provide opportunities for implicit learning; thus, players become learners, even though they might not be aware of it themselves.

In addition to examining oral fluency for ten of the students in my sample, i.e., the Low and the High OP groups, I examined the same ten students' use of advanced vocabulary in speech (Test 5). Polysyllabic words can be considered a measure of advanced vocabulary. My aim was to see whether use of polysyllabic words (here defined as words with three syllables or more) would discriminate between learners. This method, which resembles the one used by Afitskaya (2002), worked well. The results showed differences between the two groups on five word counts, where three involved polysyllabic words. The High OP group outperformed the Low OP group on all five: the total number of tokens, core tokens, polysyllabic tokens, polysyllabic types, and the students' own polysyllabic types (i.e. the polysyllabic types that were not included in and elicited by the test material). The difference was significant in the fifth and final word count ($p=.046$). The fact that I focused on certain words, i.e. those which were polysyllabic, and also included an analysis of their etymology, adds a qualitative dimension to the method of analysis used. In sum, it was concluded that students in the High OP group had acquired a more advanced vocabulary than the students in the Low OP group, since they used more polysyllabic words.

10.2.4 Extramural English and background variables

Four background variables were investigated in relation to EE: travels abroad, parents' educational background, the number of books in the home, and residency (urban versus rural). Of these, only residency was found to be connected with the EE variable. One important finding was that students living in urban areas (N=31) had significantly ($p=.027$) higher amounts of EE (22.4 hrs/w) than students living in rural areas (N=49; 15.9 hrs/w). Students living in urban areas reported spending more time on all but one of the EE activities, namely "playing video games" (rural residency 4.4 hrs/w; urban residency 3.2 hrs/w). While this difference was non-significant ($p > .05$), significant

differences were indeed found for “listening to music” and “reading books” (students in urban settings reported more time; $p < .05$). With regard to the parents’ educational background, one finding was that students who had two university-educated parents ($N=20$) reported spending more time on “reading newspapers/magazines” than students who did not have two university-educated parents ($N=53$). Reading has been shown to be of great importance for oral proficiency; parents’ educational level can be expected to be of relevance to this particular activity, since it relates to Bourdieu’s (1973) concept of cultural capital.

As mentioned previously, one of my aims was to map students’ EE. One conclusion that can be drawn from my investigation of the four background variables is that at least residency has a relationship with EE. However, in comparison with, for example, gender, residency is a peripheral factor in predicting students’ EE.

10.2.5 Oral proficiency and background variables

All four background variables (travels abroad, parents’ educational background, the number of books in the home, and residency) were clearly connected with the students’ level of oral proficiency. For instance, the students who had traveled abroad, especially if they had traveled outside Europe, had higher OP grades than the students who were not as experienced travelers. Furthermore, there was a relationship between the educational level of the parents and the students’ OP grades in school; i.e., the students with two university-educated parents had higher OP grades than the others, even though significance was not found ($p=.390$). With regard to the number of books in the home, this variable was also related to oral proficiency: students who reported having a large number of books at home had a higher mean OP grade than those with fewer books (cf. Öquist & Wikström, 2006). In fact, the more books there were in the home, the higher the students’ OP grades. Finally, regarding residency, I found that the group of students who lived in an urban setting ($N=29$) were graded significantly higher ($p=.002$) on oral proficiency than those in rural areas ($N=45$) (cf. Björnsson, 2005). In sum, the background variables that I examined, for which an alternative label could have been “socioeconomic variables”, were all clearly related to oral proficiency. This is in stark contrast to what was found for the same variables in relation to EE, where only a peripheral connection was found, and only for one variable (residency).

10.2.6 *Extramural English and motivation*

Two motivational factors were investigated in relation to EE, namely self-efficacy and anxiety related to speaking. Self-efficacy (i.e., people's judgment of their own ability to carry out specific tasks) correlated positively with EE ($r_s=.293^*$) on the sample level. However, a more interesting finding, yet again, was found for gender. Similarly to what was seen in correlations between EE and oral proficiency as well as between EE and vocabulary, there was a strong positive correlation between EE and self-efficacy for boys ($r_s=.438^{**}$). However, for girls the same correlation was weak and not statistically significant ($r_s=.107$). In other words, "the more EE, the higher the self-efficacy" (and vice versa) seems to hold for boys, whereas it is difficult to make such a claim for girls because of the lack of statistical significance. When I compared the values for self-efficacy for the five EE subsets, EE subset 1 had the lowest and subset 3 the highest. One possible interpretation of the low value for EE subset 1 relative to the other subsets is that spending time on extramural activities was perhaps not very appealing to those students, since they apparently considered their own ability to manage such activities to be low. In short, it is possible that self-efficacy determines involvement in EE, i.e., that self-efficacy affects EE.

The second motivational factor that I examined, anxiety related to speaking, had only a negligible correlation with the EE variable. However, there was a weak negative correlation with the EE activity "playing video games", which was statistically significant ($r_s= -.225^*$). This means that among those who played video games in English, the degree of anxiety decreased with an increased amount of time played. This finding agrees with interview data as well as diary data. For instance, among the five EE subsets, EE subset 5 (where the amount of time reported for playing video games was high) had the lowest value for anxiety, and EE subset 1 (where the amount of time reported for playing video games was very low) had the highest. Thus, even though findings are not overwhelmingly convincing on the matter, they suggest that EE reduces learners' level of anxiety related to speaking, in particular if it involves frequent video game playing.

10.2.7 *Oral proficiency and motivation*

Examinations of oral proficiency in relation to the two motivational factors indicated that self-efficacy was extremely important for students' oral proficiency. The values for self-efficacy increased with an increased level of oral proficiency, so that the students who were awarded the lowest OP grades had the lowest value for self-efficacy, and those who were awarded the highest OP

grades had the highest value for self-efficacy ($p=.000$). This finding is in line with Schunk et al. (2008): self-efficacy is a good predictor of subsequent grades.

The connection between level of self-efficacy and level of oral proficiency was confirmed in the interviews. For instance, a boy remembered how he was put in a group of low achievers in middle school, i.e., the group for students who were considered to have learning problems in English. It was very interesting to listen to his recollections regarding homework and vocabulary acquisition. He said that he had really had a hard time understanding why he and the other students in his group had only ten words for homework when their classmates, in the regular group, had twenty: “It was wrong. We should have had more words than the others because we were the bad ones.” Even though he was aware of his own learning difficulties, he could not see how doing less could be good for learning, and his self-efficacy was clearly affected in a negative way. His self-efficacy had been lowered when he was very young and, unfortunately, remained low.

Correlations between self-efficacy and oral proficiency were strongly positive for both boys ($r_s=.728^{**}$) and girls ($r_s=.588^{**}$), which means that the higher students’ self-efficacy was, the higher their OP grade (and vice versa).

Not surprisingly, the student levels of anxiety and oral proficiency were also closely connected. When the variable “anxiety related to speaking” was correlated with the OP grade, I found a clearly negative correlation ($r_s=-.380^{**}$), which means that the more anxious the students reported being, the lower their oral proficiency was assessed to be. This finding is in line with previous studies on anxiety related to speaking (cf. Matsuda & Gobel, 2004; Pappamihel, 2002). It should be noted that there are, of course, other factors that might affect oral proficiency as well, for example aptitude (see chapter 2).

10.2.8 Students’ views on English

In order to investigate students’ views on English, a comparison was made between questionnaire data for my sample and data from the *National Evaluation of 2003 (NU-03)* (Oscarson & Apelgren, 2005). Overall, the results were very similar, something which strengthens both the validity and the reliability of my study. In general, students like English and view it as an important subject in school. Only one important difference between the two data sets was identified. It had to do with the students’ beliefs about where they had learned most of their English, i.e., in school or outside school. In comparison with the students in the *NU-03* sample, a much larger proportion of the students in my sample thought that they had learned most of their English in school.

10.2.9 *Summary of main findings*

It is now time to summarize the main findings of my study. In doing so, I will highlight the results which I find most important. First of all, it is possible to make an argument based on empirical evidence for the students in my sample, i.e. Swedish ninth graders representing Western Svealand, that there is a relationship between EE and oral proficiency on the one hand, and EE and vocabulary on the other. In other words, the total amount of time which these adolescents spend on various EE activities correlate with their level of oral proficiency and the size of their vocabulary. However, when compared, these correlations are of a different character. For instance, the relationship between EE and vocabulary is much more straightforward than the relationship between EE and oral proficiency. Moreover, the correlation between EE and vocabulary is stronger ($r_s=.357^{**}$) than the correlation between EE and oral proficiency ($r_s=.307^{**}$). The EE variable alone accounts for approximately 13% of the variation in students' vocabulary scores, whereas it accounts for approximately 9% of the variation in their OP grades. Based on my results and findings from previous research (Forsman, 2004; Lamb, 2004; Nunan, 1991; Pearson, 2004; Pickard, 1995; Piirainen-Marsh & Tainio, 2009; Sylvén, 2004), I conclude that students' EE has an effect on both oral proficiency and vocabulary, though, the causal relationship is clearer between EE and vocabulary than it is between EE and oral proficiency.

In addition to the relationship between EE and OP/vocabulary at the sample level, there is also a correlation between EE and self-efficacy. Although this correlation may not be as evident as the others, it is, nevertheless, statistically significant ($r_s=.293^*$) and therefore worth paying attention to. One conclusion drawn was that it is possible that self-efficacy determines involvement in EE, i.e., that self-efficacy affects EE. At least this seemed to be the case for the students in EE subset 1 who had both the lowest amount of EE and the lowest value for self-efficacy in the study. Apparently, they considered their own ability to manage EE activities to be low. Nevertheless, a very strong positive and statistically significant correlation between EE and oral proficiency ($r_s=.590^*$) was found for these learners. Thus, there is convincing empirical evidence from my study to claim that it is crucial that learners such as these increase the amount of time they spend on EE activities, if only very little, since even a small increase may make an important difference.

Self-efficacy was one of two motivational factors that I investigated. I also examined four socioeconomic background variables. Here, I would like to emphasize that whereas students' OP grades were clearly connected with all the

socioeconomic variables examined in my study, this was *not* the case for the amount and type of EE. Extramural English was only connected with one background variable, namely residency, and the correlation between EE and residency was less salient and less important than the one for gender (see below). My conclusion is that EE is an independent variable. This is a key finding in my study because it pinpoints EE as a possible path to progress in English for all learners, regardless of their socioeconomic background.

Another key finding has to do with gender. An interesting pattern emerged in several of the correlation analyses between EE and various other variables. The correlations turned out to be strongly positive and statistically significant for the boys, whereas the same correlations were very weak or negligible for the girls. In addition, the correlations were always non-significant for the girls. The lack of statistical significance regarding the girls' correlations makes their results more difficult to interpret than the results for boys. Nevertheless, the results point in one direction: boys appear more sensitive to EE than girls. For example, as much as 35% of the variation in the boys' vocabulary scores could be accounted for by the EE variable, as well as 27% of the variation in their OP grades. Similarly, EE correlated strongly with self-efficacy for the boys ($r_s=.438^{**}$), but nothing similar could be found for the girls. Thus, in several places in the present study, I have said that "boys seem to be more sensitive to or affected by EE than girls". However, sensitivity, which suggests an innate ability, is not the explanation for the observed gender difference, even though I interpreted it in that way at first. With all results at hand, having looked at EE from several perspectives, my study clearly shows that *to a large extent, boys and girls engage in totally different types of EE activities*, and this is a more reasonable explanation why we arrive at contradictory results in correlation analyses involving EE than that boys and girls are inherently different regarding sensitivity towards EE. Thus, my conclusion is that the type of EE activities that learners are involved in is very important in discussions of the potential impact of EE on learners' oral proficiency and vocabulary.

With regard to the EE activities investigated in this study, I have shown that those which require learners to be active or productive are more important for learners' oral proficiency and vocabulary than activities which allow learners to remain passive. My findings provide empirical evidence in support of current second language acquisition theory, such as the interaction hypothesis and the importance of noticing in L2 acquisition. Thus, playing video games, surfing the Internet, reading books, and reading newspapers/magazines are more important for learners' English as measured in this study, than listening to

music, watching TV, or watching films. When learners play video games, surf the Internet, or read something in English, they need to rely heavily on their language skills, because otherwise these activities become pointless. Regarding videogames, it should be added that players are required to produce output, something which is also considered crucial in L2 acquisition (cf. Swain, 1995). When the time that boys and girls spent on various EE activities was compared, boys turned out to spend significantly more time on the EE activities which were shown to matter most for proficiency. *Why* boys and girls engage in different EE activities is a totally different question, beyond the scope of the present thesis to answer. However, it would be a suitable topic for future studies.

To conclude, my main findings may be condensed into six points:

- The total amount of time which ninth graders spend on EE correlates positively and significantly with both (a) their level of oral proficiency and (b) the size of their vocabulary, where the correlation between EE and vocabulary is stronger and more straightforward than the correlation between EE and oral proficiency. This means that students' EE has an impact on both oral proficiency and vocabulary, but the causal relationship is more salient regarding vocabulary.
- For students with a low amount of EE, even a small increase in time spent on EE activities may make an important difference regarding their level of oral proficiency.
- EE activities which require learners to be active/productive and to rely on their language skills (playing video games, surfing the Internet, reading books, reading newspapers/magazines) have a greater impact on learners' oral proficiency and vocabulary than EE activities where learners can remain fairly passive/receptive (listening to music, watching TV, watching films).
- Boys spend significantly more time on active/productive EE activities than girls and, therefore, EE has a greater impact on boys' oral proficiency and vocabulary, than it has on girls'.
- Oral proficiency is clearly connected with all four socioeconomic background variables investigated in this study, whereas EE is not.
- EE is an independent variable, valuable in its own right; EE is a possible path to progress in English for any learner, regardless of his or her socioeconomic background.

10.3 Implications for classroom practice

Since extramural English turned out to be a factor in L2 acquisition, the results of my study have implications for classroom practice. My study has shown that most (but not all) students spend time on EE activities. It is important that teachers become aware of the fact that EE activities which demand more active participation on the part of the learner (video games, the Internet, reading) are more beneficial to L2 acquisition than EE activities where learners usually remain fairly passive (music, TV, film). By acknowledging this, teachers may supervise individual students more easily and effectively, based on each student's interests and needs.

Likewise, it is important that teachers learn about boys' and girls' different EE habits, namely, that boys spend significantly more time on active/productive EE activities than girls do, and that they generally benefit from doing so. By contrast, such activities, in particular playing interactive video games, are generally unexplored by girls. From an English teacher's perspective, this is a pity since we can assume that girls, just like boys, would most likely benefit from participating in that type of EE activity as well.

Teachers may play a crucial role in motivating students to engage in those EE activities which are more demanding than others, for example playing video games in English, using the Internet in English, and reading various types of texts in English, because activities such as these may be perceived as less accessible by many learners. In sum, since EE is a potential path to progress in English for any learner, learning more about their own students' EE habits would facilitate teachers' daily work, for instance regarding what type of homework or task would suit a particular student. My study has shown that learners' EE varies greatly, both in amount and type. Thus, in order for teachers to optimize learning conditions for each individual student, it is necessary that they have an idea about each learner's EE.

Although my study has shown that EE activities such as listening to music and watching TV or films were not as important for proficiency as others, they might still be useful in the teaching (and learning) of English. Using music as well as TV programs or films in the classroom will most likely enhance student motivation. However, based on my results, the use of music, TV, and films in teaching should be combined with tasks that require students to interact and to produce output. Results from previous studies on lexical coverage in films (Webb & Rodgers, 2009a) together with the fact that "watching films" was ranked the lowest for both oral proficiency and vocabulary in my study (see Tables 7.12 and 8.7), indicate that the average ninth grader needs assistance in

order to learn English from films. For instance, if the learning objective is English vocabulary acquisition, my opinion is that it is better to study a short clip from a film (or TV program) and work intensively with that, than to watch the whole film (or TV program).

Finally, my study showed a low rate of both external and internal attrition, i.e., many students were interested in participating to begin with, and only a few dropped out during the course of the study. In a sense, by participating in my study, students willingly accepted to be involved in yet another EE activity. This alone testifies to the fact that extramural English matters to adolescents.

10.4 Concluding remarks and suggestions for further research

Second language acquisition is idiosyncratic and there are several variables which influence learner progress in learning the target language, in this case English. The present study has shown that extramural English is one such variable and that it can be a powerful one. I will here use one example from my study (“rural boys”) to illustrate this.

Previous research has shown that, generally, boys in rural areas have lower achievement levels in school than other groups of learners (see section 9.1.4). Other studies have revealed that English is the only subject in Swedish schools where boys seem to be catching up with girls (see section 2.4). A possible explanation for that might be the positive impact EE has on boys’ proficiency. In view of this, it was particularly interesting to see that a small group of boys who lived in rural areas excelled on the vocabulary tests in my study. A possible explanation is the positive impact of EE on these boys’ vocabulary acquisition; some boys in rural areas reported a high amount of time spent on “playing video games” (cf. section 9.1.4). I am not suggesting that EE was the only reason why these boys had very high scores, but it probably helps explain the result.

The present study has revealed that EE is certainly a multifaceted variable. It includes a whole range of linguistic activities and, most likely, will include an even wider range of activities in the future. I planned my study in the year 2005 and launched it in 2006. Obviously, extramural English has changed considerably since then. Therefore, a language diary in 2009 aimed at ninth graders would have to comprise more activities than the ones I used. For example, the EE activity called “surfing the Internet” would probably need to be supplemented with several new categories (all which require a computer and access to the Internet), such as, perhaps, “chatting in writing”, “chatting orally”, “watching YouTube”, “publishing on YouTube”, “writing a blog”, “writing

fanfiction”, and “using Facebook (or the equivalent)”. Likewise, it would probably be wise to add more reading categories, such as “reading blogs” and “reading fanfiction”. Furthermore, it is possible that an online version of the diary would be appreciated, at least by some learners. Therefore, if someone were to replicate my study, it would be necessary to revise the tools for measuring EE accordingly.

Suggestions for future research in the field of SLA might, of course, include further studies on extramural English, but using a larger sample than I did, preferably one which also includes participants from large cities. A study based on EE data from such a sample would yield results which could be generalized to a larger statistical population than the one in the present study and, thus, provide relevant information both for SLA theory and classroom practice. Furthermore, it would be interesting to map out EE for young learners as well, and study how EE develops over time for them. For instance, is the profile for specific learners and their EE activities more or less constant, or does it change with age? It would be particularly interesting to see whether gender differences can already be discerned at an early age.

Regarding my own data, it would be interesting to compare the socioeconomic background variables and motivational factors with students’ results on vocabulary, since that analysis could not be made within the scope of the present study. It would be valuable to know whether the Vocabulary Index Variable is independent, like the EE variable, or if it is clearly connected with background variables in the way that oral proficiency turned out to be. In addition, it would be valuable to conduct a more comprehensive investigation of the students’ use of polysyllabic words in speech, which was limited to only ten students and one test in the present study. In my future research, I would like to examine the use of such words by all participants, and to correlate the results with the EE variable and the separate EE activities. The outcome of such a study would deepen our understanding of the relationship between EE and vocabulary.

To conclude, my study has shown that among the ninth graders in my sample, activities in English outside of school have an impact both on their level of oral proficiency and on the size of their vocabulary. In short, the study shows that extramural English matters.

Summary in Swedish – Sammanfattning på svenska

Denna doktorsavhandling handlar om så kallad extramural engelska (EE). Ordet *extramural* kommer från latin och betyder 'utanför murarna', vilket i detta sammanhang syftar på klassrummets väggar. Med *extramural engelska* menas alltså den engelska som elever möter utanför klassrummet på sin fritid. Studiens huvudsyfte är att besvara följande forskningsfråga: "Har extramural engelska någon effekt på elevers muntliga färdighet och vokabulär?"

För att kunna besvara frågan har jag genomfört en longitudinell studie under ett läsår (2006-2007) bland elever i skolår 9. Totalt deltog 80 elever, varav 36 var pojkar och 44 flickor. De flesta elever hade svenska som modersmål, men åtta hade ett annat modersmål än svenska. De gick i fyra klasser vid tre skolor och urvalet anses som representativt för Västra Svealand. Mina resultat är alltså generaliserbara till den statistiska populationen. Såväl det externa bortfallet (10%) som det interna (8%) var lågt, vilket innebär att av de 89 elever som tillfrågades om de ville delta i studien svarade 80 "ja", medan sex av olika anledningar avbröt sitt deltagande under studiens gång.

