Stellan Sundh

Swedish School Leavers’ Oral Proficiency in English
Grading of Production and Analysis of Performance
Dissertation for the degree of Doctor of Philosophy in English presented at Uppsala University in 2003

ABSTRACT


This study deals with the testing and grading of Swedish school-leaving students’ oral proficiency in English, and with certain aspects of these students’ linguistic competence. The analyses and results are based on material drawn from an assessment project carried out at Gothenburg University in 1993.

The 29 students taking part in the project were interviewed three times by three different interviewers in tests comprising three tasks, similar in structure but different in content. The interviewers were of three categories: school teachers of English, university teachers of English and native speakers of English. The student production was graded on a five-point scale according to a set of rating criteria.

The interviewers assigned generally positive but often differing grades to the student performance. The grades were influenced by the students’ ability to communicate and speak with flow, and by gaps in vocabulary and by occurrences of grammatical errors. The students’ use of discourse phenomena and compensatory strategies was also of importance to the grades assigned. Many students were considered to have acceptable intonation and rhythm, but nevertheless an evident Swedish accent. The linguistic features studied comprised the verbal group, vocabulary, discourse markers and pronunciation.

Differences could be observed between the members of the interviewer categories regarding the grades they assigned to student production. The school teachers seem to have paid special attention to grammatical accuracy, and the native speakers appear to have had a notion of communicative competence where accuracy plays a less important role.

Differences in the grades assigned could also be explained by the order in which the interviews were made, by some students’ hesitant delivery, by the positive or negative effect of various fillers in the students’ speech, and by the interviewing methods used by the interviewers in the tests.

Stellan Sundh, Department of English, Uppsala University, Box 527, 751 20 Uppsala, Sweden

© Stellan Sundh 2003

ISSN 0562-2719
ISBN 91-554-5681-2

Typesetting: Uppsala University, Editorial Office
Printed in Sweden by Elanders Gotab, Stockholm 2003
Distributor: Uppsala University Library, Box 510, SE-751 20 Uppsala, Sweden
www.uu.se, Acta@ub.uu.se
Contents

List of Tables and Figures .................................. 11
Tables ....................................................... 11
Figures ....................................................... 14
Preface ....................................................... 15

Introduction ................................................ 17
1. Preliminary Remarks ...................................... 17
2. The Background to the Gothenburg Project .......... 18
3. The Gothenburg Project .................................. 21
4. The Organisation of the Study .......................... 22

Part 1: Testing Oral Proficiency

CHAPTER 1: Historical Background and Previous Research .... 27
1.1. Measuring Student Performance ...................... 27
1.2. Oral Proficiency Testing ............................. 38
1.3. Linguistic Analyses of Students’ Proficiency ........ 44

CHAPTER 2: Methodological Considerations .................. 57
2.1. Introduction ............................................. 57
2.2. Oral Proficiency ........................................ 57
2.3. Reliability .............................................. 61
2.4. Validity ............................................... 63
2.5. Acceptability and Intelligibility ...................... 65
2.6. Concluding Remarks ................................... 67

CHAPTER 3: Formats: Tasks and Organisation ................. 68
3.1. The Oral Test: Definition .............................. 68
3.2. Three International Oral Tests: Descriptions and Comparisons 69
3.3. Discussion .............................................. 71
Part 2: The Aims of the Study and Hypotheses to Be Tested

CHAPTER 4: The Aims and Hypotheses of the Study 75
  4.1. Aims 75
  4.2. Hypotheses 76

Part 3: Grading Students’ Oral Performance and Analysing Their Competence

CHAPTER 5: Grading Student Performance 81
  5.1. Principles in the Assessment 81
  5.2. Global and Factorial Grades 82
  5.3. Sources of Variation 83

CHAPTER 6: Analysing the Students’ Linguistic and Strategic Competence 86
  6.1. Criteria and Principles 86
  6.2. Linguistic Features 89
  6.3. Classifying Student Performance into Test Rating Categories 92
  6.4. Issues to Take into Account: The Amount of Student Speech, the Length of the Oral Tests and the Correlation Coefficients 93
  6.5. Defining Standard English and Non-standard Instances 94
  6.6. Sources of Error in the Linguistic Analysis 94

Part 4: The Practical Organisation of the Testing Session

CHAPTER 7: Selecting, Testing and Assessing: Organising the Testing and Collecting the Material 99
  7.1. The Material of the Gothenburg Project 99
  7.2. Students 101
  7.3. Interviewers 103
  7.4. Tasks 105
  7.5. Rating Criteria 107
  7.6. The Organisation of the Testing Session 110
  7.7. Summary 112

Part 5: The Results of the Project

CHAPTER 8: Grades Awarded to the Students’ Oral Proficiency 115
  8.1. Global Performance Grades 116
  8.2. Communication and Fluency Grades 118
  8.3. Grammar and Vocabulary Grades 121
8.4. Pronunciation Grades ........................................ 124
8.5. Comparative Discussion of the Four Grades Given .... 126
8.6. Grades Assigned to the Individual Students .......... 130
8.7. Grades in Relation to the Length and Order of the Interviews 132
8.8. Grades in Relation to Background Variables .......... 135
8.9. Students’ Opinions about the Tests .................. 139
8.10. Summary ..................................................... 140