Empirin omfattar enkätmaterial, språkdagböcker, inspelat tal från muntliga prov, externa bedömares betyg på deltagarnas muntliga färdighet, det nationella provet i engelska för skolår 9, samt elevintervjuer. Både kvantitativ och kvalitativ metod användes i analysarbetet, men tonvikten låg på den kvantitativa delen. Detta innebär att insamlade data huvudsakligen analyserades statistiskt. Mjukvaran SPSS användes för ändamålet.

Elevernas EE mättes med hjälp av en enkät samt två språkdagböcker, där varje språkdagbok i tid omfattade en vecka. I språkdagböckerna antecknade eleverna hur mycket tid de hade spenderat på sju givna aktiviteter ("läst bok", "läst tidning", "sett TV-program", "sett film (på bio, TV, video, DVD, dator, etc.)", "surfad på nätet", "TV-/Dataspel", "lyssnat på musik" samt en avslutande öppen kategori, "annat").

Muntliga data (cirka 46 timmar totalt) samlades in med hjälp av fem interaktiva muntliga tester (totalt 199 stycken). Eleverna var indelade i slumpmässiga par vid varje testtillfälle. De första fyra testerna spelades in med videokamera, samt med mp-3 spelare (ljud) som back-up. Vid det sista muntliga testet, som utgjordes av den muntliga delen av det obligatoriska nationella provet i engelska, användes enbart ljudinspelning. Elevernas muntliga produktion bedömdes och betygssattes av fyra externa bedömare. Alla var mycket erfarna engelsklärare; två hade engelska som modersmål och två svenska. Totalt samlades 1140 bedömningsformulär in. För varje test erhö

eleven ett medelbetyg för sin muntliga färdighet (*the oral proficiency grade/the OP grade*), vilket baserades på betygsunderlag från tre av bedömarna. Totalt samlades alltså 15 betyg (3 betyg \times 5 tester) på muntlig färdighet in för varje elev. Reliabiliteten var tillfredsställande eller hög mellan bedömarna (från $r=.451^{**}$ till $r=.703^{**}$). Baserat på de 15 betygsvärden som samlats in för varje elev kunde ett reliabelt medelbetyg för muntlig färdighet beräknas för varje individ. Medelbetyget var mellan 1 och 6, där 1 var det lägsta man kunde få och 6 det högsta.

Storleken på elevernas vokabulär mättes med hjälp av två skriftliga vokabulärprov, där det första mätte produktivt (aktivt) ordförråd och det andra receptivt (passivt). Baserat på dessa prov kunde ett reliabelt index för elevernas vokabulär skapas. För ett urval av tio elever gjordes dessutom analyser av muntligt talflöde (*oral fluency*) och användningen av avancerade ord i tal.

I det följande sammanfattar jag studiens viktigaste resultat. För det första är det möjligt att konstatera att det dels förelåg statistiskt signifikanta positiva samband både mellan EE och muntlig färdighet, och mellan EE och vokabulär. Med andra ord, den totala tid som eleverna spenderade på olika EE-aktiviteter korrelerade positivt med nivån på deras muntliga färdighet och med storleken på deras ordförråd. De båda korrelationerna (eller sambanden) såg dock olika ut i jämförelse. Sambandet mellan EE och vokabulär var till exempel mycket rakare och tydligare än det mellan EE och muntlig färdighet. Dessutom var korrelationen mellan EE och vokabulär starkare ($r=.357^{**}$) än korrelationen mellan EE och muntlig färdighet ($r=.307^{**}$). Variabeln EE förklarar 13% av variationen i elevernas resultat på vokabulär samt 9% av variationen i deras betyg på muntlig färdighet. Baserat på mina resultat och slutsatser från tidigare forskning (Forsman, 2004; Lamb, 2004; Nunan, 1991; Pearson, 2004; Pickard, 1995; Piirainen-Marsh & Tainio, 2009; Sylvén, 2004) drar jag slutsatsen att elevernas EE har effekt både på deras muntliga färdighet och på deras vokabulär. Orsakssambandet är dock tydligare mellan EE och vokabulär än det är mellan EE och muntlig färdighet.

Förutom sambanden mellan EE och muntlig färdighet respektive vokabulär fanns det en korrelation mellan EE och elevernas bedömning av sin egen förmåga att klara av uppgifter på och i engelska. En sådan typ av bedömning, som alltså har att göra med tron på sig själv, faller under begreppet *motivation* och kallas för *self-efficacy* på engelska. Detta samband var visserligen inte lika tydligt som de som beskrevs ovan, men det var trots allt statistiskt signifikant ($r=.293^{*}$) och därför värt att uppmärksamma. För eleverna som rapporterat en liten mängd EE i sina språkdagböcker (0-8 timmar per vecka)

fanns det indicier på att *self-efficacy* påverkade den tid som de ägnade åt EE. För dessa elever fanns det även en mycket tydlig och stark positiv korrelation mellan EE och muntlig färdighet ($r=.590^{**}$). Detta innebär att även en liten ökning i tid som spenderas på EE kan vara betydelsefull för just dessa elevers resultat.

Jag undersökte fyra bakgrundsvariabler för att se om de hade något samband dels med elevers EE, dels med elevers betyg på muntlig färdighet i engelska. Samtliga bakgrundsvariabler var av socioekonomisk natur, det vill säga faktorer som vanligen är betydelsefulla för studieresultat: (1) elevens erfarenhet av resor utomlands, (2) föräldrarnas utbildningsnivå, (3) hemmets så kallade kulturella kapital (baserat på enkätfråga 30, om antalet böcker i hemmet), samt (4) var eleven var bosatt, på landsbygden eller i stadsmiljö. Det visade sig att det fanns en mycket tydlig koppling mellan samtliga fyra variabler och elevernas betyg på muntlig färdighet, dock inte alls när variablerna ställdes mot elevernas totala mängd av EE eller typ av EE. Endast för en av bakgrundsvariablerna (var eleven bodde) kunde en koppling skönjas med EE, men den var svag och mindre viktig i jämförelse med exempelvis kön (se nedan) som förklaring till varför elevernas EE-vanor varierade starkt. Här bör tilläggas att det tyvärr ej fanns utrymme i denna studie att undersöka samma bakgrundsvariabler med avseende på vokabulär. Slutsatsen för denna del av studien var att EE fungerar som en självständig variabel. Det vill säga, EE är inte knuten till bakgrundsvariabler som brukar ha betydelse för elevers skolresultat. Detta är mycket viktig ny kunskap eftersom resultatet faktiskt visar att EE utgör en möjlighet för alla elever, oavsett deras socioekonomiska bakgrund, att göra framsteg i engelska.

En annan viktig slutsats gäller kön. Ett intressant mönster uppstod i korrelationsanalyser mellan EE och muntlig färdighet, vokabulär respektive *self-efficacy*. Korrelationerna visade sig vara starkt positiva och statistiskt signifikanta för pojkar, medan samma korrelationer var mycket svaga eller rent av försumbara för flickor. Bristen på statistisk signifikans avseende flickornas korrelationer gjorde dock deras resultat något svårare att tolka än resultaten för pojkar. Trots detta pekade samtliga resultat med avseende på kön i en riktning, nämligen att pojkar skulle vara mer känsliga eller mottagliga för EE än flickor. Till exempel så kunde så mycket som 35% av variationen i pojkarnas resultat på vokabulär förklaras av EE-variabeln. För flickorna var motsvarande samband försumbart. Vad gäller muntlig färdighet så förklarade EE-variabeln 27% av variationen i pojkarnas betyg, men endast 1% i flickornas. Kort sagt, sambanden mellan EE och vokabulär/muntlig färdighet för pojkar var mycket

tydliga och berodde ej på slumpen. Likaså korrelerade EE starkt med *self-efficacy* för pojkarna, medan inget liknande kunde ses för flickorna. Men, att pojkar skulle ha en högre grad av ”känslighet” för EE, eller påverkas lättare av detsamma jämfört med flickor, föreföll osannolikt. Det skulle ju implicera att det föreligger en medfödd skillnad mellan pojkar och flickor. Även om jag onekligen tolkade resultaten så till en början visade det sig i slutändan, när alla resultat för EE låg på bordet och jag hade undersökt EE ur flera perspektiv, att *pojkar och flickor i mycket stor utsträckning ägnar sig åt helt olika typer av EE-aktiviteter*. Deras fritidsengelska skiljer sig alltså avsevärt åt; skillnaderna var statistiskt signifikanta. Detta är en mer rimlig förklaring till de motsägelsefulla resultaten av korrelationsanalyserna – inte att pojkar och flickor i sig skulle vara annorlunda när det gäller ”känslighet” för EE. Slutsatsen är att den typ av EE-aktiviteter som ungdomar ägnar sig åt är en mycket viktig faktor när man diskuterar potentiella effekter av EE på elevers muntliga färdighet och vokabulär.

Mina resultat visade att den typ av EE-aktiviteter som kräver att eleven är aktiv och produktiv samt tvingar eleven att lita på sin egen språkliga färdighet i engelska (dataspelande, användningen av Internet, läsning – allt på engelska) är viktigare för elevens muntliga färdighet och vokabulär än EE-aktiviteter som kan tillåta eleven att vara mer passiv och inte lika hög grad behöva stödja sig på sin engelska (se på tv eller film, samt lyssna på musik). Min studie ger alltså empiriskt stöd till moderna språkinlärningsteorier, såsom *the interaction hypothesis* (vilken betonar interaktionens betydelse för inlärning) och teorin om *noticing* (vilken betonar att en viss grad av medvetenhet för språkliga fenomen är en förutsättning för inlärning). Det är alltså därför som dataspelande, surfning på Internet och läsning av böcker respektive tidningar är viktigare för elevernas engelska, så som den mätts i denna studie, än de andra aktiviteterna, det vill säga att lyssna på musik samt att se på tv eller film. Tilläggas bör att när man spelar dataspel måste man ofta förmedla sig både muntligt och skriftligt på engelska, något som anses mycket viktigt för språkinlärning (*the output hypothesis*). När jag jämförde den tid som pojkar respektive flickor ägnade åt olika typer av EE-aktiviteter visade det sig att pojkarna tillbringade signifikant mycket mer tid på de produktiva aktiviteterna jämfört med vad flickorna gjorde. Att utreda orsaken till denna genusrelaterade skillnad var dock inte möjligt inom ramen för denna studie, men det skulle vara ett intressant framtida forskningsprojekt.

Avslutningsvis sammanfattar jag studiens viktigaste resultat i sex punkter:

- Den totala tid som elever i nian ägnar åt EE korrelerar positivt och signifikant både med (a) nivån på deras muntliga färdighet samt (b) storleken på deras vokabulär, där korrelationen mellan EE och vokabulär kan beskrivas som starkare och rakare än korrelationen mellan EE och muntlig färdighet. Detta betyder att elevernas EE har effekt både på deras muntliga färdighet och på deras vokabulär; orsakssambandet är dock tydligare för vokabulären.
- För gruppen elever som vanligen ägnar liten tid åt EE kan redan små ökningar i tid vara av avgörande betydelse för deras muntliga färdighet.
- EE-aktiviteter som kräver att elever är aktiva/produktiva samtidigt som de måste lita till sina egna färdigheter i engelska är viktigare för den muntliga färdigheten och vokabulären än aktiviteter där eleven kan förhålla sig mer passiv.
- Pojkar ägnar betydligt mer tid åt aktiva/produktiva EE-aktiviteter jämfört med vad flickor gör; EE har därför större effekt på pojkarnas muntliga färdighet och vokabulär än den har på flickornas.
- Muntlig färdighet är tydligt knuten till samtliga fyra socioekonomiska bakgrundsvariabler som undersöktes i denna studie. Detsamma gäller dock inte för EE.
- EE är en självständig variabel som erbjuder möjligheter till utveckling i engelska för alla elever oavsett socioekonomisk bakgrund.

References

- Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. In P. Boogards & B. Laufer (eds.), *Vocabulary in a second language* (pp. 39-49). Amsterdam: John Benjamins.
- Afitskaya, N. (2002). Assessing vocabulary richness: LS* and LS**. *University of the West of England, Faculty of Languages and European Studies Research and Consultancy Bulletin (Summer)*, 22, 4-9.
- Altman, R. (1997). Oral production of vocabulary: A case study. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition* (pp. 69-97). Cambridge: Cambridge University Press.
- Arnaud, P. J. L., & Savignon, S. J. (1997). Rare words, complex lexical units and the advanced learner. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition* (pp. 157-173). Cambridge: Cambridge University Press.
- Arnold, N. (2009). Online extensive reading for advanced foreign language learners: An evaluation study. *Foreign Language Annals*, 42(2), 340-366.
- Aron, A., Aron, E. N., & Coups, E. J. (2005). *Statistics for the behavioral and social sciences: A brief course* (3 ed.). London: Prentice Hall International.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bardovi-Harlig, K. (2006). Interlanguage development: Main routes and individual paths. *AILA Review*, 19, 69-82.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow: Pearson Education.
- Bergström, I. (2001). Ordförståelse, ordinlärning och ordanvändning. In R. Ferm & P. Malmberg (eds.), *Språkboken* (pp. 51-59). Stockholm: Myndigheten för skolutveckling.
- Berns, M. (2007). The presence of English: Sociocultural, acquisitional, and media dimensions. In M. Berns, K. de Bot & U. Hasebrink (eds.), *In the presence of English: Media and European youth* (pp. 1-14). New York, NY: Springer.
- Berns, M., de Bot, K., & Hasebrink, U. (eds.). (2007a). *In the presence of English: Media and European youth*. New York, NY: Springer.
- Berns, M., Claes, M.-T., de Bot, K., Evers, R., Hasebrink, U., Huibregtse, I., Truchot, C., & van der Wijst, P. (2007b). English in Europe. In M. Berns, K. de Bot & U. Hasebrink (eds.), *In the presence of English: Media and European youth* (pp. 15-42). New York, NY: Springer.
- Bialystok, E. (1981). The role of conscious strategies in second language proficiency. *The Modern Language Journal*, 65(1), 24-35.
- Bialystok, E. (1991). Achieving proficiency in a second language: a processing description. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood Smith & M. Swain (eds.), *Foreign/Second language pedagogy research: A*

- commemorative volume for Claus Færev* (pp. 63-78). Clevedon: Multilingual Matters.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Björnsson, M. (2005). *Kön och skolframgång: Tolkeningar och perspektiv*. Stockholm: Myndigheten för skolutveckling.
- Black, R. W. (2009). English-language learners, fan communities, and 21st-century skills. *Journal of Adolescent & Adult Literacy*, 52(8), 688-697.
- Bourdieu, P. (1973). Cultural reproduction and social reproduction. In R. Brown (ed.), *Knowledge, education, and cultural change: Papers in the sociology of education* (pp. 71-112). London: Tavistock Publications.
- Bowers, K. S. (1984). On being unconsciously influenced and informed. In K. S. Bowers & D. Meichenbaum (eds.), *The unconscious reconsidered* (pp. 227-272). New York, NY: John Wiley & Sons.
- Brock, C. A. (1986). The effects of referential questions on ESL classroom discourse. *TESOL Quarterly*, 20(1), 47-59.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1-25.
- Brown, D. H. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Longman.
- Buckwalter, P. (2001). Repair sequences in Spanish L2 dyadic discourse: A descriptive study. *The Modern Language Journal*, 85(3), 380-395.
- Buss, A. H. (1986). Two kinds of shyness. In R. Schwarzer (ed.), *Self-related cognitions in anxiety and motivation* (pp. 65-75). London: Lawrence Erlbaum.
- Byström, J. (1973). *Grundkurs i statistik*. Stockholm: Natur och kultur.
- Carr, J., & Pauwels, A. (2006). *Boys and foreign language learning: Real boys don't do languages*. Basingstoke: Palgrave Macmillan.
- Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (ed.), *Individual differences and universals in language learning aptitude* (pp. 83-118). Rowley, MA: Newbury House.
- Cho, K., & Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, 37(8), 662-667.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton & Co.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: M.I.T. Press.
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80(2), 183-198.
- Cohen, A. D. (1991). Strategies in second language learning: Insights from research. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood Smith & M. Swain (eds.), *Foreign/Second language pedagogy research: A commemorative volume for Claus Færev* (pp. 107-119). Clevedon, Philadelphia: Multilingual Matters.

- Cook, V. (2005). Basing teaching on the L2 user. In E. Llurda (ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 47-61). New York, NY: Springer.
- Coxhead, A. (1998). *An Academic Word List. Occasional Publication Number 18, LALS*. Wellington, New Zealand: Victoria University of Wellington.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, D. (2003). *The Cambridge encyclopedia of the English language* (2 ed.). Cambridge: Cambridge University Press.
- d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28(3), 227-244.
- Dahlström, A. (2007). *Könet sitter i hjärnan*. Göteborg: Corpus Gullers.
- Daller, H., Milton, J., & Treffers-Daller, J. (eds.). (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Davies, J. (2003). Expressions of gender: An analysis of pupils' gendered discourse styles in small group classroom discussions. *Discourse & Society*, 14(2), 115-132.
- Dewaele, J.-M. (2004). Individual differences in the use of colloquial vocabulary: The effects of sociobiographical and psychological factors. In P. Boogards & B. Laufer (eds.), *Vocabulary in a second language* (pp. 127-153). Amsterdam: John Benjamins.
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Longman.
- Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, 19, 42-68.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dörnyei, Z. (2008). Individual differences: Interplay of learner characteristics and learning environment. Podcast retrieved from <http://www.wiley.com/bw/podcast/lang.asp>, August 14, 2009.
- Ejvegård, R. (2003). *Vetenskaplig metod*. Lund: Studentlitteratur.
- Ellis, N. (1994). Implicit and explicit language learning - An overview. In N. Ellis (ed.), *Implicit and explicit learning of languages* (pp. 1-31). London: Academic Press.
- Ellis, R. (1994a). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (1994b). A theory of instructed second language acquisition. In N. Ellis (ed.), *Implicit and explicit learning of languages* (pp. 79-114). London: Academic Press.
- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-25). Bristol: Multilingual Matters.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

- Ellis, R., & He, X. (1999). The roles of modified input and output on the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21(2), 285-301.
- Enkvist, I. (2005). *Trängd mellan politik och pedagogik: Svensk språkutbildning efter 1990*. Hedemora: Gidlund.
- Erickson, G., & Börjesson, L. (2001). Bedömning av språkfärdighet i nationella prov och bedömningsmaterial. In R. Ferm & P. Malmberg (eds.), *Språkboken* (pp. 255-269). Stockholm: Myndigheten för skolutveckling.
- Erman, B., & Warren, B. (2000). The Idiom Principle and the Open Choice Principle. *Text*, 20(1), 29-62.
- Færch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Færch & G. Kasper (eds.), *Strategies in interlanguage communication* (pp. 20-60). Harlow: Longman.
- Fassnacht, C., & Woods, D. K. (2006). Transana v.2.22 <http://transana.org>. Madison, WI: The Board of Regents of the University of Wisconsin System.
- Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler & W. S.-Y. Wang (eds.), *Individual differences in language ability and language behavior* (pp. 85-101). New York, NY: Academic Press.
- Findahl, O. (2007). *Svenskarna och Internet 2007*. Hudiksvall: World Internet Institute.
- Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton & J. Treffers-Daller (eds.), *Modelling and assessing vocabulary knowledge* (pp. 116-132). Cambridge: Cambridge University Press.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273-293.
- Forsman, L. (2004). *Language, culture and context: Exploring knowledge and attitudes among Finland-Swedish EFL-students with particular focus on extracurricular influence*. Vasa: Åbo Akademi.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3), 299-323.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin Company.
- Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (ed.), *Second language acquisition in a study abroad context* (pp. 123-148). Amsterdam: John Benjamins.
- de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, 24(1), 81-112.
- Gardner, D. (2007). Validating the construct of *Word* in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241-265.
- Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second language learning. Part II: affective variables. *Language Teaching*, 26(1), 1-11.

- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.
- Gass, S. (2006). Generalizability: What are we generalizing anyway? In M. Chalhoub-Deville, C. A. Chapelle & P. Duff (eds.), *Inference and generalizability in applied linguistics: Multiple perspectives* (pp. 209-220). Amsterdam: John Benjamins.
- Gass, S., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, 19, 3-17.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave Macmillan.
- Giles, H. (2001). Speech accommodation. In R. Mesthrie (ed.), *Concise encyclopedia of sociolinguistics* (pp. 193-197). Amsterdam: Elsevier.
- Gorski, R. A. (2000). Sexual differentiation of the nervous system. In E. R. Kandel, J. H. Schwartz & T. M. Jessell (eds.), *Principles of neural science* (pp. 1131-1148). New York, NY: McGraw-Hill.
- Grundy, P. (1995). *Doing pragmatics*. New York, NY: St Martin's Press.
- Hasselgren, A. (1996a). The EVA Project - diagnostic tests in English for the 8th class, or something more? *Språk og språkundervisning*(2), 27-31.
- Hasselgren, A. (1996b). *Kartlegging av kommunikatív kompetanse i engelsk. User compendium for teachers*. Oslo: Nasjonalt læremiddelsenter.
- Hasselgren, A. (1997). Oral test subskill scores: What they tell us about raters and pupils. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (eds.), *Current developments and alternatives in language assessment - Proceedings of LTRC 96* (pp. 241-256). Jyväskylä: University of Jyväskylä and University of Tampere.
- Hasselgren, A. (2002a). Learner corpora and language testing: Smallwords as markers of learner fluency. In S. Granger, J. Hung & S. Petch-Tyson (eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 143-173). Amsterdam: John Benjamins.
- Hasselgren, A. (2002b). Sounds a bit foreign. In L. E. Breivik & A. Hasselgren (eds.), *From the COLT's mouth, and others': Language corpora studies - In honour of Anna-Brita Stenström* (pp. 103-123). Amsterdam: Rodopi.
- Hasselgren, A. (2004). *Testing the spoken English of young Norwegians: A study of test validity and the role of 'smallwords' in contributing to pupils' fluency*. Cambridge: Cambridge University Press.
- Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford University Press.
- Hellgren, P. (1982). *Communicative proficiency in a foreign language, and its evaluation: An analysis of the concept and an experiment with oral proficiency*. Helsinki: Department of Teacher Education, University of Helsinki.
- van Hest, E. (1997). Self-repair as a measure of L2 proficiency. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (eds.), *Current developments and alternatives in language assessment - Proceedings of LTRC 96*. Jyväskylä: University of Jyväskylä and University of Tampere.

- Higgins, C. (2009). Language in and out of the classroom: Connecting contexts of language use with learning and teaching practices. *Language Teaching*, 42(3), 401-403.
- Holec, H. (1981). *Autonomy and foreign language learning*. Oxford: Pergamon.
- Hood Roberts, A. (1965). *A statistical linguistic analysis of American English*. The Hague: Mouton.
- Horwitz, E. K. (1987). Surveying student beliefs about language learning. In A. Wenden & J. Rubin (eds.), *Learner strategies in language learning* (pp. 119-129). London: Prentice Hall.
- Horwitz, E. K. (1999). Cultural and situational influences on foreign language learners' beliefs about language learning: A review of BALLI studies. *System*, 27 (4), 557-576.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning. A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.
- Håkansson, U. (2001). Moderna språk i gymnasieskolan. In R. Ferm & P. Malmberg (eds.), *Språkboken* (pp. 60-68). Stockholm: Myndigheten för skolutveckling.
- Iwashita, N. (2001). The effect of learner proficiency on interactional moves and modified output in nonnative-nonnative interaction in Japanese as a foreign language. *System*, 29(2), 267-287.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24-49.
- Jackson, H., & Amvela, E. Z. (2000). *Words, meaning and vocabulary: An introduction to modern English lexicology*. London: Continuum.
- Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger & P. Bull (eds.), *Conversation: an interdisciplinary perspective* (pp. 166-196). Clevedon: Multilingual Matters.
- Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, 11(2), 149-158.
- de Jong, J. H. A. L., & van Ginkel, L. W. (1992). Dimensions in oral foreign language proficiency. In L. Verhoeven & J. H. A. L. de Jong (eds.), *The construct of language proficiency: Applications of psychological models to language assessment* (pp. 187-205). Amsterdam: John Benjamins.
- de Jong, J. H. A. L., & Verhoeven, L. (1992). Modeling and assessing language proficiency. In L. Verhoeven & J. H. A. L. de Jong (eds.), *The construct of language proficiency: Applications of psychological models to language assessment* (pp. 3-19). Amsterdam: John Benjamins.
- Kachru, Y. (2006). Pedagogical grammars: Second language. In K. Brown (ed.), *Encyclopedia of Language & Linguistics* (pp. 248-254). Oxford: Elsevier.
- Karlsson, M. (2004). *An ITiS teacher team as a community of practice*. Göteborg: Acta Universitatis Gothoburgensis.
- Katamba, F. (1989). *An introduction to phonology*. London: Longman.

- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58(2), 285-325.
- Klapp Lekholm, A. (2008). *Grades and grade assignment: Effects of student and school characteristics*. Göteborg: Acta Universitatis Gothoburgensis.
- Knutsson, R. (2006). *Formulaic language in L1 and L2*. Lund: Lund University.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145-164.
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70(4), 366-372.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Krashen, S. D. (1987). Applications of psycholinguistic research to the classroom. In M. H. Long & J. C. Richards (eds.), *Methodology in TESOL: A book of readings* (pp. 33-44). New York, NY: Newbury House.
- Lam, W. S. E. (2000). L2 literacy and the design of self: A case study of a teenager writing on the Internet. *TESOL Quarterly*, 34(3), 457-482.
- Lamb, M. (2004). 'It depends on the students themselves': Independent language learning at an Indonesian state school. *Language, Culture and Curriculum*, 17(3), 229-245.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 75(4), 440-448.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. (2005). Lexical Frequency Profiles: From Monte Carlo to the Real World. A response to Meara (2005). *Applied Linguistics*, 26(4), 582-588.
- Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA (European Second Language Association) Yearbook 1* (pp. 7-28). Amsterdam: John Benjamins.
- Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: the case of CASE. *Language Testing*, 13(2), 151-172.
- Leech, G. N., Wilson, A., & Rayson, P. (2001). *Word frequencies in written and spoken English based on the British National Corpus*. Harlow: Pearson Education.
- Leńko-Szymańska, A. (2000). Passive and active vocabulary knowledge in advanced learners of English. In B. Lewandowska-Tomaszczyk & P. J.

- Melia (eds.), *PALC'99: Practical applications in language corpora*. Frankfurt am Main: Peter Lang.
- Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach. *Language Learning*, 40(3), 387-417.
- Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice* (pp. 245-270). Basingstoke: Palgrave Macmillan.
- Lightbown, P. M., & Spada, N. (2006). *How languages are learned* (3 ed.). Oxford: Oxford University Press.
- Lin, M. S. P. (2006). *The role of formulaic language in the spoken performances of native and nonnative speakers of English*. Hong Kong: The Chinese University of Hong Kong.
- Lindblad, T. (1992). Oral tests in Swedish schools: a five-year experiment. *System*, 20(3), 279-292.
- Lindeman, S. (2006). What the other half gives: the interlocutor's role in non-native speaker performance. In R. Hughes (ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice* (pp. 23-49). Basingstoke: Palgrave Macmillan.
- Linderöth, J., & Bennerstedt, U. (2007). *Att leva i World of Warcraft*. Stockholm: Medierådet, Regeringskansliet.
- Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Lund: Lund Studies in English.
- Littlewood, W., Liu, N.-F., & Yu, C. (1996). Hong Kong tertiary students' attitudes and proficiency in spoken English. *RELC Journal*, 27(1), 70-88.
- Long, M. H. (1981). Input, interaction and second language acquisition. In H. Winitz (ed.), *Native language and foreign language acquisition* (Vol. 379, pp. 259-278). New York, NY: Annals of the New York Academy of Sciences.
- Long, M. H. (1987). Native speaker/Non-native speaker conversation in the second language classroom. In M. H. Long & J. C. Richards (eds.), *Methodology in TESOL: A book of readings* (pp. 339-354). New York, NY: Newbury House.
- Longman Dictionary of Language Teaching and Applied Linguistics* (3 ed.). (2002). J. C. Richards, R. Schmidt, H. Kendrick & Y. Kim (eds.). London: Longman.
- Lorenzo-Dus, N. (2007). The best of both worlds? Combined methodological approaches to the assessment of vocabulary in oral proficiency interviews. In H. Daller, J. Milton & J. Treffers-Daller (eds.), *Modelling and assessing vocabulary knowledge* (pp. 220-233). Cambridge: Cambridge University Press.
- Lorenzo-Dus, N., & Meara, P. (2004). Role-plays and the assessment of oral proficiency in Spanish. In R. Márquez Reiter & M. E. Placencia (eds.), *Current trends in the pragmatics of Spanish* (pp. 79-97). Philadelphia, PA: John Benjamins.
- Lorenzo-Dus, N., & Meara, P. (2005). Examiner support strategies and test-taker vocabulary. *IRAL*, 43(3), 239-258.

- Lundahl, B. (2009). *Engelsk språkdiraktik: Texter, kommunikation, språkutveckling*. Lund: Studentlitteratur.
- Macmillan English Dictionary for Advanced Learners of American English*. (2002). Oxford: Macmillan Publishers.
- Malmberg, P. (2000a). De moderna språken i grundskolan och gymnasieskolan från 1960. *Språk Gy2000:18. Grundskola och gymnasieskola. Kursplaner, betygsriterier och kommentarer* (pp. 7-26). Stockholm: Skolverket.
- Malmberg, P. (ed.). (2000b). *I huvudet på en elev. Projektet STRIMS: Strategier vid inläring av moderna språk*. Stockholm: Bonnier Utbildning.
- Malvern, D. D., & Richards, B. J. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85-104.
- Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Marcos-Llinás, M., & Juan Garau, M. (2009). Effects of language anxiety on three proficiency-level courses of Spanish as a foreign language. *Foreign Language Annals*, 42(1), 94-111.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, 32(1), 21-36.
- McCarthy, M., & O'Keeffe, A. (2006). Second language speaking. In K. Brown (ed.), *Encyclopedia of Language & Linguistics* (pp. 95-102). Oxford: Elsevier.
- McDonough, S. (2007). Motivation in ELT. *ELT Journal*, 61(4), 369-371.
- McKee, G., Malvern, D. D., & Richards, B. J. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3), 323-337.
- Meara, P. (1992). *EFL vocabulary tests*. Swansea: Centre for applied language studies, University College Swansea.
- Medierådet (2008). *Unga och medier 2008 - Fakta om barns och ungas användning och upplevelser av medier*. Stockholm: Regeringskansliet.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20(1), 83-107.
- Miliander, J. (2003). *We get the answer we deserve: A study of vocabulary in a spoken and written corpus of advanced learner English*. Karlstad: Karlstad University Studies.
- Minkova, D., & Stockwell, R. (2006). English words. In B. Aarts & A. McMahon (eds.), *The handbook of English linguistics* (pp. 461-482). Oxford: Blackwell.
- Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2 ed.). London: Arnold.
- Mitra, S., Dangwal, R., Chatterjee, S., Jha, S., Bisht, R. S., & Kapur, P. (2005). Acquisition of computing literacy on shared public computers: Children and the "hole in the wall". *Australasian Journal of Educational Technology*,

- 21(3), 407-426. Available from <http://www.ascilite.org.au/ajet/ajet21/mitra.html> (accessed May 25, 2009).
- Mizera, G. J. (2006). *Working memory and L2 oral fluency*. Pittsburgh, PA: University of Pittsburgh.
- Modiano, M. (2005). Cultural studies, foreign language teaching and learning practices, and the NNS practitioner. In E. Llorca (ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 25-43). New York, NY: Springer.
- Murray, G. (2008). Pop culture and language learning: Learners' stories informing EFL. *Innovation in Language Learning and Teaching*, 2(1), 2-17.
- Naiman, N., Frölich, M., Stern, H. H., & Todesco, A. (1996). *The good language learner*. Clevedon: The Ontario Institute for Studies in Education.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12-25.
- Nation, P. (2001a). How many high frequency words are there in English? In M. Gill, A. W. Johnson, L. M. Koski, R. D. Sell & B. Wärvik (eds.), *Language, learning, literature: Studies presented to Håkan Ringbom* (pp. 167-181). Turku: Åbo Akademi University Press.
- Nation, P. (2001b). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. (2005). Teaching and learning vocabulary. In E. Hinkel (ed.), *Handbook of research in second language teaching and learning* (pp. 581-595). Mahway, NJ: Lawrence Erlbaum.
- Nation, P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton & J. Treffers-Daller (eds.), *Modelling and assessing vocabulary knowledge* (pp. 35-43). Cambridge: Cambridge University Press.
- Naughton, D. (2006). Cooperative strategy training and oral interaction: Enhancing small group communication in the language classroom. *The Modern Language Journal*, 90(2), 169-184.
- Norton, B. (2000). *Identity and language learning: Gender, ethnicity and educational change*. Harlow: Pearson Education.
- Nunan, D. (1991). *Language teaching methodology: A textbook for teachers*. Hemel Hempstead: Prentice Hall International.
- Nycander, M. (2006). *Pojkars och flickors betyg: En statistisk undersökning*. Uppsala: Institutionen för lärarutbildning.
- O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing*, 12(2), 217-237.
- O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277-295.
- Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (1999). Factors associated with foreign language anxiety. *Applied Psycholinguistics*, 20(2), 217-239.
- Oscarson, M. (1998). Om självbedömning av språkfärdighet - empiri och reflektioner. In B. Ljung & A. Pettersson (eds.), *Perspektiv på bedömning av kunskap* (pp. 133-156). Stockholm: Institutionen för pedagogik, Lärarhögskolan i Stockholm.

- Oscarson, M., & Apelgren, B. M. (2005). *Nationella utvärderingen av grundskolan 2003 (NU-03). Engelska. Ämnesrapport till rapport 251*. Stockholm: Skolverket.
- Oxford English Dictionary Online (1989). Available from <http://dictionary.oed.com>.
- Pappamihiel, N. E. (2002). English as a second language students and English language anxiety: Issues in the mainstream classroom. *Research in the Teaching of English, 36*(3), 327-355.
- Pearson, N. (2004). The idiosyncrasies of out-of-class language learning: A study of mainland Chinese students studying English at tertiary level in New Zealand. In H. Reinders, H. Anderson, M. Hobbs & J. Jones-Parry (Eds.), *Supporting independent learning in the 21st century*. Proceedings of the inaugural conference of the Independent Learning Association, Melbourne AUS, 13-14 September 2003. Available from http://independentlearning.org/ILA/ila03/ila03_pearson.pdf (accessed June 11, 2009).
- Peregoy, S. F., & Boyle, O. F. (1991). Second language oral proficiency characteristics of low, intermediate, and high second language readers. *Hispanic Journal of Behavioral Sciences, 13*(1), 35-47.
- Pickard, N. (1995). Out-of-class language learning strategies: Three case studies. *Language Learning Journal, 12*(1), 35-37.
- Piirainen-Marsh, A., & Tainio, L. (2009). Other-repetition as a resource for participation in the activity of playing a video game. *The Modern Language Journal, 93*(2), 153-169.
- Plejer, C. (2004). *To fix what's not broken: Repair strategies in non-native and native English conversation*. Linköping: Linköping University.
- Pollitt, A., & Murray, N. L. (1996). What raters really pay attention to. In M. Milanovic & N. Saville (eds.), *Performance testing, cognition and assessment. Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem* (pp. 74-91). Cambridge: Cambridge University Press.
- Poole, G. (2005). Noam Chomsky. In S. Chapman & C. Routledge (eds.), *Key thinkers in linguistics and the philosophy of language* (pp. 53-56). Edinburgh: Edinburgh University Press.
- Purushotma, R. (2006). Wake up CALL. In P. Svensson (ed.), *Från vision till praktik: Språkutbildning och informationsteknik* (pp. 127-149). Härnösand: Myndigheten för nätverk och samarbete inom högre utbildning.
- Quaglio, P., & Biber, D. (2006). The grammar of conversation. In B. Aarts & A. McMahon (eds.), *The handbook of English linguistics* (pp. 692-723). Oxford: Blackwell Publishing.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Reinders, H. (2000). *Do it yourself? A learners' perspective on learner autonomy and self-access language learning*. Groningen: University of Groningen.
- Reinders, H. (2006). *The effects of different task types on L2 learners' intake and acquisition of two English grammatical structures*. Auckland: University of Auckland.

- Riggenbach, H. (1998). Evaluating learner interaction skills: Conversation at the micro level. In R. Young & A. W. He (eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 53-67). Amsterdam: John Benjamins.
- Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, 45(2), 283-331.
- Robinson, P., Ting, S. C.-C., & Urwin, J. J. (1995). Investigating second language task complexity. *RELC Journal*, 26(2), 62-79.
- Rubin, J. (2005). The expert language learner: A review of good language learner studies and learner strategies. In K. Johnson (ed.), *Expertise in second language learning and teaching* (pp. 37-63). New York, NY: Palgrave Macmillan.
- Sachs, J., Bard, B., & Johnson, M. L. (1981). Language learning with restricted input: Case studies of two hearing children of deaf parents. *Applied Psycholinguistics*, 2(1), 33-54.
- Sandlund, E. (2004). *Feeling by doing: The social organization of everyday emotions in academic talk-in-interaction*. Karlstad: Karlstad University Studies.
- Sax, L. (2007). *Boys adrift: the five factors driving the growing epidemic of unmotivated boys and underachieving young men*. New York, NY: BasicBooks.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmitt, N., Dörnyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (ed.), *Formulaic sequences* (pp. 55-86). Amsterdam: John Benjamins.
- Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: theory, research, and applications* (3. ed.). Upper Saddle River, NJ: Pearson Education.
- Seliger, H. W., & Shohamy, E. (1989). *Second language research methods*. Oxford: Oxford University Press.
- Silverman, D. (2005). *Doing qualitative research: A practical handbook*. London: Sage.
- Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (ed.), *Individual differences and instructed language learning* (pp. 69-94). Amsterdam: John Benjamins
- Skolverket (1993). *Den nationella utvärderingen av grundskolan våren 1992. Engelska*. Stockholm: Skolverket, Liber.
- Skolverket (2000). *Språk Gy2000:18. Grundskola och gymnasieskola. Kursplaner, betygsriterier och kommentarer*. Stockholm: Skolverket.
- Skolverket (2004a). *Engelska i åtta europeiska länder: En undersökning av ungdomars kunskaper och uppfattningar (Rapport 242)*. Stockholm: Skolverket.
- Skolverket (2004b). *National evaluation of the compulsory school in 2003: A summary main report (A summary of report 250)*. Stockholm: Skolverket.
- Stansfield, C. W., & Reed, D. J. (2004). The story behind the Modern Language Aptitude Test: An interview with John B. Carroll (1916-2003). *Language Assessment Quarterly*, 1(1), 43-56.
- Sundh, S. (2003). *Swedish school leavers' oral proficiency in English: Grading of production and analysis of performance*. Uppsala: Acta Universitatis Upsaliensis.

- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.
- Sylvén, L. K. (2004). *Teaching in English or English teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*. Göteborg: Göteborg University.
- Sylvén, L. K. (2006). How is extramural exposure to English among Swedish school students used in the CLIL classroom? *VIEWES - Vienna English Working Papers*, 15(3), 47-53.
- Tarone, E. (2005). Speaking in a second language. In E. Hinkel (ed.), *Handbook of research in second language teaching and learning* (pp. 485-502). Mahwah, NJ: Lawrence Erlbaum.
- Taylor, D. (1988). The meaning and use of the term "competence" in linguistics and applied linguistics. *Applied Linguistics*, 9(2), 148-168.
- Tornberg, U. (2005). *Språkdiradaktik* (3 ed.). Malmö: Gleerups.
- Warren, M. (2006). *Features of naturalness in conversation*. Amsterdam: John Benjamins.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79-95.
- Webb, S. (2009). L2 vocabulary: Perspectives on teaching, learning, testing, and formulaic sequences. *Language Teaching*, 42(3), 405-414.
- Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407-427.
- Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, 59(2), 335-366.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16(2), 180-205.
- Wenden, A. L. (1987). How to be a successful language learner: Insights and prescriptions from L2 learners. In A. Wenden & J. Rubin (eds.), *Learner strategies in language learning* (pp. 103-117). London: Prentice Hall.
- Vermeer, A. (1992). Exploring the second language learner lexicon. In L. Verhoeven & J. H. A. L. de Jong (eds.), *The construct of language proficiency: Applications of psychological models to language assessment* (pp. 147-162). Amsterdam: John Benjamins.
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17(1), 65-83.
- Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Boogaards & B. Laufer (eds.), *Vocabulary in a second language* (pp. 173-189). Amsterdam: John Benjamins.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.