CHAPTER 9: The Individual Linguistic Features Included in
the Analysis: An Introduction ................................. 142
9.1. Verbal Group ................................................. 142
  9.1.1. The Third Person Singular Marker .................. 143
  9.1.2. The Progressive ........................................ 144
9.2. Vocabulary .................................................. 147
  9.2.1. Abstract Nouns ....................................... 148
  9.2.2. Modifiers of Adjectives .............................. 150
  9.2.3. Students’ Lexical Errors and Compensatory Strategies... 151
9.3. Discourse Markers ......................................... 157
  9.3.1. Well ................................................... 158
  9.3.2. You know .............................................. 162
9.4. Aspects of Pronunciation ................................ 165
  9.4.1. Intonation .............................................. 165
  9.4.2. Stress .................................................. 166
  9.4.3. Speech Rate .......................................... 167

CHAPTER 10: Verbal Group ....................................... 168
10.1. The Third Person Singular Marker .................... 168
  10.1.1. Overall Distribution ................................ 168
  10.1.2. The Occurrence of the Third Person Singular Marker
        in Individual Students’ Production .................. 169
  10.1.3. Linguistic Factors ................................... 170
  10.1.4. Occurrences of the Third Person Singular Marker
          in Relation to Interviewer Categories, Tasks and
          Background Variables .................................. 172
  10.1.5. Summary ............................................... 175
10.2. The Progressive ........................................... 175
  10.2.1. Overall Distribution ................................ 175
  10.2.2. Linguistic Factors ................................... 177
  10.2.3. The Occurrence of the Progressive in Individual Students’
          Production .............................................. 179
  10.2.4. Occurrences of the Progressive in Relation to
          Interviewer Categories, Tasks and Background Variables ... 180
  10.2.5. Summary ............................................... 182
CHAPTER 11: Vocabulary .................................................. 184
11.1. Abstract Nouns ..................................................... 184
  11.1.1. Overall Distribution ........................................... 184
  11.1.2. Linguistic Factors ............................................. 186
  11.1.3. Occurrences of Abstract Nouns in Relation to
           Interviewer Categories, Tasks and Background Variables . . 191
  11.1.4. Summary ..................................................... 193
11.2. Modifiers of Adjectives .......................................... 194
  11.2.1. Overall Distribution ........................................... 194
  11.2.2. The Types of Modifier ....................................... 196
  11.2.3. Occurrences of Modifiers of Adjectives in Relation
           to Interviewer Categories and Background Variables . . . . 197
  11.2.4. Summary ..................................................... 199
11.3. Students’ Lexical Errors and Compensatory Strategies ....... 199
  11.3.1. Occurrences of Lexical Errors ............................... 199
  11.3.2. Students’ Compensatory Strategies ......................... 201
  11.3.3. Summary ..................................................... 204

CHAPTER 12: Discourse Markers ........................................ 205
12.1. Well... .............................................................. 205
  12.1.1. Overall Distribution ........................................... 205
  12.1.2. Discoursal Factors ........................................... 206
  12.1.3. Occurrences of well in Relation to Interviewer
           Categories, Tasks and Background Variables ................. 207
  12.1.4. Summary ..................................................... 210
12.2. You know... ....................................................... 211
  12.2.1. Overall Distribution ........................................... 211
  12.2.2. Collocational and Discoursal Factors ....................... 212
  12.2.3. Occurrences of you know in Relation to Interviewer
           Categories, Tasks and Background Variables ................. 214
  12.2.4. Summary ..................................................... 217

CHAPTER 13: Aspects of Pronunciation ............................... 219
13.1. Intonation .......................................................... 219
  13.1.1. Results from the Observations of Students’ Intonation ... 219
  13.1.2. Summary ..................................................... 220
13.2. Stress: Some Relevant Issues ................................... 221
  13.2.1. Overall Distribution of Non-standard English Stress:
           Sentences and Words ........................................... 221
  13.2.2. The Occurrence of Non-standard English Stress in
           Relation to Background Variables ............................. 223
  13.2.3. Summary ..................................................... 223
13.3. Speech Rate ....................................................... 224
13.3.1. Overall Distribution ........................................... 224
13.3.2. Speech Rate in Relation to Interviewer Categories
and Background Variables. ......................................... 225
13.3.3. Summary ....................................................... 226

Part 6: Discussion

CHAPTER 14: Grading Swedish Students’ Oral Proficiency
in English ............................................................... 229
  14.1. Global Performance Grades ................................. 229
  14.2. Factorial Grades ............................................. 231
  14.3. Interviewer Categories ...................................... 232
  14.4. Individual Students’ Grades ................................ 234
  14.5. Results in Relation to Background Variables .......... 242
  14.6. Concluding Remarks ....................................... 243

CHAPTER 15: Swedish Students’ Oral Proficiency in English .... 245
  15.1. Verbal Group .................................................. 245
  15.2. Vocabulary ................................................... 248
  15.3. Discourse Markers .......................................... 251
  15.4. Aspects of Pronunciation ................................... 254
  15.5. Interference from Swedish ................................ 255
  15.6. Concluding Remarks ....................................... 257

Conclusion .............................................................. 260
  1. Summary and Conclusions .................................... 260
  2. Reflections ...................................................... 266

Bibliography .............................................................. 273

Appendix

APPENDIX 1: The Organisation of the Project:The Numbers Given
to the Tests .......................................................... 285

APPENDIX 2: The Distribution of the Interviewers in the Tests on
8 May, 1993 ............................................................ 286

APPENDIX 3: The Organisation of the Tests on 8 May, 1993 ...... 287

APPENDIX 4A: The Instructions and Tasks of the Students in Tests
with School Teachers ................................................. 288