- Viberg, Å. (2000). Tvåspråkighet och inlärnin g av språk i och utanför skolan. *Språk 2000:18. Kursplaner, betygskriterier och kommentarer* (pp. 27-41). Stockholm: Skolverket.
- Williams, J. M. (1975). *Origins of the English language: A social and linguistic history*. New York, NY: Free Press.
- Wolf, J. P. (2008). The effects of backchannels on fluency in L2 oral task production. *System, 36*(2), 279-294.
- Wood, D. (2001). In search of fluency: What is it and how can we teach it? *The Canadian Modern Language Review, 57*(4), 573-589.
- Vygotskij, L. (1999). Problem i undervisningen och den intellektuella utvecklingen i skolåldern. In G. Lindqvist (ed.), *Vygotskij och skolan: Texter ur Lev Vygotskij's Pedagogiska psykologi kommenterade som historia och aktualitet* (pp. 259-277). Lund: Studentlitteratur.
- Xiu Yan, J., & Horwitz, E. K. (2008). Learners' perceptions of how anxiety interacts with personal and instructional factors to influence their achievement in English: A qualitative analysis of EFL learners in China. *Language Learning, 58*(1), 151-183.
- Xue, G.-y., & Nation, P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215-229.
- Yi, Y. (2005). Asian adolescents' out-of-school encounters with English and Korean literacy. *Journal of Asian Pacific Communication, 15*(1), 57-77.
- Yule, G., & Tarone, E. (1991). The other side of the page: Integrating the study of communication strategies and negotiated input in SLA. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood Smith & M. Swain (eds.), *Foreign/Second language pedagogy research: A commemorative volume for Claus Færch* (pp. 162-171). Clevedon: Multilingual Matters.
- Öquist, O., & Wikström, H. (2006). *Individ- och klassvariation i grundskolan åk. 9. Studier av individ- och klassvariation i NU-materialet 2003*. Stockholm: Skolverket.

Appendix 1: Letter to the parents



Till föräldrar med barn i klass 8X, läsåret 2005-06, Xskolan

Information angående deltagande i forskarstudie om elevers muntliga engelska läsåret 2006-07

Jag heter Pia Sundqvist och är doktorand i engelska vid Karlstads universitet. Min avhandling kommer att handla om elevers muntliga engelska. Under läsåret 2006-07 kommer jag att samla in material från elever på ett flertal skolor. En av de skolor jag samarbetar med är **Xskolan**. Engelskläraren **NNs** elever i nuvarande klass **8X** kommer att ingå i studien och det är därför ni som målmän får denna första skriftliga information under vårterminen i årskurs 8.

För att få ett brett underlag för framtida analys kommer bland annat fem muntliga par-övningar att genomföras under årskurs 9. Dessa övningar kommer att ingå som en naturlig del i lärarens ordinarie undervisning i engelska. Det som är annorlunda är att jag kommer att videospela samtliga övningar och för detta behöver jag ert medgivande (se "Målsmans medgivande", där även en mer detaljerad beskrivning av studien ingår).

Som jag nämnde ovan ingår de muntliga övningarna i lärarens planering, så på det sättet är de "obligatoriska". *Varje enskild elevs deltagande i denna studie är dock helt frivilligt.* Givetvis är det min förhoppning att så många som möjligt vill vara med!

Var vänlig, läs igenom "Målsmans medgivande". Skriv under om du/ni tycker att ert barn får (eller inte får) delta i studien, och lämna sedan in medgivandet till skolan. Jag kommer att höra av mig igen skriftligen, och vi syns vid höstens första föräldramöte.

Ta gärna kontakt med mig eller min handledare om ni vill veta mera!

Med vänliga hälsningar,

Pia Sundqvist, tel 054/7001508 eller pia.sundqvist@kau.se

Handledare, professor Solveig Granath, tel 054/7001379 eller solveig.granath@kau.se

Appendix 2: Form of consent

Målsmans medgivande

”Forskarstudie om elevers muntliga engelska”

Årskurs 9

Läsåret 2006-07

Innan du ger ditt medgivande att låta din son/dotter delta i studien ber jag dig läsa igenom följande information.

Forskarstuderande och ansvarig för studien: Pia Sundqvist, doktorand i engelsk lingvistik, Karlstads universitet. Huvudhandledare är professor Solveig Granath, Karlstads universitet. Biträdande handledare är universitetslektor June Miliander.

Syfte med studien: Syftet med studien är att samla in data för att kunna analysera elevers muntliga engelska utifrån olika aspekter.

Beskrivning av studien: Studien bygger på att elever i årskurs 9 vid ett antal skolor deltar och genomför att antal uppgifter. I studiens inledning ombeds eleverna att besvara en allmän enkät samt en självbedömning. Under höstterminens början kommer eleverna även att tilldelas en så kallad språkdagbok, vilken de ombeds fylla i. Språkdagboken ska ifyllas även under en vecka mot slutet av vårterminen. Den största delen av studien består av videoinspelningar av elevers muntliga engelska. Samtliga deltagare kommer att spelas in med videokamera vid fem olika tillfällen, spridda över läsåret, då olika muntliga parövningar eller tester genomförs. Paren kommer att utses slumpmässigt bland klassens elever och vara olika vid varje inspelningstillfälle. De muntliga uppgifterna ingår som en del av lärarens ordinarie undervisning. Det muntliga material som ingår i studien kommer att transkriberas för att därefter analyseras. Enstaka exempel på elevens skriftliga produktion kommer att samlas in för att därefter analyseras. Vid slutet av studien kan det bli aktuellt med intervjuer av enstaka elever. Allt insamlat material kommer att behandlas med största aktsamhet. Inga personliga identiteter eller detaljer kommer att avslöjas. Studien genomförs i enlighet med Vetenskapsrådets regler för god forskningssed (för mer information: <http://www.vr.se/>).

Fördelar med att delta: Genom att delta i studien bidrar eleven till forskningen om elevers muntliga engelska. Ett deltagande innebär också att eleven får vara med om någonting speciellt, vilket på det personliga planet kan upplevas som utvecklande. Det är möjligt att eleven inspireras av att han/hon spelas in vid ett flertal tillfällen, vilket i sådana fall torde ha en motivationshöjande effekt. Deltagandet kan också innebära att eleven ökar medvetenheten om sin studiesituation, speciellt med avseende på engelska. Eftersom de muntliga övningarna spelas in och administreras av den forskarstuderande medför det att elevens engelsklärare får tillgång till ett bredare underlag vad gäller elevens muntliga färdigheter än vad som normalt är möjligt.

Nackdelar med att delta: Nackdelarna med att delta är få. För enstaka individer är det möjligt att närvaron av en kamera samt en ny vuxen (den forskarstuderande) kan kännas hämmande, åtminstone inledningsvis. Detta brukar normalt gå över; de flesta vänjer sig helt enkelt. Det är dock möjligt att vissa individer trots allt upplever omständigheterna som jobbiga, vilket till exempel kan medföra att de muntliga övningarna känns onödigt svåra att genomföra. Avsikten är att samtliga inspelningar sker under normal lektionstid för engelska, men av inspelningstekniska skäl kan det uppstå en situation som gör att eleven kommer för sent till efterföljande lektion. Förhoppningen är givetvis att tekniken och tidsplaneringen för inspelningarna ska fungera normalt hela tiden, så att eventuella förseningar undviks.

Sekretess: Insamlade data kommer att behandlas med största aktsamhet och förvaras säkert av den forskarstuderande. Videoinspelningarna kommer enbart att ses av den forskarstuderande och hennes handledare, samt eventuellt av medlemmar i Karlstads universitets forskargrupp för engelska. Alla

medverkande elever och skolor kommer att vara anonyma i efterföljande publikationer och presentationer, allt i enlighet med ovan nämnda regler för god forskningsetik.

Frivillighet: Deltagande i denna studie är frivilligt. Beslut att delta eller att inte delta påverkar inte målsmans eller elevs relation till engelskläraren eller den forskarstuderande. Målsman har rätt att när som helst dra tillbaka sitt medgivande och eleven utgår då ur studien.

Information: Om du har några frågor om studien, kontakta Pia Sundqvist (pia.sundqvist@kau.se eller 054/7001508) eller huvudhandledaren Solveig Granath (solveig.granath@kau.se eller 054/7001379). Pia Sundqvist kommer att närvara vid första föräldramötet på höstterminen i 9:an för att ge ytterligare information.

Medgivande: Genom att kryssa för 'Ja' och att skriva under detta dokument ger målsman sitt medgivande till sonens/dotterns deltagande i "Forskarstudie om elevers muntliga engelska", enligt vad som beskrivits ovan. Ett 'Nej' samt underskrift innebär att eleven ej deltar.

JA, jag tillåter att mitt barn deltar i "Forskarstudie om elevers muntliga engelska."

NEJ, jag tillåter *inte* att mitt barn deltar i "Forskarstudie om elevers muntliga engelska."

Elevens namn

Klass

Målsmans underskrift

Målsmans namnförtydligande

Ort och datum

Målsmans e-post*

* Om någon av målsmännen har e-post, var vänlig ange denna för att underlätta eventuell framtida korrespondens!



Inlämnas till mentor eller engelsklärare!

Appendix 3: Western Svealand

The sample in this study is representative of the geographical region Western Svealand, see the marked area on the map of Sweden below. Map available from Google Maps Sverige, <http://maps.google.se/>, accessed Oct. 20, 2009.



Appendix 4: Questionnaire

Elevenkät

För- och efternamn: _____ Klass: _____

Skola: _____

1. a) Vilket är ditt modersmål (= ditt första språk)? _____

-

b) Om ditt modersmål är något annat än svenska, ungefär när började du prata svenska? (Kryssa ett alternativ)

- Innan jag började skolan.
- När jag gick i 1:an-3:an.
- När jag gick i 4:an-6:an.
- När jag gick i 7:an-9:an.

2. a) Har du besökt/varit/bott i något *engelsktalande land*? Ja Nej

b) Om ja, vilket/vilka land/länder? _____

c) Hur länge sammanlagt? mindre än 3 månader
 3-12 månader
 mer än 12 månader

d) Vilka *andra länder* har du besökt/varit/bott i? _____

e) Hur länge sammanlagt? mindre än 3 månader
 3-12 månader
 mer än 12 månader

3. Hur ofta får du hjälp av någon hemma med läxor? (Kryssa ett alternativ)

- Vid varje läxtillfälle.
- Någon gång i veckan.
- Någon gång i månaden.
- Sällan eller aldrig.

4. Bor du så långt ifrån skolan att du åker med skolskjuts? Ja Nej

5. a) Pratar du regelbundet engelska med någon nära släkting, kompis eller liknande?

Ja Nej

b) Om ja, med vem/vilka (mamma, pappa, syskon, kusin, bästis, etc.)?

c) Om ja, markera med ett kryss hur ofta du brukar prata med denna person.

- dagligen
 någon eller några gånger veckan
 någon eller några gånger i månaden
 någon eller några gånger om året

6. Hur bor du?

Lägenhet Radhus/Parhus Fristående hus/Villa Studentrum

7. a) Brukar du läsa på din fritid (ej inräknat läxarbete)? Ja Nej

b) Om ja, *vad* läser du och *i vilken omfattning*?

- Tänk på att man kan läsa böcker/tidningar både ”som vanligt” och på datorn!
- Kryssa för allt som stämmer för dig!

	<i>Jag läser...</i>	Dagligen	Någon eller några gånger per vecka	Någon eller några gånger per månad	Aldrig eller nästan aldrig
1	<i>böcker på svenska</i>				
2	<i>nyhetstidningar på svenska</i>				
3	<i>serietidningar på svenska</i>				
4	<i>annat på svenska, t.ex. manualer, sångtexter m.m.</i>				
5	<i>böcker på engelska</i>				
6	<i>nyhetstidningar på engelska</i>				
7	<i>serietidningar på engelska</i>				
8	<i>annat på engelska, t.ex. manualer, sångtexter m.m.</i>				

8. Var bor du?

På landet I småstad I mellanstor stad I storstad
(Stockholm/Göteborg/Malmö)

9. a) Har du tillgång till dator i hemmet? Ja Nej
 b) Om ja, har du tillgång till Internet? Ja Nej
10. a) Bor dina föräldrar tillsammans? Ja Nej
 b) Om nej, var bor du? Hos mamma eller mest hos mamma
 Hos pappa eller mest hos pappa
 Ungefär lika mycket hos både mamma och pappa
 Annat boende:.....
11. a) Har du tillgång till TV i hemmet? Ja Nej
 b) Har ni fler kanaler än SVT1, SVT2 och TV4? Ja Nej
12. Vilket betyg tror du att du får i engelska när du slutar 9:an?
 MVG VG G Ännu inte godkänd
13. Vilken utbildningsnivå har dina föräldrar? (Kryssa)

	Grundskola (1-9:an)	Gymnasieskola	Högskola/Universitet	
Mamma				<input type="checkbox"/> Jag vet inte
Pappa				<input type="checkbox"/> Jag vet inte

14. a) Har du något engelskspråkigt favoritprogram på TV? Ja Nej
 b) Om ja, vilket är programmet? _____
15. a) Använder du dator på din fritid? Ja Nej
 b) Om ja, *vad* använder du dator till och *i vilken omfattning*? (Kryssa allt som stämmer!)

	<i>Jag använder datorn till</i>	Dagligen	Någon eller några gånger per vecka	Någon eller några gånger per månad	Aldrig eller nästan aldrig
1	<i>att göra läxor</i>				
2	<i>att surfa på Internet</i>				
3	<i>att sköta mina sociala kontakter, t.ex. Lunarstorm, MSN etc.</i>				
4	<i>att lyssna på musik etc.</i>				
5	<i>att se på film etc.</i>				
6	<i>att spela dataspel</i>				



16. a) Hur ofta ser du på engelskspråkiga TV-program som är *textade på svenska*?

- Dagligen
- Någon eller några gånger i veckan
- Någon eller några gånger i månaden
- Aldrig eller nästan aldrig

b) Ge något eller några exempel på sådana program: _____

17. Hur tror du att du lärt dig det mesta du kan i engelska?

- Allt eller nästan allt genom skolarbetet.
- Det mesta genom skolarbetet.
- Det mesta vid sidan av skolarbetet.
- Allt eller nästan allt vid sidan av skolarbetet.

18. a) Hur ofta ser du på engelskspråkiga TV-program som *inte är textade på svenska*?

- Dagligen
- Någon eller några gånger i veckan
- Någon eller några gånger i månaden
- Aldrig eller nästan aldrig

b) Ge något eller några exempel på sådana program: _____

19. Vad gör dina föräldrar? (Kryssa!)

	Studerar	Jobbar	Studerar och jobbar	Är hemma	Är arbetslös	
Mamma						<input type="checkbox"/> Jag vet inte
Pappa						<input type="checkbox"/> Jag vet inte

20. a) Hur ofta ser du på engelskspråkiga filmer (på TV/dator/dvd/video/bio)?

- Dagligen
- Någon eller några gånger i veckan
- Någon eller några gånger i månaden
- Aldrig eller nästan aldrig

b) Om du har möjlighet att välja textning till film, vad väljer du vanligtvis?

- Otextat
- Med svensk textremsa
- Med engelsk textremsa
- Jag har aldrig haft denna möjlighet, så frågan kan ej besvaras.

21. Hur bedömer du själv ditt arbete i engelska och dess resultat? (Kryssa)

	Stämmer mycket bra	Stämmer ganska bra	Stämmer ganska dåligt	Stämmer mycket dåligt
a) Jag anstränger mig för att göra mitt bästa i engelska	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Jag tycker att jag tar ansvar för mitt arbete i engelska	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Jag får visa vad jag kan i engelska	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

22. Är du rädd för att göra misstag när du pratar engelska?

- Ja, mycket.
- Lite grand.
- Nej, inte alls.

23. Hur ofta brukar du lyssna på musik där sångtexten är på engelska?

- Dagligen
- Någon eller några gånger i veckan
- Någon eller några gånger i månaden
- Aldrig eller nästan aldrig

24. a) Utövar du någon sport på fritiden? Ja Nej

b) Om ja, vilken/vilka? _____

c) Spelar du något instrument på fritiden? Ja Nej

d) Om ja, vilket/vilka? _____

e) Spelar du någon form av rollspel *på svenska* på fritiden? Ja Nej

f) Om ja, vilket/vilka? _____

g) Spelar du någon form av rollspel *på engelska* på fritiden? Ja Nej

h) Om ja, vilket/vilka? _____

i) Har du några övriga fritidsintressen? Ja Nej

j) Om ja, vilket/vilka? _____

25. Vad gör du om du inte kommer på vad du ska säga på engelska?
Kryssa alla alternativ som passar! Du kan även lägga till något 'Annat'
 längst ned.

- Jag gör ingenting.
- Jag använder kroppsspråket.
- Jag använder något annat ord eller uttryck på engelska.
- Jag ber den person jag pratar med om hjälp.
- Jag använder svenska (eller något annat språk)
- Annat:.....

26. Här kommer några påståenden om ämnet engelska som du ska ta ställning till genom att sätta kryss.

	Stämmer mycket bra	Stämmer ganska bra	Stämmer ganska dåligt	Stämmer mycket dåligt
a) Engelska intresserar mig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Jag tycker det är viktigt att ha bra kunskaper i engelska	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) De vuxna jag bor tillsammans med tycker att engelska är viktigt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Jag arbetar med engelska bara för att klara av proven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e) Kunskap i engelska kommer jag att behöva för att klara mina fortsatta studier	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f) Kunskap i engelska är bra för det jag tänker arbeta med i framtiden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- g) Jag är ofta borta från engelskan
- h) Det är för lite tid i engelska
- i) Det är för mycket tid i engelska
- j) Engelska är ett svårt ämne

27. a) Studerar du, eller har du studerat, något annat främmande språk än engelska i skolan? Ja Nej

b) Om ja, vilket/vilka? (kryssa)

- Franska
- Spanska
- Tyska
- Annat språk:.....

28. Hur tycker du det känns att prata engelska i klassrummet med kompisarna eller läraren?

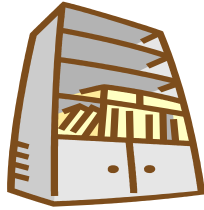
- Jag tycker det känns bra och har inga större problem med det.
- Jag tycker det känns ganska bra eller ok.
- Jag tycker det känns sådär och skulle föredra att slippa prata på engelska.
- Jag tycker det känns jobbigt och tycker inte alls om det.

29. Vad gör du om du inte förstår vad någon säger till dig på engelska? *Kryssa alla alternativ som passar!* Du kan även lägga till något 'Annat' längst ned.

- Jag gör ingenting.
- Jag frågar en kompis/läraren/föräldern.
- Jag kollar upp det hela i ett lexikon eller en grammatikbok.
- Jag kollar upp det hela med hjälp av datorn efteråt.
- Jag ber den person jag pratar med att repetera vad han/hon sa.
- Jag ber den person jag pratar med att förklara eller formulera om sig.
- Jag försöker att gissa.
- Annat:.....

30. Tänk efter hur det ser ut i ditt hem och gör en uppskattning av hur många böcker som sammanlagt finns där. Du kan räkna med att det går ungefär 50 böcker på en vanlig hylla som är 1 meter lång.

- Inga eller väldigt få böcker (0-10 böcker)
- 11-50 böcker (50 böcker ryms på ett hyllplan)
- 51-200 böcker (200 böcker ryms på fyra hyllplan)
- 201-500 böcker (500 böcker ryms på tio hyllplan)
- Fler än 500 böcker (fler än 500 böcker kräver mer än tio hyllplan)



Tack för hjälpen!

Appendix 5: Language diary (sample page)

MÅNDAG¹

ENGELSKA

Fyll i varje typ av kontakt du har haft med engelska under dagen, både i skolan och på fritiden.

		Totaltid Skriv timmar & minuter. Om du t.ex. läser i 25 min skriver du 0 tim 25 min.	
		I skolan	På fritiden
Läst bok (Skönlitteratur, t.ex. biblioteksbok. Inte vanliga skolböcker!)	Titel:		
Läst tidning:	Namn:		
Sett TV-program <i>Skriv S, E eller O efter varje program! Se nedan*</i>	Namn:		
Sett film (på bio, TV, video, DVD, dator, etc) <i>Skriv S, E eller O!*</i>	Namn:		
Surfat på nätet	Sidor:		
TV-/Dataspel	Namn:		
Lyssnat på musik	Artist(er):		
Annat	Exempel:		

* S = Svensk text. E = Engelsk text. O = Otextrad.

"Denna måndag har jag hållit på med...

	<i>...som vanligt"</i>	<i>...mindre än vanligt"</i>	<i>...mer än vanligt"</i>
ENGELSKA			
SVENSKA			
Ev. ÖVR. SPRÅK			

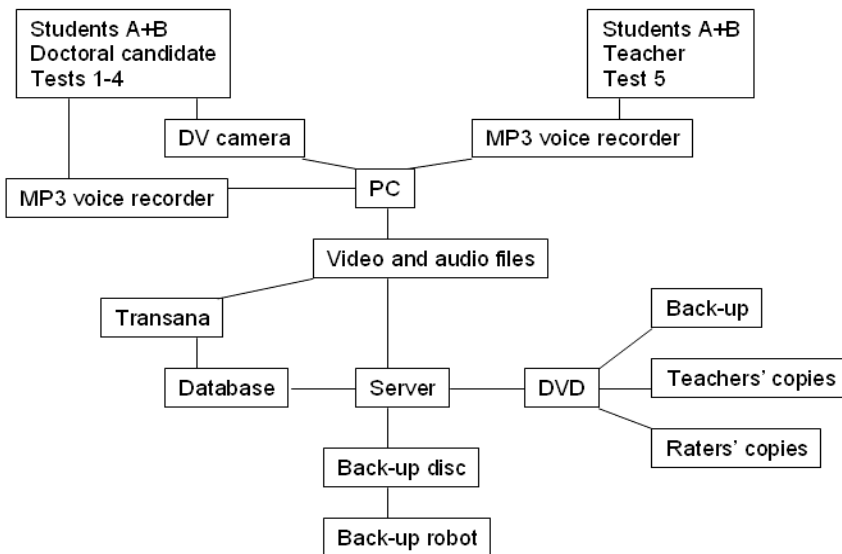
(Markera med ett kryss vad som stämmer för respektive språk.)

¹ This is an example of one part (English) of the language diary for one day (Monday). The other two parts are identical, except for the language heading (Swedish; Other languages). Following these three parts, students were asked to rate their extramural contacts with languages in relation to an average Monday (Tuesday, etc.).

Appendix 6: Technical details

Speech from tests 1-4 was recorded on video.¹ The video recordings were made with a DV camera (Panasonic NV-GS180E) using a wide lens (Panasonic VW-LW3707M3). A microphone (Stageline ECM-302B) was used in a few recordings to improve sound quality, but for the most part the camera microphone sufficed. A portable digital voice recorder (iRiver iF-799) was used as back-up. Some of the video recordings were transferred to the computer with Pinnacle software (Pinnacle Studio Version 9.4) and others with Panasonic software (Panasonic Motion DV Studio). When the former was used, files were saved directly in the MPEG-1 format. When the latter was used, files were by default saved in the AVI-format and then converted into MPEG-1. MPEG-1 was chosen as the desired format since it minimizes storage space and complies with the recommendations of Transana (Fassnacht & Woods, 2006), which was the qualitative software for analysis of video and audio data used in the present study.

For test 5, only audio recording was used. Audio recordings were also used as back-up for tests 1-4. Audio files were transferred to the computer using iRiver software (iRiver Music Manager Version 3.16) and then converted into and saved in the MP3-format, using the same software.



¹ Marjolijn Deunk, University of Groningen, was consulted in the technical set-up. I am very grateful for all her help.

Appendix 7: Expository speech (Test 4)

EXPOSITORY SPEECH (”förklarande tal”)

- Prepare an **expository speech** at home. (‘Expository’ = ‘förklarande’)
- Time: **2-5 minutes**.
- You may use **keywords**, if you would like to. Max: 10 keywords.
- You may **bring something** that suits your speech. For example, if you want to explain what is typical for a good snowboard, you can bring the snowboard and use it in your speech.
- Your audience: You will give your speech to **a friend and Pia**. Afterwards, there will be some time for comments and questions.
- When you are in the audience: Be a **good listener! Participate actively** in the conversation that follows your classmate’s speech.

Some piece of advice: Choose a topic you are familiar with! Prepare at home. Talk aloud in English when you practice. Try to time yourself when you prepare, to make sure you don’t talk too little or too much! It is often good to practice in front of a mirror; pretend it is your audience. If you want to use a few keywords, write them on a piece of paper. Remember to have an introduction and an end. And, speak English all the time!

Here are some examples of topics students have used over the years:



- “My rabbit” (brought a picture to accompany the speech)
- “My snake Sally” (Please, do *not* bring any live pets to school!!!)
- “My big brother”
- “What a floorball court looks like” (brought drawing of the court)
- “My idol”
- “How tennis ranking works”
- “The best lyrics ever written”
- “My favorite song”
- “How to put on a snowboard” (brought a snowboard plus a boot!)
- “My moped”
- “The best/worst book (film) I have read (seen)”
- “How to bake muffins” (brought the ingredients, a bowl etc.)
- “My computer”
- “What I have learned in.....(History/Geography/P.E. etc)”
- “My best friend” (brought a picture)
- “How to shoot a film” (brought a camera)
- “My hobby: to shoot trap”
- “Power play in ice hockey”
- “How my mobile phone works” (used the phone to demonstrate functions)
- “When I went to Spain with my family”
- “My instrument” (brought a flute)



Good luck!

Appendix 8: Test-instructor's manuscript for Test 4

EXPOSITORY SPEECH Test-instructor's manuscript

Welcome! This is the fourth time we meet. It is a bit different now, compared to the other times, since you have prepared what you will talk about at home.

Now, let's see, we have to make sure we have the right set-up. First of all, Student A, that is /NAME/, you can sit /here/ etc.

- ✓ *Arrange seating* for A and B and *mention their names* for the recording.
- ✓ *Explain* that A will begin by giving his/her speech for a few minutes, to friend+me.
- ✓ Say that it is going to be very *interesting* to listen to both students.
- ✓ Mention that while we are in "*the audience*", we *listen carefully*.
- ✓ Afterwards, the audience will comment on the speech, ask questions, and so forth.
- ✓ Do both of you understand?

Please, start A.

A gives his/her speech. (Try to keep track of the time; should last 2-4 minutes.)

Thank you.

B (A), please, you can start. What comments do you have for A (B)?

- ✓ Add my own comments.
- ✓ Questions from B (A) and test-taker.

List of comments and questions that can be used:

- That was a/an...
interesting/different/entertaining/fascinating/new/(un)usual/scary/fun/educational etc. topic.
- I have (never) heard about...
- Why did you choose this topic? How come you chose this topic?
- Where did you prepare your speech?
- How did you prepare your speech?
- The next time you give a speech in English, is there something you would do different? What/Why/Why not?

Fine, let's move on, now it's B's turn.

B gives his/her speech. (Try to keep track of the time; should last 2-4 minutes.)

Time for comments and questions, see above.

Wind off the test and thank the students for their participation.

Appendix 9: Instructions for the assessment of speaking tests

INSTRUCTIONS for evaluation of oral tests

The process of evaluation is in three stages. The first stage entails getting a picture, or profile, of how the student performed on the test. The next stage involves placing the student's performance according to the criteria scales (see below), leading to the awarding of a grade. Finally, the grade is adjusted if necessary according to the criteria relating to pronunciation and intonation.

The evaluator first watches/listens to the performance of both students (A and B) while filling in a performance profile scheme for each student. You have to choose between five alternative levels of performance for each of the questions in the profile scheme and put a cross in one of the boxes.

By working through the profile when listening to the performance, the evaluator's attention is drawn to all the different aspects of speech that need to be considered when assessing the student. Without this type of guidance it is tempting to listen for certain limited aspects of speaking.

Setting grades on the oral test

After listening and filling in the performance profile for each pupil (A and B), the next task is to decide on a grade, from 1 to 6. The criteria scales (shown at the end) are used for this purpose. The following advice is offered:

Using the performance profile as a guide:

- 1) Try to match the performance with a grade on each of the two scales: *message and fluency* and *language structures and vocabulary*. No performance will match perfectly—try to find the nearest description! It is a good idea to begin by deciding which of the three broad 'bands' the performance best fits into, and then seeing whether the performance would be 'weak' or 'strong' within this band—this will indicate the approximate grade.
- 2) Find an 'overall grade' that best reflects the grades on the two scales.
- 3) Use pronunciation/intonation as a final adjuster (see below) in order to arrive at a final grade.

In placing a student on an overall grade, the general principle employed is that the evaluator should aim at choosing the grade which best reflects the placings on both scales. However, when this decision is difficult to make, the scale *message and fluency* should be weighted more heavily than the scale related to *language*.

There are two related reasons for this: first, in the speaking situation, the success of communication is dependent on numerous factors of which linguistic 'correctness' is only one. Being able to take initiative and 'keep going', showing politeness and interest, knowing how to cope when language limitations arise are crucially important in speaking. The second reason for rating 'communicativeness' over 'correctness' lies in the nature of speaking itself. Speaking is done under time pressure, and with no means of 'crossing out' what was incorrect. Even native speaker speech is 'untidy' compared to writing. We make false starts and rarely produce perfect 'sentences'. Learners have the additional burden of coping with second language grammar under pressure, and even very advanced foreign speakers of English make mistakes they would never submit on paper. It seems reasonable therefore that students, who will normally be assessed most heavily on formal skills in their written English, should be judged more

mildly for these in the oral test, and *credited for other, more relevant skills*. This will give a more balanced total picture of the student's overall ability. If we were to use the same criteria for judging spoken and written performance, the oral test would have little function! The criteria scales are shown at the end.

The place of intonation and pronunciation

Because there is little direct correlation between pronunciation and intonation on one hand and general performance on the other, pronunciation and intonation will be assessed independently, and used as adjusters in setting the final grade given to a student.

The following guidelines are given:

- *At level 6*, both pronunciation and intonation must be 'very good', i.e. the pronunciation is such that no sounds could be misinterpreted, and the intonation both supports the message and indicates friendliness and interest.
- *At level 5*, both pronunciation and intonation must be 'acceptable', i.e. do not 'block' the communication to a significant extent.
- *At all other levels*, students on the 'borderline' between two grades should be upgraded when pronunciation and intonation are acceptable, and downgraded when they are not acceptable.

<p style="text-align: center;">Criteria scales for evaluating spoken English</p>

MESSAGE AND FLUENCY

5-6

- This level is characterized by a good and independent performance on all tasks tested. The student should take initiative and willingly supply original, detailed contributions, which are linked logically into cohesive 'wholes'. S/he should be able to 'keep going' with the minimum of help. Speech will be 'flowing' with little hesitation.
- *For the final awarding of '5' or '6'*, see the section on pronunciation and intonation.

3-4

- At this level the student should manage an adequate performance on most of the tasks, with some simple linking of ideas.
- *At level 3*, the student may be very hesitant and need a good deal of help to keep going. S/he should manage 'the essentials' of tasks which involve concrete ideas, such as information, actions and physical features in pictures (if applicable; pictures are not included in all oral tests).
- *At level 4*, the student will sometimes be able to keep going quite well without help. S/he will from time to time take the initiative and contribute more than 'the essentials' of a task. Both concrete and more abstract ideas such as opinions, reasons, and feelings, will be tackled fairly well.

1-2

- What the student contributes as this level will largely be in response to questions from the tester, and answers will be very short and generally inadequate.
- *At level 2*, however, the student should make at least a minimal response to most tasks.

LANGUAGE STRUCTURES AND VOCABULARY

5-6

- At this level the language should be characterized as idiomatic, varied, independent with few errors. The vocabulary will have very few ‘gaps’. The pupil should be able to use the appropriate style and degree of politeness. The student will rarely resort to Swedish (or his/her first language, if other than Swedish).
- *For the final awarding of ‘5’ or ‘6’*, see section on pronunciation and intonation.

3-4

- At this level there will be errors and examples of ‘Swenglish’, but the message should be understandable in most tasks. Ideas will be linked simply.
- *At level 4*, the language should show some originality and independence from the given material. Swedish (or another first language) will only be resorted to from time to time.
- *At level 3*, the wording may not be very original, being very dependent on what is ‘given’ in the task. Swedish (or other L1) words may often be used.

1-2

- At this level, the language will be insufficient to cope with the more demanding tasks.
- However, *at level 2*, despite ‘gaps’, many errors and frequent reliance on Swedish (or other L1) terms, the language should be sufficient to provide simple short responses to most tasks.

These instructions are almost identical to the instructions given in *Kartläggning av kommunikativ kompetans i engelska* (Hasselgren, 1996b). The instructions have been adapted to suit Swedish conditions and the current study.

Appendix 10: Performance profile scheme

PERFORMANCE PROFILE SCHEME GENERAL

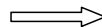
Audio/Video file ID: _____

Pupil A / B: _____ Evaluator/Assessor: _____

Use separate schemes for both pupils (A/B) simultaneously as you go through the audio/video file. Put a cross in one of the five given alternatives.

In the test generally

- a) *when difficulties in communication arose, did the pupil make an independent attempt to overcome these, in English?*
- virtually always
- sometimes
- rarely
- b) *was the pupil able to take the initiative, and make relevant independent contributions*
- yes
- on the whole, but hesitant at times, often needing encouragement
- no, was only able to respond to questions
- c) *the pupil's 'flow' of speaking was*
- very good, with a comfortable 'speed'
- slow, so that the pace of the conversations was affected
- frequently so broken up that the message was not coherent
- d) *the pronunciation was*
- very good, with no sounds that could be misinterpreted
- good enough to get the message across
- difficult to understand (or would be for a non-Swede!)
- e) *the intonation*
- was good and supported the message on the whole
- was not very good, but did not interfere with the message
- was poor, and interfered with the message



Pupil A / B: _____

- f) *did the tone indicate friendliness, politeness and interest?*
 very much so

 to some extent

 no
- g) *the language structures and vocabulary were*
 appropriate and fully understandable, without many errors

 on the whole understandable, despite errors and gaps

 so full of errors and gaps that the message was not easily understood
- h) *the language choices were*
 adventurous, independent and idiomatic

 independent and idiomatic at times

 very dependent on the 'given' material/input
- i) *the student's ability to interact with the other student was*
 excellent (takes initiatives, adapts speech to suit partner/situation etc)

 acceptable

 very poor (e.g. doesn't respond to cues etc)
- j) *the way the student treated the subject/topic was*
 excellent (focussed/in depth, with rich content)

 acceptable

 very poor (brief/shallow)

Grade (1-6) on the oral test (see criteria scales on instruction sheet):

<i>Message and fluency</i>	<i>Language structures and vocabulary</i>	<i>Overall</i>

Appendix 11: Interrater reliability

Pearson (r) for Tests 1-5: the overall grade (the OP grade) and factorial grades for fluency and vocabulary.

		Pearson Correlation: Overall grade (OP grade)			
		Rater 1	Rater 2	Rater 3	Rater 4
Test 1	Rater 1	1	.451**	.703**	
	Rater 2	.451**	1	.512**	
	Rater 3	.703**	.512**	1	
	Rater 4				
Test 2	Rater 1	1		.764**	.585**
	Rater 2				
	Rater 3	.764**		1	.624**
	Rater 4	.585**		.624**	1
Test 3	Rater 1	1	.593**		.602**
	Rater 2	.593**	1		.574**
	Rater 3				
	Rater 4	.602**	.574**		1
Test 4	Rater 1	1		.664**	.530**
	Rater 2				
	Rater 3	.664**		1	.565**
	Rater 4	.530**		.565**	1
Test 5	Rater 1	1	.546**	.473**	
	Rater 2	.546**	1	.569**	
	Rater 3	.473**	.569**	1	
	Rater 4				

**Correlation is significant at the .01 level (2-tailed)

		Pearson Correlation: Factorial grade - Fluency			
		Rater 1	Rater 2	Rater 3	Rater 4
Test 1	Rater 1	1	.569**	.693**	
	Rater 2	.569**	1	.550**	
	Rater 3	.693**	.550**	1	
	Rater 4				
Test 2	Rater 1	1		.743**	.581**
	Rater 2				
	Rater 3	.743**		1	.641**
	Rater 4	.581**		.641**	1
Test 3	Rater 1	1	.597**		.677**
	Rater 2	.597**	1		.593**
	Rater 3				
	Rater 4	.677**	.593**		1
Test 4	Rater 1	1		.672**	.557**
	Rater 2				
	Rater 3	.672**		1	.555**
	Rater 4	.557**		.555**	1
Test 5	Rater 1	1	.610**	.541**	
	Rater 2	.610**	1	.651**	
	Rater 3	.541**	.651**	1	
	Rater 4				

**Correlation is significant at the .01 level (2-tailed)

Pearson Correlation: Factorial grade - Vocabulary

		Rater 1	Rater 2	Rater 3	Rater 4
Test 1	Rater 1	1	.367**	.689**	
	Rater 2	.367**	1	.524**	
	Rater 3	.689**	.524**	1	
	Rater 4				
Test 2	Rater 1	1		.699**	0.572**
	Rater 2				
	Rater 3	.699**		1	.628**
	Rater 4	.572**		.628**	1
Test 3	Rater 1	1	.598**		.620**
	Rater 2	.598**	1		.600**
	Rater 3				
	Rater 4	.620**	.600**		1
Test 4	Rater 1	1		.633**	.502**
	Rater 2				
	Rater 3	.633**		1	.498**
	Rater 4	.502**		.498**	1
Test 5	Rater 1	1	.537**	.554**	
	Rater 2	.537**	1	.578**	
	Rater 3	.554**	.578**	1	
	Rater 4				

**Correlation is significant at the .01 level (2-tailed)

Appendix 12: Productive Levels Test (page 1)

Complete the underlined words as has been done in the example.

Ex) He was riding a bicycle.

Good luck!



Do NOT write
in the boxes!

- 1) I'm glad we had this opp_____to talk. 1
- 2) There are a doz_____eggs in the basket. 2
- 3) The pirates buried the treas_____ on a desert island. 3
- 4) Her beauty and cha_____had a powerful effect on men. 4
- 5) La_____of rain led to a shortage of water in the city. 5
- 6) He takes cr_____and sugar in his coffee. 6
- 7) The rich man died and left all his we_____to his son. 7
- 8) Pup_____must hand in their homework by the end of the week. 8
- 9) This sweater is too tight. It needs to be stret_____. 9
- 10) Ann intro_____her boyfriend to her mother. 10
- 11) Teenagers often adm_____and worship pop singers. 11
- 12) If you blow up that balloon any more it will bur_____. 12
- 13) In order to be accepted into the university, he had to impr_____his grades. 13
- 14) The package was deli_____two hours after it had been sent. 14
- 15) We try to go abr_____at least once a year. 15
- 16) The dress you're wearing is lov_____. 16
- 17) He wasn't very popu_____when he was a teenager, but he has many friends now. 17
- 18) I found the book deeply distu_____. 18

Appendix 13: Teacher's instructions (VLT)

Läarinstruktioner *Vocabulary Levels Test*

Tid: Testet tar ca. 30 min. Elever får sitta längre, dock max 40 minuter.

Hjälpmedel: Inga, men om någon elev frågar hur ett ord uttalas får läraren säga ordet högt.

Extra stort tryck: 2 prov-ex i förstorat format bifogas för elever som har speciella svårigheter att läsa liten text. Kopiera upp fler exemplar om behov finnes!

När? Testet genomförs förslagsvis i *vecka 17 (23-27/4)*, men läraren får själv bestämma om en annan tidpunkt är lämpligare, dock får testet ej genomföras förrän *tidigast i vecka 14 (mån 2 april)* och *senast i vecka 19 (ons 16 maj)*.

Rättning + hantering av prov: Pia rättar samtliga test och gör egna kopior (för vidare analys). Läraren (och eleverna) får tillbaka originalproven så snart rättningen är klar. Använd bifogade svarskuvert för inskickning av proven! Ha gärna provet redo ovanstående veckor och genomför det under en lektion när många elever är på plats. Eftersom det är ett vanligt oförberett ordprov får gärna hela klassen delta i provet (även de som ej ingår i studien). Jag rättar samtliga!