APPENDIX 4B: The Instructions and Tasks of the Students in Tests
with University Teachers .............................................. 291
APPENDIX 4C: The Instructions and Tasks of the Students in Tests with Native Speakers. ......................... 294
APPENDIX 5: The Instructions for the Interviewers ............ 298
APPENDIX 6: The Questionnaire to the Interviewers ........... 299
APPENDIX 7: The Questionnaire to the Students ............... 300
APPENDIX 8: The Rating Criteria ......................... 301
APPENDIX 9: The Oral Mark Sheet. ......................... 302
APPENDIX 10: The Length of the Tests ..................... 303
APPENDIX 11: The Number of Students’ Words in the Tests. 304
APPENDIX 12: Illustrative Samples of Three Oral Tests ........ 305
Index ................................................................. 319
List of Tables and Figures

Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 6.1.</td>
<td>Sub-hypotheses of the ten linguistic features</td>
<td>91</td>
</tr>
<tr>
<td>Table 6.2.</td>
<td>Oral tests across the test rating categories</td>
<td>92</td>
</tr>
<tr>
<td>Table 7.1.</td>
<td>Students’ background variables</td>
<td>101</td>
</tr>
<tr>
<td>Table 7.2.</td>
<td>Students’ school grades in English</td>
<td>102</td>
</tr>
<tr>
<td>Table 7.3.</td>
<td>The rating criteria of Communication and Fluency versus the rating criteria of CEF</td>
<td>110</td>
</tr>
<tr>
<td>Table 7.4.</td>
<td>The rating criteria of Grammar and Vocabulary versus the rating criteria of CEF</td>
<td>111</td>
</tr>
<tr>
<td>Table 8.1.</td>
<td>The Global Grades of the 29 students (A-WW) by School Teachers (ST), University Teachers (UT) and Native Speakers (NS)</td>
<td>117</td>
</tr>
<tr>
<td>Table 8.2.</td>
<td>Factorial Grade 2a (Communication and Fluency) of the 29 students (A-WW) by School Teachers (ST), University Teachers (UT) and Native Speakers (NS)</td>
<td>120</td>
</tr>
<tr>
<td>Table 8.3.</td>
<td>Factorial Grade 2b (Grammar and Vocabulary) of the 29 students (A-WW) by School Teachers (ST), University Teachers (UT) and Native Speakers (NS)</td>
<td>122</td>
</tr>
<tr>
<td>Table 8.4.</td>
<td>Factorial Grade 2c (Pronunciation) of the 29 students (A-WW) by School Teachers (ST), University Teachers (UT) and Native Speakers (NS)</td>
<td>125</td>
</tr>
<tr>
<td>Table 8.5.</td>
<td>Number of tests for which the Global grades were the same as or different from the Factorial grades respectively</td>
<td>127</td>
</tr>
<tr>
<td>Table 8.6.</td>
<td>Factorial Grades 2a, 2b and 2c versus interviewer categories (School Teachers (ST), University Teachers (UT) and Native Speakers (NS))</td>
<td>128</td>
</tr>
<tr>
<td>Table 8.7.</td>
<td>Correlations between the categories of interviewers’ Global and Factorial grades</td>
<td>128</td>
</tr>
<tr>
<td>Table 8.8.</td>
<td>Number of students whose Global or Factorial grades were the same or different in their three oral tests</td>
<td>130</td>
</tr>
</tbody>
</table>
Table 8.9. Number of students who were assigned the same, two different or three different Global or Factorial grades in the three tests by the three different interviewers  

Table 8.10. Global grades assigned to boys’ and girls’ production 

Table 8.11. Global grades by age 

Table 8.12. Global grades by course programme 

Table 8.13. Global grades versus school grades in English 

Table 8.14. Global grades versus grades in the National Test of English 

Table 10.1. Uses of the third person singular marker by test rating categories 

Table 10.2. Proportions of the third person singular marker in individual students’ production 

Table 10.3. Proportions of non-standard English types of the third person singular marker by test rating categories 

Table 10.4. Proportions of types of subject in non-standard English instances of the third person singular marker by test rating categories 

Table 10.5. Instances of the third person singular marker ending in sibilants by test rating categories 

Table 10.6. The third person singular marker by interviewer categories 

Table 10.7. The third person singular marker versus the tasks of the tests 

Table 10.8. The third person singular marker in comparison with school grades in English 

Table 10.9. The progressive by test rating categories 

Table 10.10. The progressive simple and the progressive complex by test rating categories 

Table 10.11. Standard English and non-standard instances of the progressive by test rating categories 

Table 10.12. Types of standard English instances of the progressive by test rating categories 

Table 10.13. Non-standard English instances of the progressive by test rating categories 

Table 10.14. Non-standard English instances of the progressive by test rating categories 

Table 10.15. The progressive by interviewer categories 

Table 10.16. The progressive by the tasks of the tests 

Table 10.17. The progressive in comparison with school grades in English 

Table 11.1. Abstract nouns (tokens) by test rating categories
Table 12.12. The introduction of *you know* by students or by interviewers by test rating categories .......................... 215
Table 12.13. *You know* and the tasks of the tests by test rating categories .......................................................... 216
Table 12.14. *You know* by the students’ sex .......................................................... 216
Table 12.15. *You know* by school grades in English ............................. 217
Table 13.1. *And so on* by test rating categories .......................... 222
Table 13.2. The stress of *and so on* by test rating categories ........... 223
Table 13.3. “Pruned words” per minute by test rating categories .... 224
Table 13.4. Average scores of pruned words per minute by interviewer categories .............................................. 225
Table 13.5. “Pruned words” per minute by school grades in English 226
Table 14.1. Global and Factorial grades to the twelve students who were assigned remarkably different grades ............ 235
Table 15.1. Rejected or confirmed sub-hypotheses as regards the ten linguistic features ........................................... 258

Figures

Figure 8.1. Number of students receiving various mean Global grades ......................................................... 118
Figure 8.2. Number of students receiving various mean Communication and Fluency grades .................... 121
Figure 8.3. Number of students receiving various mean Grammar and Vocabulary grades ..................... 123
Figure 8.4. Number of students receiving various mean Pronunciation grades ........................................ 126
Figure 8.5. The distribution of Global and Factorial grades to the tests ......................................................... 126
Figure 8.6. Agreement or non-agreement in grades assigned to the individual students’ three tests by their three interviewers ......................................................... 132
Figure 8.7. Students’ Global grades versus length of students’ production in words ........................................ 133
Figure 8.8. Students’ Global grades versus length of student production in minutes ..................................... 134
Figure 8.9. Students’ Global grades versus the order of the students’ interviews ........................................ 135
Figure 8.10. Factorial grades for Pronunciation versus school grades in English ............................................. 138
At an early stage of my postgraduate studies in Uppsala, I wanted to write a thesis about Swedish students’ proficiency in English. Now that the thesis is completed, I wish to express my thanks to those people without whose support, advice and encouragement the completion of this dissertation would not have been possible.

This project would never have been started if it had not been for the help from Docent Torsten Lindblad at the Language Teaching and Testing Research Unit, the LTTC (Avdelningen för språkpedagogik), at the Department of Education and Educational Research, Gothenburg University. Docent Lindblad initiated and carried out the experiment from which the data used in the present dissertation emanate. I am grateful to Docent Lindblad for permission to use the material and for invaluable advice and perceptive criticism of various parts of the manuscript. Thank you, Torsten.

At the early stages of my research, the guidance and the encouragement of Professor Mats Rydén were valuable. The discussions with my former teacher Professor Gunnel Tottie inspired me to carry on my studies and also brought to my attention the fascinating characteristics of spoken English.

I would like to express my sincere thanks to my thesis supervisor Professor Merja Kytö. She has generously guided me from the start to the finish by making detailed and relentlessly accurate comments on the many draft versions. I have profited greatly from our discussions regarding the plan and method of my work. Her inspiring enthusiasm, friendliness and humour have greatly contributed to bringing this work to a conclusion. Thank you, Merja.

I would like to thank the persons who participated in the collection of the material of the Gothenburg project, both the 29 students interviewed and the nine interviewers. I am also grateful to Marianne Demaret, who transcribed all the student production.

The participants in the English Linguistics seminars at Uppsala University have been of tremendous help throughout the years. Docent Ingegerd
Bäcklund, Dr. Margareta Westergren Axelsson, Dr. Tore Nilsson and Docent Inger Ruin took a great interest in all my drafts and painstakingly read the final manuscript. I also thank Dr. Christine Johansson, Donald MacQueen, Göran Rönnéradal and Dr. Pia Norell for valuable comments on my drafts at numerous seminars. Their ideas for improving the text were of great help. Special thanks are also due to Dr. Pia Norell and Dr. Francis Nolan, Cambridge University, for participating in the investigation of the students’ intonation. I also wish to thank Donald MacQueen, who kindly went through my manuscript scrutinizing my English.

Warm thanks are due to Gudrun Erickson for valuable suggestions and comments and to Malin Tyllered, who read and commented on parts of the manuscript. I owe thanks to Dr. Per Malmberg whose comments led to a great improvement in several chapters. I also wish to thank Charlotte Lilja, headmaster at Richard Steffengymnasiet in Visby, for encouraging me to finish this work and for generously giving me permission to print numerous drafts.

For financial support I thank the Department of English at Uppsala University for their trust and help during my postgraduate years. I am also greatly indebted to Professor Erik Tengstrand’s fund for financial help. I would also like to thank Hans-Eric Ekengren, Utvecklingsavdelningen at Barn- och Utbildningsförvaltningen, Gotlands Kommun, for providing financial support for the printing of the manuscript.

Warm thanks go to Mrs Eivor Back, who helped me and my family with many practical arrangements so that I could work on the manuscript. Special thanks are due to Bengt and Svea Sundh, who also helped me to finish the writing.

Finally, my deepest gratitude goes to Rosita who showed her unfailing love and patience throughout the many years devoted to my doctoral work. I also thank Greta, Lydia, Natanael and Aron for all the inspiration they gave to me.

Uppsala June, 2003

Stellan Sundh
Introduction

1. Preliminary Remarks

The present study investigates the testing and grading of Swedish school-leaving students’ oral proficiency in English, and certain aspects of these students’ linguistic competence in English. The analyses and results are based on material drawn from an assessment project that will be referred to as “the Gothenburg project”.

For the purposes of the Gothenburg project, the data collection phase took place at Gothenburg University in May 1993 under the supervision of Docent Torsten Lindblad, head of the Foreign Language Teaching and Testing Research Unit, Gothenburg University, and initiator of the project. The 29 students taking part in the project were interviewed three times by three different interviewers in oral proficiency tests comprising three tasks, similar in structure but different in content. The student production was graded on a five-point scale according to a set of rating criteria. Each test was assigned overall (Global) grades and three Factorial grades (for the aims and design of the Gothenburg project, see Section 3 below).

Grading oral proficiency raises questions such as:

– What realistic ways are there of testing and grading oral proficiency in English?
– Is there a consensus on how to define “good oral proficiency in English”?
– If not, what do assessors disagree about?
– If yes, what can teachers do to help students improve their proficiency?