- 1) Dela ut provet till eleverna.
- 2) Be alla elever skriva namn och klass.
- 3) Läs instruktionerna högt: "Vocabulary levels test. This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning."
- 4) Be eleverna att titta på exemplet. Läs de 6 orden högt, läs de tre förklaringarna högt.
- 5) Be eleverna att titta på exemplets svar: Läs högt. "You answer the example in the following way: wall =6 part of a house, horse=3 animal with four legs, pencil=4 something used for writing".
- 6) Läs resten av instruktionen: "Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are *business, clock, shoe*."
- 7) Säg att om de är osäkra på något svar bör de ändå chansa på det alternativ de tror är rätt! Man får inga avdrag om man skriver fel svar.
- 8) Säg att eleverna har 40 minuter på sig att genomföra provet. Betona att de ska titta igenom sina svar på alla sidor innan de lämnar in. Berätta vad de ska göra när de är klara (t.ex. om de ska jobba med något, eller om de får gå).
- 9) Fråga om det finns några frågor och besvara dessa.
- 10) Önska dem lycka till och sedan får de börja.
- 11) Vid inlämning: *Kontrollera noga att det står för- och efternamn* på varje prov.

Tack för hjälpen!

Pia

Appendix 14: Vocabulary Levels Test (page 1)

Name: Class:

VOCABULARY LEVELS TEST

TOTAL: P

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning.

Ex)

You answer the example in the following way:

- 1 business
- 2 clock _____ part of a house
- 3 horse _____ animal with four legs
- 4 pencil _____ something used for writing
- 5 shoe
- 6 wall

- 1 business
- 2 clock 6 part of a house
- 3 horse 3 animal with four legs
- 4 pencil 4 something used for writing
- 5 shoe
- 6 wall

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, the words that are supposed to be more difficult are *business*, *clock*, and *shoe*.

- 1.
- 1 copy
 - 2 event _____ end or highest point
 - 3 motor _____ this moves a car
 - 4 pity _____ thing made to be like another
 - 5 profit
 - 6 tip

- 6.
- 1 admire
 - 2 complain _____ make wider or longer
 - 3 fix _____ bring in for the first time
 - 4 hire _____ have a high opinion of someone
 - 5 introduce
 - 6 stretch

- 2.
- 1 accident
 - 2 debt _____ loud deep sound
 - 3 fortune _____ something you must pay
 - 4 pride _____ having a high opinion of yourself
 - 5 roar
 - 6 thread

- 7.
- 1 arrange
 - 2 develop _____ grow
 - 3 lean _____ put in order
 - 4 owe _____ like more than something else
 - 5 prefer
 - 6 seize

- 3.
- 1 birth
 - 2 dust _____ game
 - 3 operation _____ winning
 - 4 row _____ being born
 - 5 sport
 - 6 victory

- 8.
- 1 blame
 - 2 elect _____ make
 - 3 jump _____ choose by voting
 - 4 manufacture _____ become like water
 - 5 melt
 - 6 threaten

- 4.
- 1 clerk
 - 2 frame _____ a drink
 - 3 noise _____ office worker
 - 4 respect _____ unwanted sound
 - 5 theatre
 - 6 wine

- 9.
- 1 brave
 - 2 electric _____ commonly done
 - 3 firm _____ wanting food
 - 4 hungry _____ having no fear
 - 5 local
 - 6 usual

- 5.
- 1 dozen
 - 2 empire
 - 3 gift _____ chance
 - 4 opportunity _____ twelve
 - 5 relief _____ money paid to the government
 - 6 tax

- 10.
- 1 bitter
 - 2 independent _____ beautiful
 - 3 lovely _____ small
 - 4 merry _____ liked by many people
 - 5 popular
 - 6 slight

Appendix 15: Figures

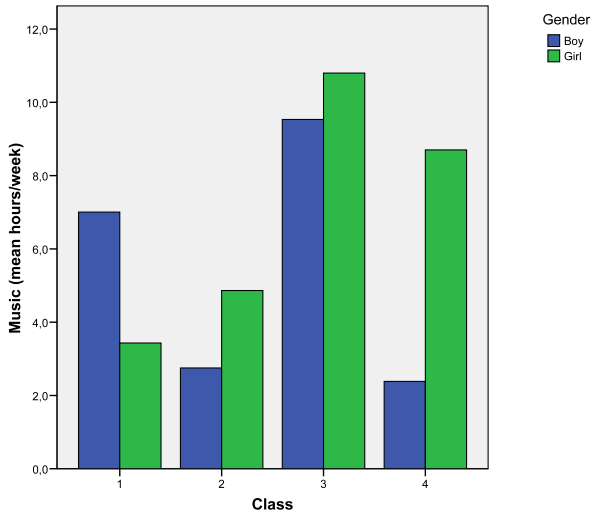


Figure 7.2. EE “listening to music” for the four classes and gender.

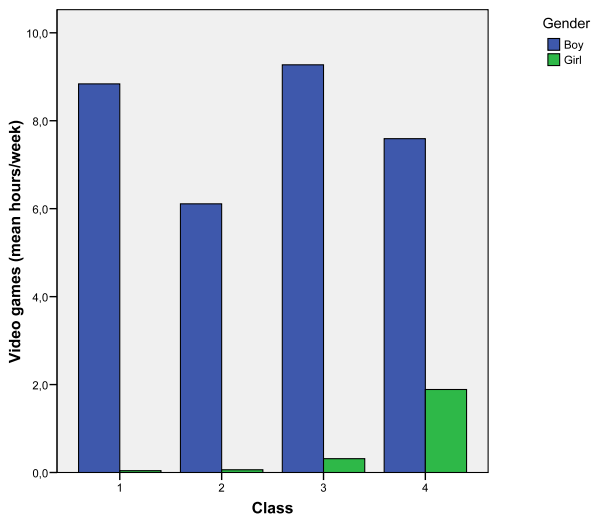


Figure 7.3. EE “playing video games” for the four classes and gender.

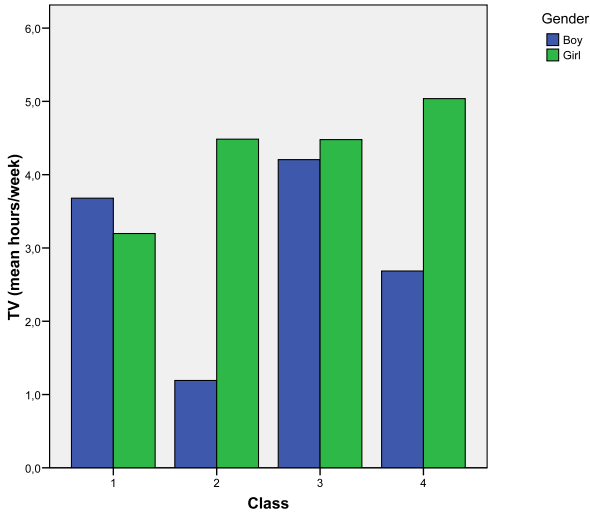


Figure 7.4. EE "watching TV" for the four classes and gender.

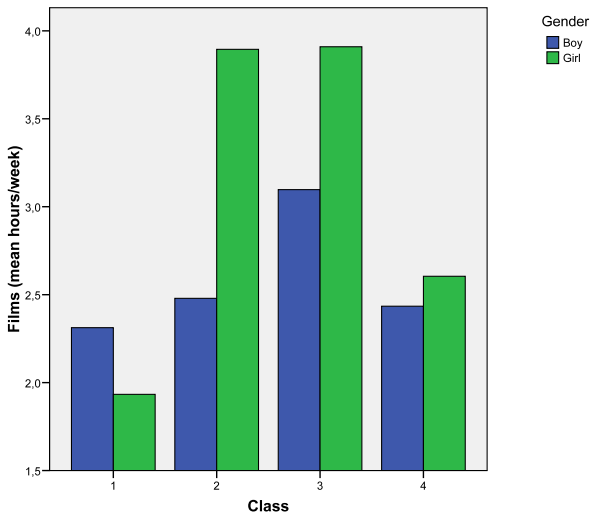


Figure 7.5. EE "watching films" for the four classes and gender.

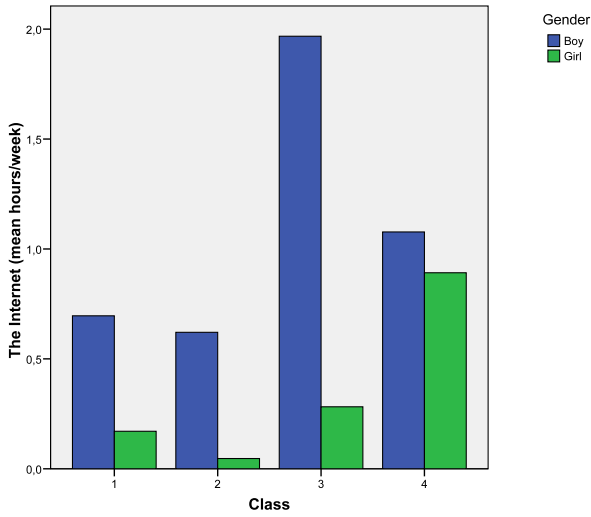


Figure 7.6. EE “surfing the Internet” for the four classes and gender.

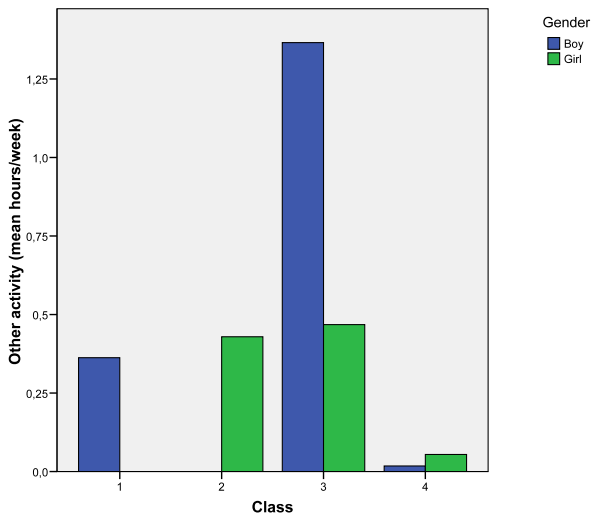


Figure 7.7. EE “other activity” for the four classes and gender.

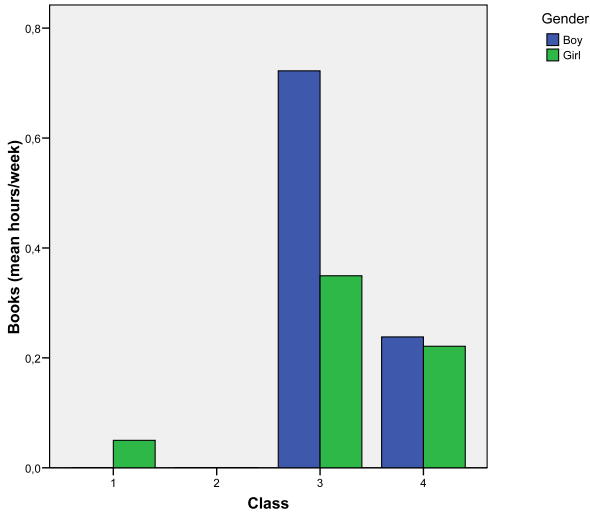


Figure 7.8. EE “reading books” for the four classes and gender.

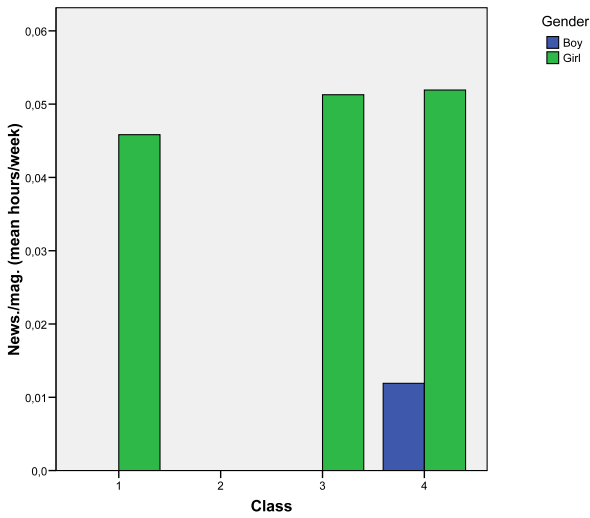


Figure 7.9. EE “reading newspapers and/or magazines” for the four classes and gender.

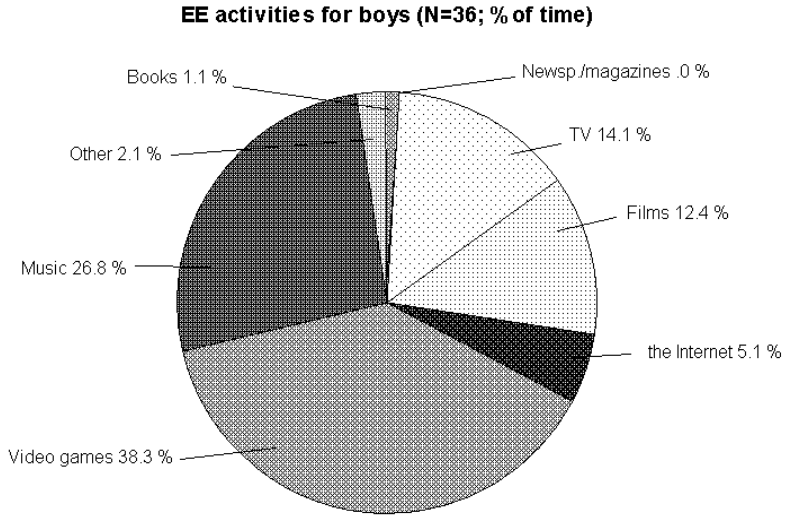


Figure 7.10. EE activities for boys as percentages of boys' time spent on EE per week (20.8 hours).

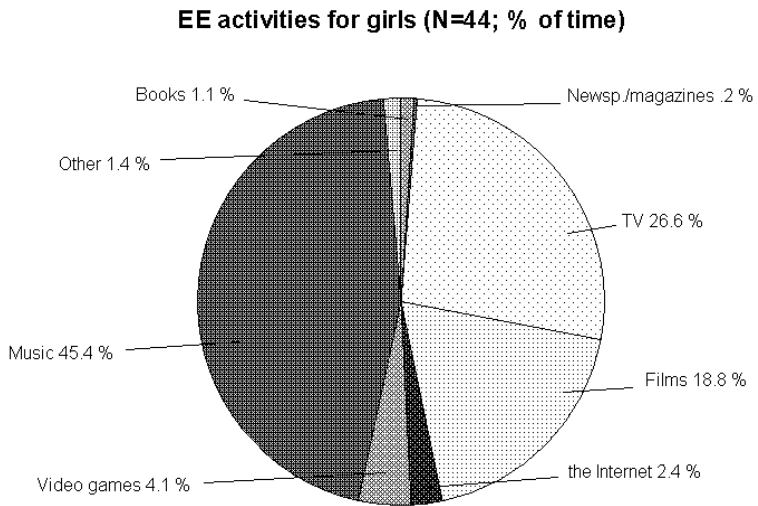


Figure 7.11. EE activities for girls as percentages of girls' time spent on EE per week (16.4 hours).

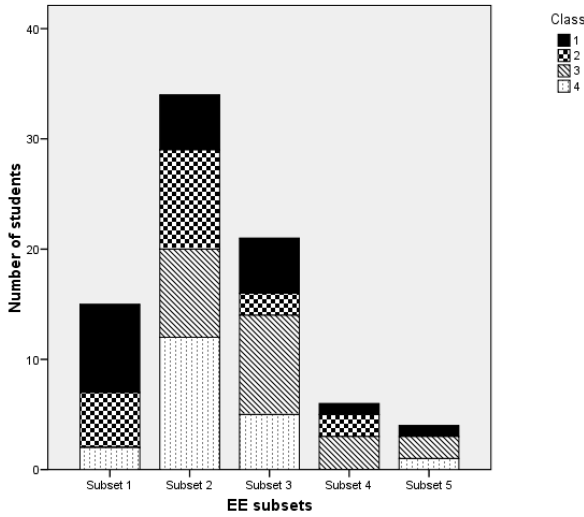


Figure 7.13. The distribution of the four classes in the five EE subsets.

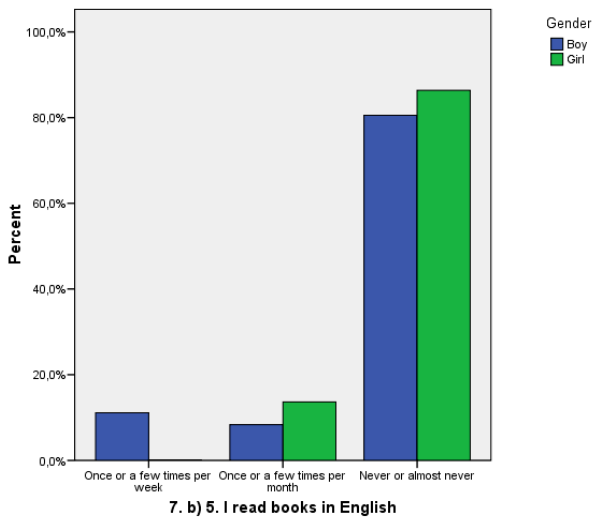


Figure 7.14. Responses to question 7 (item b: 5), "I read books in English" – gender distribution.

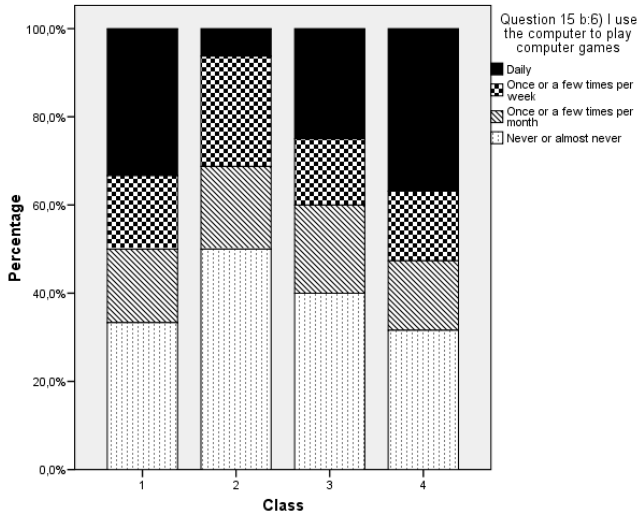


Figure 7.15. Responses to question 15 b, item 6, “I use the computer to play computer games” – the four classes.

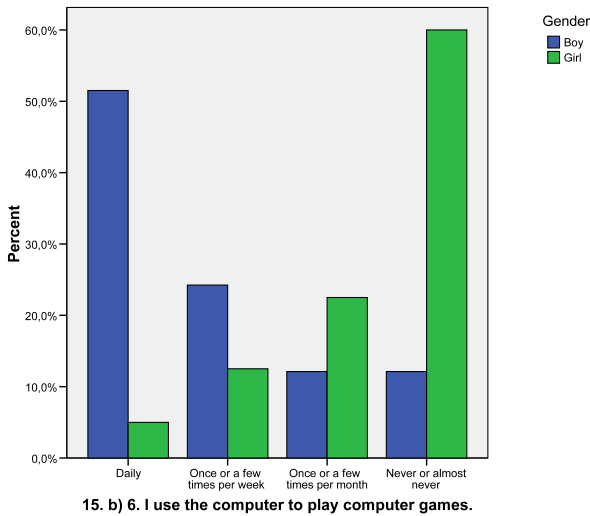


Figure 7.16. Responses to question 15 b, item 6, “I use the computer to play computer games” – gender distribution.

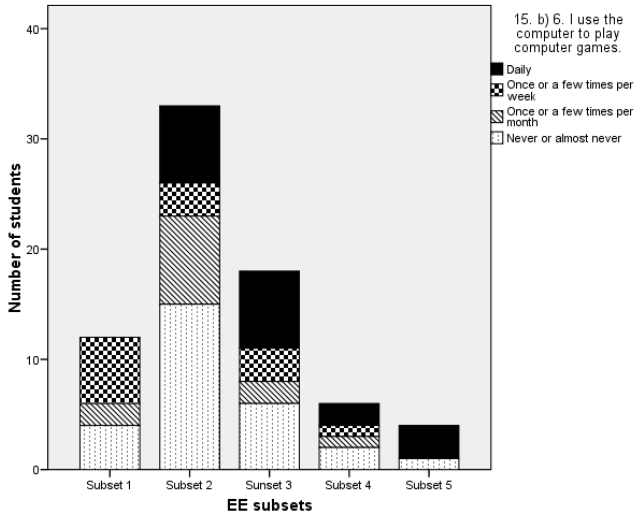


Figure 7.17. Responses to question 15 b, item 6, “I use the computer to play computer games” – the five EE subsets.

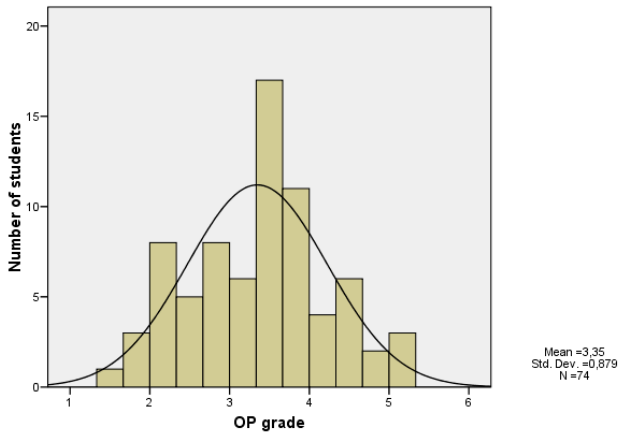


Figure 7.18. Histogram of the OP grade, with normal curve.

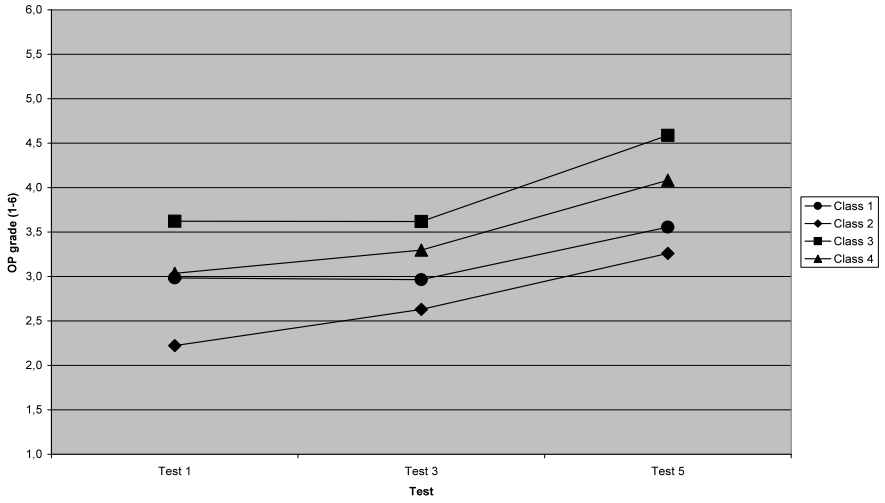


Figure 7.19. OP grade means for the four classes on Tests 1, 3, and 5.

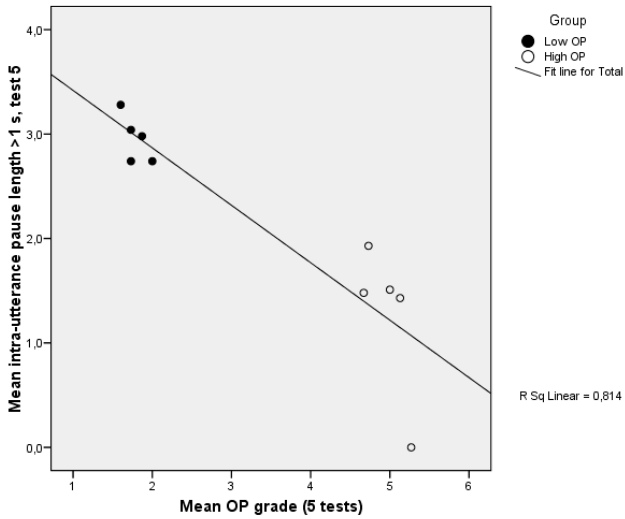


Figure 7.20. Mean OP grade (5 tests) versus mean pause length beyond one second (Test 5) for the Low and High OP group.

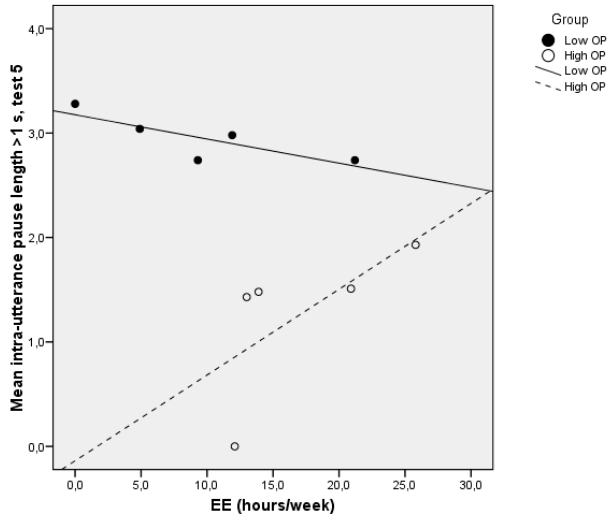


Figure 7.21. EE versus mean intra-utterance pause length beyond 1 second, with fit lines for the Low and High OP group.

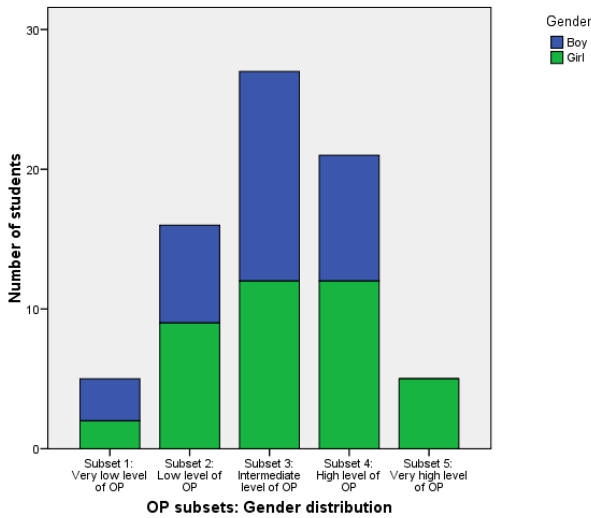


Figure 7.22. Five OP subsets and gender distribution.

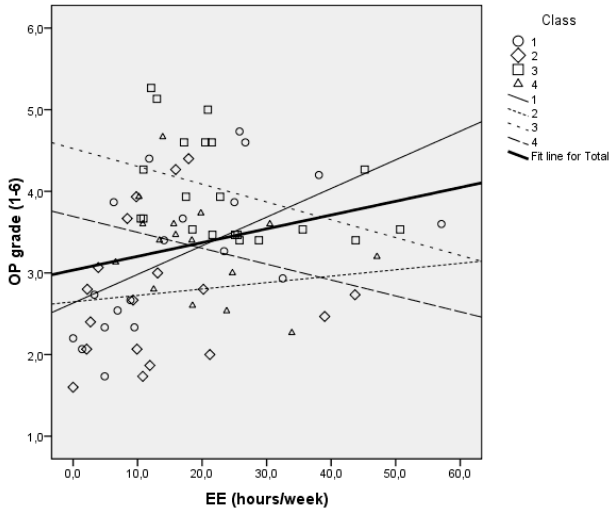


Figure 7.23. A correlation of students' EE and their OP grades, with markers for all students, fit lines for the four classes, and a fit line for the total sample.

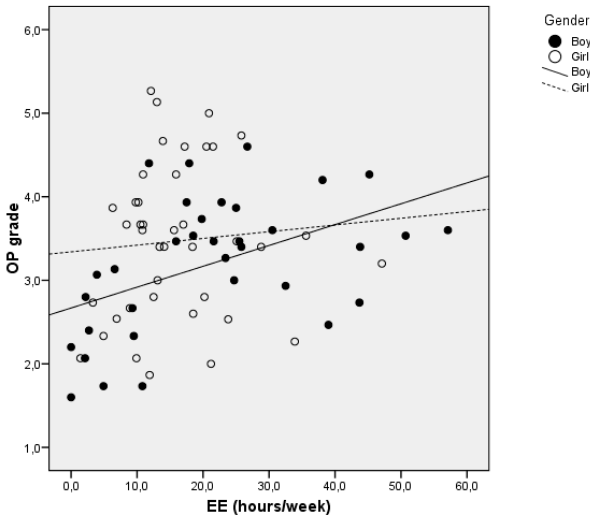


Figure 7.24. Mean OP grades and total EE (hours/week) for boys and girls, with added fit lines.

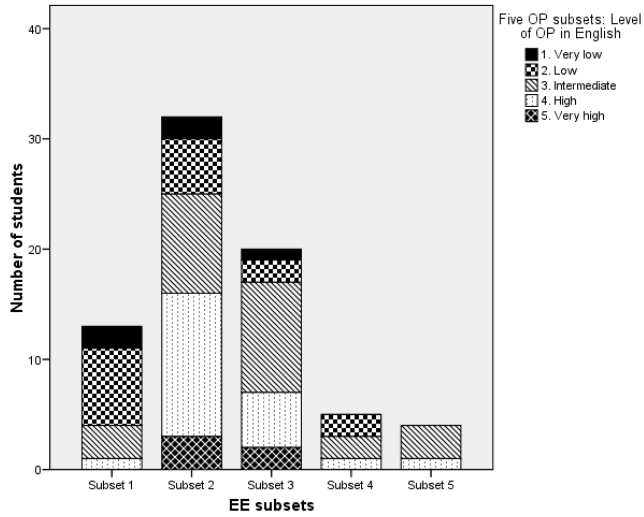


Figure 7.25. The five OP subsets dispersed in the five EE subsets.

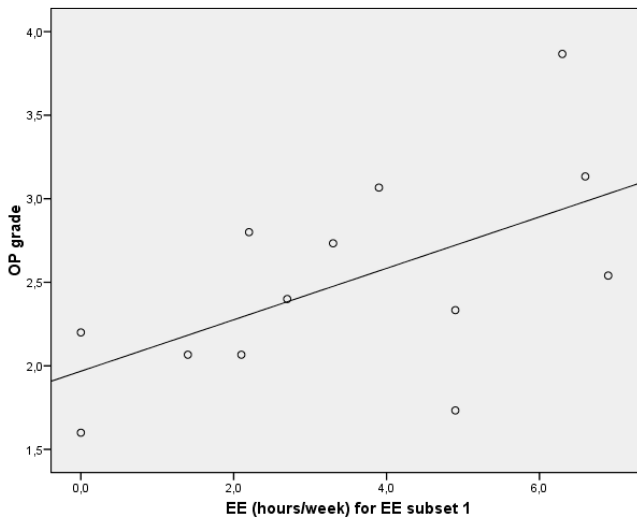


Figure 7.26. The correlation between EE and the OP grade for EE subset 1.

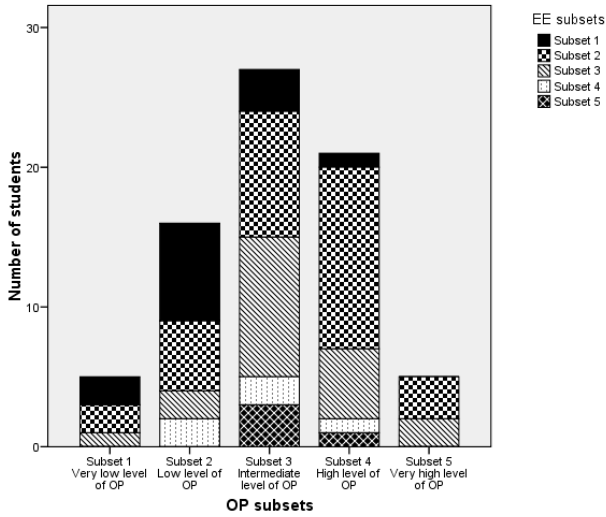


Figure 7.27. The five EE subsets dispersed in the five OP subsets.

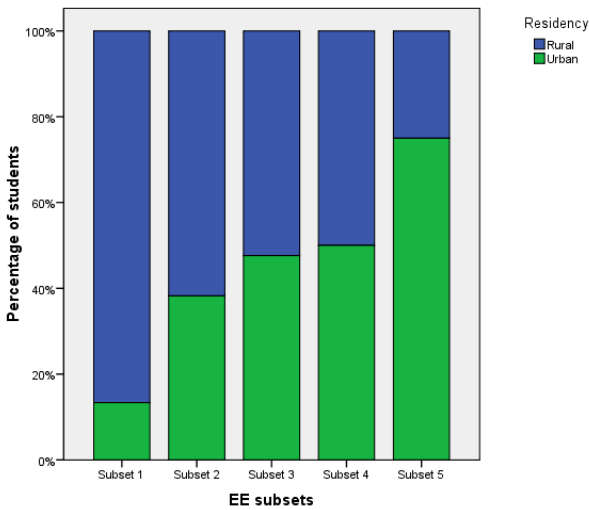


Figure 9.1. The dispersion of students depending on residency (EE subsets).

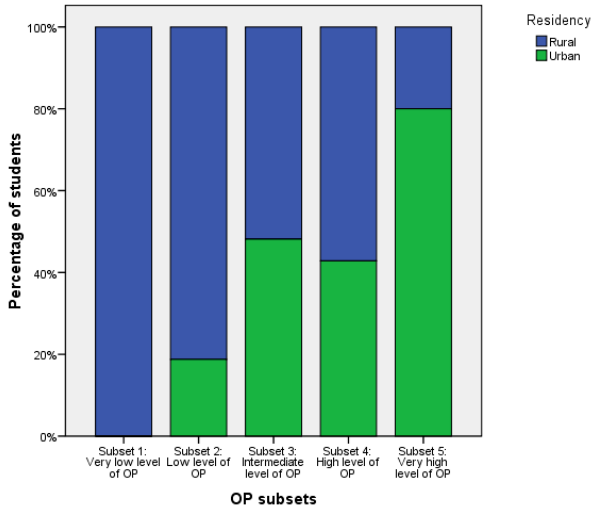


Figure 9.2. The dispersion of students depending on residency (OP subsets).

Appendix 16: Polysyllabic types and their etymology

Student ID	Polysyllabic types	Etymology			Total
		Germanic	Non-Germanic	Hybrid	
Low 1	animals		1		
	anymore	1			
	different*		1		
	disagree*		1		
	education*		1		
	environment*		1		
	family		1		
	restaurants		1		
	Total	1	7	0	8
% of polysyll. types	12	88	0	100	
Low 2	accepted		1		
	computer		1		
	disagree*		1		
	everywhere	1			
	family		1		
	violence*		1		
	Total	1	5	0	6
	% of polysyll. types	17	83	0	100
Low 3	animal		1		
	disagree*		1		
	important*		1		
	Total	0	3	0	3
% of polysyll. types	0	100	0	100	
Low 4	countryside*			1	
	grandparents		1		
	important*		1		
	understand	1			
	unusual			1	
	vegetarian*		1		
	Total	1	3	2	6
% of polysyll. types	17	50	33	100	
Low 5	anything	1			
	different*		1		
	difficult		1		
	disagree*		1		

Student ID	Polysyllabic types	Etymology			Total
		Germanic	Non-Germanic	Hybrid	
Low 5, ctd.	education*		1		
	everyone	1			
	Total	2	4	0	6
	% of polysyll. types	33	67	0	100
Low OP group	Total	5	22	2	29
	% of polysyll. types	17	76	7	100

* The word is given in the test instructions before it is used in the conversation by the speaker.

Student ID	Polysyllabic types	Etymology			Total
		Germanic	Non-Germanic	Hybrid	
High 1	animals		1		
	anything	1			
	area		1		
	countryside*			1	
	dangerous*		1		
	disagree*		1		
	discussion		1		
	everybody	1			
	everything	1			
	families		1		
	history*		1		
	important*		1		
	individual*		1		
	individually*			1	
	mobilephone*		1		
	mobilephones*		1		
	salary		1		
	trampoline		1		
	unfortunately			1	
	vegetarian*		1		
Total	3	14	3	20	
% of polysyll. types	15	70	15	100	
High 2	actually			1	
	animals*		1		
	anything	1			
	captivity*		1		
	different*		1		
	disturbing			1	
	education*		1		

Student ID	Polysyllabic types	Etymology			Total
		Germanic	Non-Germanic	Hybrid	
High 2, ctd.	everyone	1			
	everywhere*	1			
	example*		1		
	families		1		
	horrible		1		
	interesting			1	
	society		1		
	violence		1		
	whatever	1			
	Total	4	9	3	16
	% of polysyll. types	25	56	19	100
High 3	addicted		1		
	another	1			
	anyone's	1			
	bicycle		1		
	countryside*			1	
	dangerous		1		
	different		1		
	difficult		1		
	educated		1		
	education		1		
	environment		1		
	everyone	1			
	everything	1			
	family		1		
	fireplace			1	
	glass bottles	1			
	liquor store		1		
	living-room	1			
	mobilephones*		1		
	mud-houses	1			
	neighborhood	1			
	opportunities		1		
	otherwise	1			
	paper bags			1	
	privacy		1		
	recycle		1		
	recycled		1		
	tinnitus		1		
	together	1			

Student ID	Polysyllabic types	Etymology			Total
		Germanic	Non-Germanic	Hybrid	
High 3, ctd.	Total	10	16	3	29
	% of total polysyll. types	34	55	10	100
High 4	actually			1	
	anymore	1			
	area		1		
	civilized		1		
	connection		1		
	countryside			1	
	dangerous*		1		
	different*		1		
	easier		1		
	education*		1		
	everything	1			
	expensive		1		
	fashion clothes			1	
	forever	1			
	Germanish			1	
	Germany		1		
	grandchildren			1	
	massage therapist		1		
	mobilephones*		1		
	organization		1		
	otherwise	1			
situation		1			
teenagers			1		
Total	4	13	6	23	
% of polysyll. types	17	57	26	100	
High 5	accepted		1		
	another	1			
	apartment		1		
	cigarettes		1		
	dangerous*		1		
	difference		1		
	different		1		
	disagree*		1		
	emergency		1		
	everybody	1			
	everyone	1			
	excellent		1		
	execute		1		

Student ID	Polysyllabic types	Etymology			Total	
		Germanic	Non-Germanic	Hybrid		
High 5, ctd.	family		1			
	favorite		1			
	important*		1			
	mobilephones*		1			
	neighborhood	1				
	newspaper		1			
	reminding			1		
	represents		1			
	separated		1			
	solution		1			
	usually			1		
	violence*		1			
	Total		4	19	2	25
	% of polysyll. types		16	76	8	100
High OP group	Total	25	71	17	113	
	% of polysyll. types	22	63	15	100	

* The word is given in the test instructions before it is used in the conversation by the speaker.