These sociolinguistic and pedagogic questions are fundamental to the present study which seeks to analyse and compare the grades given by the different interviewers to the student production in the oral tests and to investigate how some linguistic features produced by the students tally with the grades assigned to their production. At the same time it is of interest to scrutinize how the students use some particular linguistic features in their
production. The investigation of the students’ uses of certain linguistic features may provide interesting information as regards some aspects of the students’ spoken English. The aims and hypotheses of the study are elaborated on in Part 2 below.

2. The Background to the Gothenburg Project

As a background to the Gothenburg project and to the present study, a brief account of the status of English in the school system in Sweden and an overview of testing and test research are relevant. This survey is based on selected literature on the topic and personal communication with the leader of the Gothenburg project, Docent Lindblad (September 2002).

The status of English as a school subject has changed considerably over the last 40 years, in both compulsory schools and in upper secondary schools in Sweden. In 1962, the first Curriculum for Compulsory Schools (Lgr 62) was introduced. All children were to attend nine years of compulsory schooling, and English was introduced as the first foreign language in grade 4 and a compulsory subject for all children in grades 4 to 7. It was optional in grades 8 and 9 (Lgr 62). In the next curriculum, introduced in 1969 (Lgr 69), the study of English was introduced in grade 3 and was made a compulsory subject all the way through grade 9 (Malmberg 1985: 35). In the curriculum of 1980 (Lgr 80), pupils still started studying English in grade 3 (Marklund 1987: 142; Malmberg 1990; Lindblad 1982). In the most recent curriculum for the compulsory school system (Lpo 94) and the revised versions of the syllabi of 2000, individual schools decide when to introduce English in pupils’ timetables (Lpo 94: 6).

Changes concerning the teaching of English also applied to upper secondary schools. In 1965, new curricula for the upper secondary schools (Läroplan för gymnasiet 1965; Läroplan för fackskolan 1965) were introduced, and a few years later, in 1970, additional changes were made in the system of upper secondary education in Sweden (Lgy 70). The new gymnasieskola, as it is called in Swedish, included both the old upper secondary school (the gymnasium) and the vocational schools (fackskolan, previously yrkesskolorna). To begin with, the new gymnasieskola was organised in two-, three- or four-year-streams (Marklund 1987: 335). As far as English was concerned, there were then three different courses: in the two-year streams, there were two, one more advanced (known as särskild kurs, i.e.

---

1 In 1998/99, 32% of the pupils in grade 1 of the Swedish comprehensive school studied English (Skolverket 1999: 14).
2 The streams were study programmes of a theoretical or practical character.
‘special course’) and one less demanding (allmän kurs, i.e. ‘general course’), and in the three/four-year streams, there was a third one, the most advanced one and originally intended to prepare for higher education. This meant that the gymnasieskola in fact offered three different courses in English. The curriculum of 1994 (Lpf 94) changed the structure of the upper secondary school: all students now follow the same course of study at the lower levels. Regarding English, this is the so-called English A course. The more theoretically oriented students go on to English B and, in the last year, there is a more advanced elective course, English C. In 2000, the syllabi of English of the upper secondary school were again revised.

In the late 1940s and 1950s, there was a change worldwide in the view of language testing. New so-called standardized tests were introduced. In Sweden, they were first used in compulsory schools in mathematics and Swedish, later in English. The tests were set by the National Board of Education (Skolöverstyrelsen). The purpose of these tests was to help teachers give fair and nationally comparable grades. The construction and trying-out of these tests as well as follow-up work was entrusted to a group of education and subject specialists at the Teachers Training College (Lärarhögskolan) in Stockholm, and the tests were of course to be designed in line with the latest professional standards, such as ‘objectivity’ and the minimizing of chance factors, i.e. with a strong stress on reliability. This meant that the tests were of the multiple-choice type and the student’s answers were scored 1 (correct) or 0 (erroneous). As far as the teaching and testing of foreign languages, primarily English, are concerned, it is of interest to notice that at this time, there were also discussions concerning the significance of the oral exam in language testing and the desire to develop the oral part in the studentexamen\(^3\) (that is, the comprehensive school-leaving examination) (Thoren 1957: 136). Thorén also quotes a number of official documents in which teachers are recommended to use English as often as possible in their classes.

Later, the task of producing and following up the standardized tests was taken over by the National Board of Education itself, where a testing department was set up. The norm-referenced grading system used at the time prescribed the distribution of the five grades 1–5. They were to be given to, respectively, 7 – 24 – 38 – 24 – 7 per cent of the age cohort in the country. Guidelines for individual schools and teachers were provided in accordance with this and by means of nationwide test results. More or less identical standardized tests were used for 10 to 15 years.

\(^3\) The studentexamen (the baccalaureate) was taken at the very end of the last year of the gymnasium and consisted of two steps: written tests and a final oral exam. In this latter step, each student was tested in four or five subjects. This meant that only some students were tested in English. In the new kind of upper secondary school, there was no final examination.
Towards the end of the 1960s, the need was felt for new tests. The task of constructing these new tests was entrusted to the Foreign Language Teaching and Testing Unit, Gothenburg University, under the supervision of Docent Lindblad. Initially, the tests were of the multiple-choice kind and followed the dominating testing theories of that period (see for example Lado 1961 and Valette 1967). This meant that reliability (cf. 2.3. Reliability) was given high priority. Gradually the tests were revised, however, to include so-called open-ended items. In these tasks students were asked to write one word, a string of words, or full sentences. In the early 1980s, completely new tests were produced under the leadership of Bror Andered, a language-teaching expert employed at the National Board of Education. These tests were almost a kind of trend-breakers in the Swedish schools as they followed the language-testing theories and practices of that time. Besides reliability, there was now a greater interest in questions concerning validity (cf. 2.4. Validity). In these new tests, open-ended items with answers produced by the students were much more numerous (Lindblad 1990).

Great changes have also been introduced in foreign language testing in upper secondary schools. The old studentexamen disappeared in the 1960s when the new kind of upper secondary school (gymnasieskola) was created. New national tests were introduced and, to minimize the dramatic status of the old final tests, they were, as far as English was concerned, given in the last year but one. From the beginning these new national tests, originally only given in the three/four-year streams, consisted of nothing but multiple-choice items. In the 1970s, Docent Lindblad and his research team were entrusted with the task of producing tests for the two-year streams. Studies showed considerable differences in the mean achievement obtained for students enrolled in the different stream programmes when they were given the same tests (which was done as a large-scale research project in the 1970s), but considerable overlap in test results could also be noticed: some students of the two-year programmes even had better results than some students of the three-year programmes. After 1994, all students in the English A course took the same national tests.

According to Docent Lindblad, the view of languages as school subjects has changed radically over the last 30 to 40 years. He sees two major sources that have influenced the teaching and learning of foreign languages in Sweden: the American and the European influence. To quote Docent Lindblad,

\[\text{In the late 1960s, there were lively discussions regarding language teaching in Sweden. These discussions were of great influence for the developments in the fields of language teaching and testing in the 1970s and 1980s in Sweden. Summaries of these discussions are found in Edwardsson (1970).}\]
The American influence – to simplify things slightly for the sake of brevity – consists of the introduction of the concept of Communicative Competence as the main, but not the only, goal and objective of foreign language study. The European influence has come from the language project of the Council of Europe. This started out on a fairly small scale and its first widely known and discussed product was the book *The Threshold Level for Modern Language Learning in Schools* by J.A van Ek from 1977 (Lindblad, personal communication, 2002).

The ideas by van Ek were further developed, for instance, concerning definitions of communicative competence (van Ek 1986: 36) (cf. 2.1. Oral Proficiency) and regarding the psychological, emotional and social conditions of the language learner. The idea of adding this aspect to foreign language learning, and not only addressing the learners’ cognitive powers in language learning, has been of significance (Malmberg 2001). Furthermore, the Council of Europe made contributions regarding ideas and practical illustrations of classroom activities in foreign languages (Shiels 1988).

In the 1980s, Docent Lindblad and his colleagues in Gothenburg had the full responsibility for all national foreign-language tests. The research team then considered it necessary to launch studies of how to give and grade written and oral production tests, i.e. essays and oral proficiency interviews, and a ten-year project was started. During the first five years, starting in 1982, various topics were tried out in compulsory essays in the national tests, and instructions were given for the grading of the essays by a group of experts. During the following five years, from 1988, oral tests were made a part of the national tests. The tests were compulsory for the teacher and his/her class but only for a limited number of students in each class. During the last five years of the experiment, tasks for oral proficiency testing as well as detailed instructions for the grading of the students were produced and tried out (Lindblad 1992).

The oral part has been a fully compulsory part of the tests of the upper secondary school since 2000 (Erickson and Börjesson 2001), and in compulsory schools since 1998 (Erickson 1999).

### 3. The Gothenburg Project

As has been mentioned above, the material used in the present study was collected in 1993 by Docent Lindblad and his team of researchers at the Department of Education and Educational Research, Gothenburg University. In what follows, this material and the way in which it was collected will be described. The author of the present study participated in practical arrangements on the day the material was collected (May 8, 1993).\(^5\)

\(^5\) The practical arrangements included, for instance, showing the students to the interviewers’ rooms and assisting the interviewers with the tape-recorders.
A total of 29 students took part in the project. Each of them was tested three times by three different interviewers. Each oral test consisted of three different tasks, constructed along the same lines but using different topics (cf. 7.4. Tasks). Each interview lasted some 10 minutes. The students were assessed by three types of interviewers, School Teachers (ST), University Teachers (UT) and Native Speakers (NS). Consequently, the material to be presented and discussed in the present study comprises three assessments of each one of the 29 students in three separate tests, all in all 87 student tests.

The interviewers were instructed to give each student four different grades: one grade for the overall Global impression of the student’s performance, and three grades for three different factors according to the criteria the interviewers had to adhere to. These three so-called Factorial grades were given for Communication and Fluency, Vocabulary and Grammar, and Pronunciation.

The Gothenburg project material was thus collected in order to study the assessment of oral proficiency in English. Among the questions raised were, for instance, whether students’ demonstrated skills were assessed differently by the three different categories of interviewers, and if so, to what degree and in what respects interviewers differed from each other in their assessments.

The present investigation used the material of the Gothenburg project (a) to compare the grades given by the different interviewers, (b) to analyse the students’ uses of certain linguistic features in their production, and (c) to investigate the correlation between the occurrence of these features and the grades assigned.