Index

- Ability ... 12-13, 18-20, 25, 31, 39, 41, 43-46, 54, 60, 65, 91, 97, 99, 134, 175-176, 200, 202-203, 244, 247
- Academic skills..... 16
- Academic Word List 64
- Accent..... 1, 15, 35, 72
- Accuracy..... 5, 11, 42-45, 48-51, 53, 55, 94
- Acquisition/Learning Hypothesis..... 11
- Active vocabulary 60
- Add-on strategy..... 42
- Adolescents..... 1, 27, 29-30, 46, 202, 206
- Advanced vocabulary 8, 66, 68, 71-72, 100, 107, 134, 137, 157-158, 161, 164, 191, 198
- Affective Filter Hypothesis 13
- Aloud 1, 53, 84, 92-93, 134, 160, 241
- Anecdotalism..... 87
- Anglo-Norman..... 109
- ANOVA..... 55, 112, 120, 136, 167
- Anxiety .. 13, 16-19, 35, 110, 165, 175, 177-179, 188, 200-201
- Language anxiety 17-18
- Related to speaking 110, 165, 175, 177-178, 188, 200-201
- Applied linguistics... 10, 13, 33, 57, 89, 112, 114
- Aptitude..... 16, 18-19, 201
- Assessment 2, 7-8, 27, 39, 43-45, 48, 50-57, 59, 66, 68-69, 78, 86, 88, 95, 98, 100, 103, 115-116, 132, 137, 142, 145, 147, 178, 188, 190-193, 195, 201, 243-244
- Associative memory 18
- Attention . 15-16, 20, 24, 40, 45, 57, 72, 85, 182, 198, 202, 243
- Attrition..... 83-84, 87, 206
- External 79, 83-85, 206
- Internal..... 83-84, 90, 190, 206
- Aural comprehension..... 40
- Automatization 49
- Autonomic nervous system..... 17
- Avoidance 13, 18, 62
- Awareness 10, 14-16, 26, 33, 36
- Backchannels..... 44
- Backward linear regression analysis 115, 144, 146, 156, 194
- BALLI 27, 75, 99
- Beliefs about language learning . 7, 27, 75, 110-111, 175, 180-181, 183, 188, 201
- Bilingual..... 65-66, 84
- Bourdieu..... 22, 89, 170, 199
- Breakdown..... 14
- Channel-bound aspects..... 44
- Chi-square 113, 126, 129, 167, 181, 183
- Chomsky 12, 42
- Class 1..... 84, 95, 119-120, 123, 126, 131-132, 135, 137-138, 148-149, 151, 154, 168-169, 185, 193, 195
- Class 2 ... 120, 123, 126, 130, 132, 135, 137-138, 149, 152, 154, 168-169, 193, 195
- Class 3..... 80, 83, 93, 119-120, 123, 130-132, 135, 137-138, 147, 149, 154, 167, 169, 185, 192-193, 195
- Class 4 80, 83-84, 120, 123, 132, 135, 137, 147, 149, 151, 153-154, 167, 169, 184, 193
- Classroom..... 1, 3, 10, 17, 20, 24-27, 30, 36-37, 59, 69, 89, 177, 190, 205, 207
- Cloze tests..... 65, 78, 80, 98, 100, 147, 153, 162-163
- Rational cloze 65, 78, 97, 106
- Standard cloze 65
- Clustering..... 99-102, 122, 132, 140-141, 183
- Cognitive load 48-49
- Collocation 25, 59
- Colloquial..... 17, 25
- Communication 13-14, 17, 24, 27, 31, 39, 41-42, 44, 49, 51, 54, 56, 58, 69, 82, 96, 100, 167, 243-244, 246
- Communicative competence 12
- Communicative skills 3, 16
- Community..... 7, 10, 62
- Competence..... 12-14, 26, 34, 39, 44, 92
- Complexity 48, 55
- Compounds..... 24, 68, 108-109
- Comprehensibility 44-45, 56
- Comprehensible input 13, 61
- Comprehensible Output Hypothesis 14
- Compulsory school 3, 7, 27, 36, 168
- Computer..... 5, 29, 32, 36, 71, 119, 130-131, 161, 206, 240-241, 259-260, 267
- Computer games..... 36, 130-131, 259-260
- Consciousness 1, 11, 15, 25-26, 42
- Consonant-Vowel tiers 71
- Continuum..... 12, 60, 98, 110
- Controlled productive ability..... 65
- Core vocabulary..... 70, 73, 107, 109, 160
- Corpus..... 1, 18, 63, 66, 68, 70-71, 73, 105
- Correlation analysis 98, 113, 141, 143, 145-146, 154-155, 174, 177, 191, 193-194, 196, 203
- Cultural capital 10, 22, 89, 170-172, 199
- CV-tier theory 71
- Descriptive statistics..... 111, 122-124, 132, 140
- Discourse..... 13, 28, 42, 44
- Disyllabic words..... 71-72, 160
- Dysfluency 41, 49
- Normal dysfluency..... 41
- Education..... 3, 7, 21-22, 27-28, 32, 53, 65, 82, 84, 89, 97-99, 110, 125, 161, 165, 168-169, 171, 186-187, 198-199, 242, 267-270
- EE subset 1 .. 136, 142-143, 145, 154, 163, 174, 176, 178, 180, 187-188, 193, 196, 200, 202, 264
- EE subset 2 136, 143, 145, 153

EE subset 3.....	143, 145, 176, 187
EE subset 4.....	143, 145, 193
EE subset 5.....	142-143, 145, 151, 153-154, 163, 174, 178, 185, 188, 193, 196, 200
Effect.....	6, 20-21, 32, 35, 37, 45, 48, 50, 52, 54, 87, 90, 101, 104, 136, 141-142, 177, 180, 193, 202, 250
Ellipsis	42
English as a foreign language (EFL)....	8, 10, 25
English as a second language (ESL)	2, 10
English for Academic Purposes	33, 55
Error correction.....	11
Ethics.....	52, 75, 81, 85, 135
Etymology... ..	57, 72-74, 107, 109, 160, 164, 198, 267-271
EVA project	46
Evaluation.....	3-4, 17, 19, 33, 45-46, 110, 165, 171, 180, 190, 201, 243
Explicit learning	14-15
Expository speech	78, 80, 92-93, 241
Exposure.....	3, 11, 13-14, 25-26, 58
Extramural English (EE), definition	25
Extrinsic measure	66
Extroversion.....	16-17
Factorial grade.....	51, 53-54, 96-97, 103, 132, 137, 147, 154, 248-249
Fear	13, 17, 19-20, 177
Feedback	14, 88, 99
Film.....	1, 28, 77, 90-91, 117-118, 121, 124-127, 144, 146, 152, 156, 163, 191-192, 194, 197, 204-205, 208, 211, 233-234, 239, 241, 254
First Certificate in English	78
Fluency.....	1, 7, 32, 39-40, 44-49, 51-56, 59, 94, 96, 103-104, 106, 132, 137-139, 154, 157, 195, 243, 247-248
Broad sense	46
Narrow sense	46
Formulaic sequences	47, 49, 58
General Certificate of Secondary Education.	69
Generative theory	12
Germanic	73, 109, 160, 267-271
Goal attainment	3
Good language learners	30, 35
Grading criteria	4, 51, 54
Grammar.....	11, 12, 40, 42, 44, 54, 97, 243
Grammatical sensitivity.....	18
Guidelines	7, 80, 98-99, 244
Hawthorne Effect.....	52
Head word	61, 64, 67, 107
Hedges.....	42
Hesitation.....	46, 48, 244
High OP group	138-139, 145, 157, 159-162, 164, 195, 198, 261-262, 271
Hole-in-the-wall experiments	32
Holistic scoring	52-54
Hybrid.....	109, 160, 267-271
Idiosyncratic	35, 206
Immigrants.....	17, 35, 82-83, 121
Impact.....	2, 5-6, 32, 55-56, 140, 154, 172, 190, 193, 203-204, 206-207
Implicit learning.....	15-16, 198
Incidental learning	15
Independent samples <i>t</i> test..	112, 159, 170, 172, 179, 185
Index variable.....	115, 155, 175-178, 191, 196
Individual difference variables.....	10, 16-19
Inductive language learning ability.....	18
Inferential statistics.....	55, 111, 190
Information technology (IT)	2, 31
Inhibition	16-17
Input.....	8, 13-15, 25, 28, 31, 37, 61, 194, 247
Input Hypothesis	13, 31, 61, 194
Instrumental	17, 19, 184
Intake.....	15
Intelligibility.....	51, 56
Interaction Approach.....	14
Interaction Hypothesis	14, 42, 197, 203, 211
Interactional skills.....	40
Interface position	11
Interlanguage.....	12, 43
Interlocutor	35, 39, 41, 43-44, 47, 49-50, 56, 92, 103, 133, 160
Internalize.....	11-12, 42
Internet.....	1, 5, 8, 28-29, 31, 35-36, 90-91, 117-120, 124, 130-131, 144, 146, 152, 156-157, 163, 176, 179-180, 191-192, 194, 196, 203-206, 211, 233, 255
Interpretation.....	6, 9, 13, 29, 31, 57, 59, 63, 87, 103-104, 113, 141-142, 151, 155, 163, 175, 177, 184, 190-192, 194-197, 200, 203
Interrater reliability.....	54, 96, 103, 192, 248
Interval data.....	110
Intrinsic measure	66, 72, 108
Introversion.....	16-17
L2 acquisition ..	10-19, 21, 27, 30-31, 36, 38, 42, 88, 175, 178, 193, 198, 203, 205
Language diary	76, 80, 83, 85, 89-91, 100, 102-103, 116, 119, 121-122, 125, 128-131, 144-145, 152, 190-192, 197, 206, 239
Language proficiency	8, 12, 27, 32-33, 39, 59, 163
Latin.....	24, 73-74, 109
Learner autonomy	25-26, 30, 32
Learner language	12
Learner speech ..	8, 39, 41-43, 45, 48, 51, 55-57, 68, 72, 107, 158
Learner strategies.....	2, 75
Learning outcomes.....	2, 5, 8, 24, 33, 35, 37-38, 101
Leaving certificate.....	79-80, 98, 165
Lemma	63, 67
Lexeme	67
Lexical bundles	58
Lexical coverage.....	28, 205
Lexical diversity	62
Lexical Frequency Profile (LFP).....	63
Lexical individuality.....	69
Lexical item	18, 57-64, 66-67, 108, 157, 161, 164
Lexical richness.....	62-63, 69, 74, 108

Lexical sophistication.....	62, 68-69, 74, 108, 159
Lexical variation.....	62, 66, 72, 74
Lingua franca.....	27-28
Listening comprehension.....	3, 5, 39
Low OP group.....	138-139, 145, 157-162, 164, 195, 198, 268
Magazines.....	31, 34, 36, 90-91, 117, 124, 144, 152, 156, 168-169, 176, 187, 191-192, 194, 199, 203-204, 256
Mandatory.....	7, 20, 27, 79, 84, 92, 138
Measurement validity.....	86
Media.....	24, 27-30, 37, 124, 126-127
Memory.....	16, 18, 91, 134
Long-term memory.....	49
Short-term memory.....	16
Middle French.....	109
Mixed methods.....	6, 75, 86, 99-100, 190
Modern Language Aptitude Test.....	18
Monitor Hypothesis.....	13
Monosyllabic words.....	71-72
Mother tongue.....	82-83
Motivation.....	100, 110, 175
Instrumental motivation.....	17, 19, 184
Integrative motivation.....	19
Intrinsic motivation.....	19, 31, 33, 66, 72, 108
Multi-word expressions.....	58
Music.....	1, 5, 15, 22, 27-29, 31, 34, 36, 90-91, 117-118, 120, 124-127, 144, 146, 152, 156, 163, 173, 176, 191-192, 194, 199, 204-205, 240, 253
Native speaker (NS).....	1, 10, 31, 35, 47, 49, 51-52, 95, 105, 197, 243
Natural Order Hypothesis.....	13
Naturalistic learning.....	24
Negotiation for meaning.....	14, 42
Newspapers.....	30, 34, 90-91, 117, 124, 128-130, 144, 152, 156, 161, 168-169, 176, 187, 191-192, 194, 199, 203-204, 256, 271
Nominal data.....	110, 113, 181
Non-Germanic.....	73, 109, 157, 160, 267-271
Non-interface position.....	11
Non-native speaker (NNS).....	1, 47, 49, 51, 68, 95, 105
Normal distribution.....	96, 111-112, 116, 137, 154
Noticing.....	15, 59, 198, 203, 211
Null-hypothesis.....	6, 111-112
Old English.....	70, 73, 109
Old Norse.....	109
OP grade, definition.....	103
OP subset 1.....	143, 168, 177, 180
OP subset 2.....	178, 180
OP subset 3.....	143
OP subset 4.....	143, 167
OP subset 5.....	142-143, 167, 174, 177, 180
Oral fluency.....	7, 32, 39, 43, 45-49, 51-52, 56, 59, 94, 100, 103-105, 108, 116, 132, 139, 145, 191, 195, 198, 209
Oral proficiency, definition.....	39
Ordinal data.....	110, 114
Outline.....	8, 75, 79-81
Out-of-class learning.....	10, 24-26, 30, 32-34, 37
Out-of-school learning.....	8, 14, 25, 27, 32-33, 176, 181
Output.....	12, 14, 25, 40, 42, 49, 52, 54, 57, 61, 74, 198, 204, 205, 211
Output Hypothesis.....	14, 61, 211
Overall grade.....	21, 43-44, 51, 53, 54, 82-83, 96-98, 103, 135, 165, 170-173, 184-186, 188, 191, 243, 248
Paired samples <i>t</i> test.....	112, 135-136, 145
Parents.....	7, 22, 26, 78-79, 83-85, 90, 110-111, 135, 165, 168-169, 171, 184, 186-187, 198-199, 227
Passive vocabulary.....	60
Pauses	
Filled.....	53, 104
Intra-utterance.....	49, 100, 104-105, 137-139, 145, 195, 262
Length.....	7, 49, 104-105, 137-139, 145, 195, 261-262
Unfilled.....	49, 55, 104
Pearson correlation coefficient.....	96, 114
Peers.....	4, 17, 26, 31, 35, 50, 54, 88
Perception.....	17, 20, 49, 53, 87
Performance.....	11-12, 14, 40, 43, 48, 50-54, 87, 96, 98, 103, 192, 243-244, 246
Performance profile scheme.....	96, 192, 243, 246
Peripheral vocabulary.....	8, 57, 68-74, 86, 107, 109, 157-158, 161
Personality.....	10, 16-19, 35
Phonemic coding ability.....	18
Picture-cued story-telling.....	41, 47, 93, 133
Planned.....	24, 85, 206
Planning time.....	43, 48, 94, 134
Polysyllabic words.....	8, 54, 62, 69, 72, 94, 100, 103, 106-109, 147, 157-162, 164, 198, 207, 267-271
Polysyllabicity.....	66, 107, 109, 158, 163
Popularity.....	4, 117, 192
Pragmatic meaning.....	104
Prefabs.....	58
Prefix.....	24
Probability.....	107, 111, 135, 151
Productive Levels Test (PLT).....	59, 63, 65, 74, 78, 80, 83, 97, 106-107, 147-151, 153-155, 157, 163, 170, 173, 195, 250
Productive skills.....	40
Productive vocabulary.....	8, 57, 59-62, 65, 74, 97, 147, 150, 191
Pronunciation.....	1, 40, 41, 44-45, 49, 52-55, 70, 243-246
Prosody.....	1, 40
Radio.....	15, 22, 30, 34
Rarity.....	1, 2, 54, 63, 66, 69, 130
Rarity measure	
Extrinsic.....	66
Intrinsic.....	66, 72, 108
Raters.....	7, 49, 51-54, 57, 86, 88, 95-98, 103, 137, 154, 190-192

Reading comprehension	3, 5, 87	SPSS.....	89, 91, 94, 97, 102-103, 106, 110, 112, 115, 208
Reading tests.....	3, 48	Standard maximum tolerance	105
Receptive skills.....	40, 97	Statistical significance.....	111
Receptive vocabulary .. 60, 62, 74, 97, 147, 149-150, 191, 197		Stress.....	51, 68, 71, 172, 193
Referential meaning.....	104	Suffix	109
Reliability.. 6, 36, 63, 75, 86, 88, 91, 96, 99, 103, 106, 115, 145, 155, 171, 177, 185, 188, 192, 201		Suppletive inflection.....	67
Research validity	86	Swedish national test of English ...	97, 106, 153, 190
Residency ...22, 36, 110, 165, 172-174, 186-187, 198-199, 203, 265-266		Syllable margin	72
Retention.....	34	Syllable nucleus	72
Reticence.....	17	Syllable onset.....	72
Risk	17	Syntax	70
Role-play	41, 44, 53, 93	Target language (TL)..	11, 13-15, 18, 24, 26, 28, 32-33, 35, 42, 63, 198, 206
Second language acquisition (SLA)....2, 7-8, 10-13, 15-16, 18, 58, 156, 164, 181, 203, 206-207		Teacher.....	1, 3, 17, 24, 30-31, 57-58, 65, 76-78, 80-85, 90, 100, 135, 141, 182, 205, 251
Self-assessment.....27, 30-31, 37, 110, 165, 180, 182, 194		Teaching.....	1, 2, 7-9, 11, 13, 20, 26, 28, 39, 45, 64-65, 81, 92, 99, 205
Self-confidence.....	13, 17	Teaching methodology	2, 20
Self-directed naturalistic learning	24-25	Terminology ..	10, 12, 14, 24-26, 56, 67, 98, 160
Self-efficacy.... 16, 18-19, 31, 110, 114-115, 165, 175-178, 187, 200-203, 209-210		TOEFL	55
Self-esteem.....	13, 17	Transana.....	76, 94, 240
Self-instruction.....	24	Transcription.....	57, 94-95, 99, 105, 108-109
Shakespeare	67	Traveling	1, 82, 166-168, 187
Shyness	16-17	Travels abroad.....	22, 110, 124, 165-167, 171, 186, 198-199
Skewed distribution.....	116, 122, 178, 191	Triangulation	86
Smallwords.....	46-47	Trisyllabic words.....	72
Social interaction.....	17	T-Unit.....	53
Socioeconomic2, 82, 141, 166-167, 199, 202, 204, 207		TV1, 5, 8, 16, 22, 28-31, 34, 36, 77, 80, 90-93, 117-119, 121, 124-125, 130, 144, 146, 152, 156, 162-163, 176, 191-192, 194, 196, 204-205, 208, 211, 233-234, 239, 254	
Sonority	72	Type/token ratio	62-63, 66
Spare time	2, 4, 25, 29-32, 77, 80, 116-117, 119, 125, 128-129, 152, 168-169, 179-180, 190, 192	Unintentional learning	25-26, 37
Spare time English.....	4, 25	University Word List (UWL).....	64, 97, 106
Speaking		University-educated.....	28, 169, 187, 199
Basic types of learner speech.....	41, 93	Unplanned	43
Extensive speaking.....	41, 47, 52-55	Utterance.....	55, 104
Imitative speaking	41	Vague referencing.....	42
Intensive speaking.....	41, 53	Validity	14, 53-54, 63, 75, 86-88, 94, 99, 103, 171, 188, 201
Interactive speaking.....	41, 44, 47, 53-54	Verbal report	15
Responsive speaking	41	Video games	1, 5, 28-30, 34-35, 37, 90-91, 117-121, 124-125, 144, 146, 152, 156-157, 163, 173, 178-180, 188, 191-192, 194, 196, 198, 200, 203-206, 253
Speaking tests		Vocabulary Levels Test (VLT)	64-66, 74, 78, 80, 83, 97, 106-107, 147, 149-151, 153-155, 157, 163, 170, 173, 195, 251-252
Test 1.....	92-93, 133, 135, 145, 248-249	Vocabulary tests....	37, 59-60, 63, 76, 78, 87, 97, 100, 111-112, 115, 147, 149, 151-152, 162, 190-191, 195-197, 206, 251
Test 2.....	46-47, 93, 133-134, 248-249	Withdrawal.....	17
Test 3.....	92-93, 133, 135, 248-249	Word family.....	61, 106-107
Test 4.....	41, 92-93, 103, 133-134, 241-242, 248-249	Writing.....	1-5, 30-31, 33, 40-41, 44, 48, 55, 57-59, 62-63, 66, 68-69, 74, 85, 91, 97, 107, 119, 134, 197-198, 206, 243, 251
Test 5.....	84, 92-94, 97, 104, 107, 133, 135, 138, 145, 147, 157-161, 163, 195, 198, 248-249, 261	Yardstick	47, 51
Spearman correlation coefficient	114		
Specific referencing	42		
Speech length	105, 108		
Speech rate.....	46, 48-49, 53, 55		

Extramural English Matters

There is plenty of anecdotal evidence that young people in Sweden learn a great deal of English outside school. Empirical studies which examine the relationship between adolescents' out-of-school contacts with English and their learning outcomes in school are, however, very rare. The present study aims to fill that void.

Based on data from 80 Swedish ninth graders, i.e. boys and girls aged 15-16, Pia Sundqvist carried out a longitudinal study spanning one school year. In total, four school classes from three schools participated. The purpose of the study was to find out whether extramural English has an effect on learners' oral proficiency and vocabulary in English. Extramural English is defined as the English that learners come in contact with or are involved in outside the walls of the classroom.

With the help of a questionnaire and so-called language diaries, Sundqvist maps students' extramural English, both as regards the total amount of time spent on EE, and the type of activities engaged in. She also measures their level of oral proficiency and size of vocabulary. The students took five speaking tests in the course of the study, all which were recorded. External raters were engaged in order to assess the students' level of oral proficiency, using speech data from the tests. The assessment data were then used in subsequent analyses. An index variable based on the students' results on two written tests was used to measure the size of learner vocabulary.

Using inferential statistics, Sundqvist identifies which extramural English activities are more important than others for oral proficiency and vocabulary, respectively. The concept of extramural English is investigated from several perspectives and the combined results of the study provide an up-to-date picture of how extramural English matters.