4. The Organisation of the Study

The present study is an inter-disciplinary product that brings together the frameworks of linguistic and pedagogical study and is structured in six parts. In Part 1 (Ch. 1–3) the theoretical and empirical background of the study is dealt with. Apart from a brief survey of previous research, special attention will be paid to five central concepts in the discussion of language testing, namely proficiency, reliability, validity, acceptability and intelligibility. Part 1 also includes a comparison of the Gothenburg project test with three oral tests from other countries. Part 2 (Ch. 4) is devoted to a presentation of the aims and the hypotheses of the study. In Part 3 (Ch. 5–6), the principles of the analyses are clarified, first the principles of the investigation of the grading of the student production, then the principles of the analyses of the linguistic features. Part 4 (Ch. 7) includes a discussion of
the practical organisation of the testing session, with reference to the participants, the students’ tasks and the principles of the assessment. Part 5 (Ch. 8–13) presents the results, both as regards the grades assigned and the linguistic features studied. In this part, the results of the study of a number of factors that may have influenced the assessment, such as linguistic features related to accuracy and fluency and the different background of the interviewers are discussed. The results are expounded but not interpreted. Part 6 (Ch. 14–15) provides a discussion and interpretations of the results presented in Part 5. Finally, conclusions and some reflections are presented.
Part 1:
Testing Oral Proficiency

In this part of the study, three issues concerning oral proficiency tests will be elaborated on: historical background and previous research, methodological considerations and formats of oral tests. The purpose is to provide a background to the discussion of the testing and assessment methods used in the oral proficiency interviews of the Gothenburg project.
CHAPTER 1

Historical Background and Previous Research

The present chapter will describe the historical background of language testing, previous research regarding measuring student performance and certain issues with reference to the testing of oral proficiency. Furthermore, results from previous linguistic investigations not only of non-native students’ but also of native speakers’ use of certain linguistic features in the English language will be discussed. These results are relevant to discuss as a background to the principles adopted in the linguistic analyses of the student performance, for instance regarding the classification of “standard English” and “non-standard English” instances.

1.1. Measuring Student Performance

In this section the following issues are treated:

- the history of language testing,
- basic distinctions in measuring student performance,
- factorial assessments of students’ oral proficiency,
- students’ achievements in different skills, and
- students’ results in relation to the assessor’s background.

A Brief Survey of the History of Language Testing

By way of background to the principles adopted in the Gothenburg project (cf. Chapter 5), and in order to provide a historical context for oral proficiency testing investigated in the present study, a brief survey is given on the history of language testing. The discussion is based mainly on McNamara (1996, 2000).

To begin with, Spolsky (1976, 1985) distinguishes three stages in the development of language testing:
1. The pre-scientific or traditional phase
   The measurement was based on the subjective judgement of one examiner or a group of examiners.

2. The psychometric-structuralist or modern phase
   The measurement was characterised by efforts made to emphasize objectivity and reliability (cf. 2.3. Reliability). Knowledge of language was generally broken up into small parts and measured by standardised tests, for instance multiple choice items. Testing focused on the students’ knowledge of the grammatical system, of vocabulary and of aspects of pronunciation (McNamara 2000: 14).

3. The psycholinguistic-sociolinguistic or post-modern phase
   The main concern in the measurement was to secure not only objectivity and reliability, but also validity (cf. 2.4. Validity). The relationship between the students’ competence and performance was of interest, and it was important to take into account the variability of language use. The new orientation resulted in tests which integrated knowledge of the systematic features of language with an understanding of the situational context (McNamara 2000: 15).

Three different trends in the history of language testing are also distinguished by Simensen (1998: 268). They are the traditional trend of the 1950s and before (sometimes described as pre-scientific), the discrete-point trend in the 1960s and 1970s (the ideal was to measure isolated skills and small elements in separate tests) and the communicative and meaning-oriented trend from the 1980s onwards (the focus was, and is still, on measuring language in context). An interest gradually developed in the use of qualitative methods in language testing research in order better to understand students’ test-taking behaviours. The focus shifted to the student’s perspective; how did they arrive at correct or incorrect answers (Alderson 1993: 4)?

The communicative language tests of this late period displayed two principal features. Firstly, the tests were performance tests, i.e. the students were engaged in an act of communication. Secondly, the tests paid attention to the social context and the roles the students would have in real-life settings (McNamara 2000:17). There was thus a move from a focus on language structure and knowledge towards measuring students’ abilities to communicate. The emphasis was on the message as much as the form, and the students’ ability to convey meaning appropriately and fluently was crucial. Furthermore, there was a desire for authenticity in teaching activities and this desire had great impact on test content, too. The test content attempted to reflect the students’ likely future language needs (Alderson 1993: 9–11). In particular, it is the testing of oral proficiency that has changed:
Nowhere is the contrast between the old and the new approaches to language testing clearer than in the testing of speaking, or more precisely, of oral interaction. (Carroll 1985: 49)

Testing the oral skill in oral proficiency interviews, an approach initiated in US government agencies, such as the Foreign Service Institute (FSI), in the 1950s, spread to the assessment of foreign language skills at universities in the 1980s through the American Council of the Teaching of Foreign Languages (ACTFL). The oral proficiency interview (OPI) and its rating scale were used to determine the level of the students’ oral performance (McNamara 1996: 76; ACTFL 1989). Influences from these tests can still be found, for instance in the descriptions of rating scales. In describing the top level of the scale, the rating criteria of many rating scales make specific reference to the assumed performance of native speakers. This practice dates from the FSI Oral Proficiency Interview, where the highest level was defined as: “speaking proficiency equivalent to that of an educated native speaker” (Clark and Clifford 1988: 131; McNamara 1996: 184). The oral proficiency interview has remained the most commonly used means for the direct testing of spoken language skills (McNamara 2000: 23).

Basic Distinctions in Measuring Student Performance

The procedure of grading student performance can be organised in various ways and alternative methods can be selected. In what follows, twelve distinctions related to methods of measuring student performance are briefly presented. The presentation is, partly, based on the Common European Framework of References for Languages (2001: 183–192). The twelve distinctions are:

1. informal and formal assessment
2. norm-referenced and criterion-referenced assessment
3. mechanical and human assessment
4. simultaneous and delayed assessment
5. direct, semi-direct and indirect assessment
6. proficiency and achievement assessment
7. assessment using global and factorial grades (holistic/analytical)
8. performance and knowledge assessment
9. subjective and objective assessment
10. continuous and fixed point assessment
11. impression and guided assessment
12. self-assessment and assessment by others

1 In a review of research from 1965, the testing of spoken English is claimed to be growing in importance and in volume (Hitchman 1965).
Informal and formal assessment. Informal assessment is carried out on a daily basis through observations of how students cope in various types of oral and written activities. In formal assessment, various types of tests are used (Simensen 1998: 265).

Norm-referenced and criterion-referenced assessment. In norm-referenced assessment, the student’s performance is judged in relation to the performance of the other students in the same group. In criterion-referenced assessment, the student’s performance is measured against certain given criteria (Simensen 1998:266).

Mechanical and human assessment. In mechanical assessment of student production, responses are converted automatically to the appropriate score, whereas in human assessment, a certain degree of thought or judgement is necessary in order to arrive at the appropriate grade (Clark 1975: 12).

Simultaneous and delayed assessment. This distinction concerns the occasion of the assessment. Simultaneous assessment is carried out during or directly after the test. In delayed assessment, the test responses are evaluated on a later occasion.

Direct, semi-direct and indirect assessment. In the distinction between direct, semi-direct and indirect tests, the appeal to ‘real-life performance’ provides the basis for the distinction. Clark (1975:10) discusses direct tests and states that

in direct proficiency testing, the test format and procedure attempt to duplicate as closely as possible the setting and operation of the real-life situation in which the proficiency is normally demonstrated.

Direct proficiency tests should thus be very close to real-life language situations with respect to setting and language. Indirect proficiency tests do not have this requirement of a real-life situation and can be very different from the situations the students may encounter in real life:

Indirect measures are not required to reflect authentic language-use contexts and, indeed, they may in many cases bear little formal resemblance to linguistic situations that the student would encounter in real life (Clark 1978: 26).

The fact that language is both the object and the instrument of the measurement is also of significance in the discussion of these notions. A third level can be distinguished in so-called semi-direct tests, resulting in the three-level classification given above: direct, semi-direct and indirect tests (Bachman 1990:127; Clark 1979). The difference between direct, semi-direct and
indirect tests lies in the form of input used to elicit student production. Direct tests involve the use of the skills being tested. This is evident in an oral test or in a writing sample. In indirect tests other methods are used to check students’ knowledge, for instance a multiple-choice test to check vocabulary. Semi-direct tests are different from direct tests in the form of the input used to elicit student production. In direct tests, the human input is live whereas semi-direct tests could contain some kind of indirect input, for instance tape-recordings. It could be argued, however, that all language tests are indirect since they are “indirect measures of the underlying traits” (Bachman 1990: 33).

**Proficiency and achievement assessment.** Proficiency tests are based on a theory of language proficiency or a specific domain of content and look to the future situation of language use (without necessarily having any reference to the process of teaching). Achievement tests, also called attainment tests, are syllabus-based tests, drawing from the content of the course and are associated with the process of instruction related to the past (Bachman 1990: 71; McNamara 2000: 6–7; Alderson, Clapham and Wall 1995: 12).²

**Assessment using global and factorial grades.** In a system of global (holistic) grades, the interviewer gives one overall grade to the impression he or she has of the production as a whole. In a system of factorial (analytical) grades, several grades are assigned to the student production. The factorial grades may focus on, for instance, pronunciation, fluency or grammatical accuracy. Oller (1979: 321) discusses global versus factorial assessments and offers a third possibility. In this third type, the interviewer has a number of factorial grades to pay attention to according to the instructions but only gives a global grade in the final assessment of the student production. This method, i.e. analytic rating in which the scores for the separate aspects are combined into a single overall score, is now common (McNamara 2000: 43–44).³ Dyson (1972) compared the procedures of using exclusively global grades or only factorial grades in the assessment of oral proficiency in French. One of the purposes of the investigation was to see whether the latter procedure led to more objective and reliable grades. The study came to the conclusion that there was no significant difference in reliability between the two types of assessments (ibid. 1972: 35). Another study showed similar results (Walker 1981).

² Two additional terms can be distinguished. They are progress tests and diagnostic tests. Progress tests are given at various stages throughout a language course. Diagnostic tests seek to identify those areas in which a student needs further help (Alderson 1995: 12).

³ In the oral part of the Swedish National Test of English (Course A or B) of 2001, one Global grade is to be given, but factors are given in the instructions to facilitate the assessment. These factors are, for instance, Content and Language.
Performance and knowledge assessment. Performance assessment requires the student to give a sample of speech or writing in a direct test (see above). In knowledge assessment, the student answers questions to provide evidence of their linguistic knowledge and control.

Subjective and objective assessment. Subjective assessment is a judgement by one assessor, normally of the quality of a student’s performance. In objective assessment, subjectivity is controlled, often by means of indirect tests (see above) in which items have one correct answer, for instance, multiple choice tests.

Continuous and fixed point assessment. Continuous assessment evaluates performance which reflects a whole course or semester. Fixed point assessment is based on the performance on a particular occasion.

Impression and guided assessment. Impression assessment is a fully subjective judgement based on the student’s performance in class, for instance, and without reference to specific criteria. In guided assessment, the subjectivity is reduced by conscious assessment in relation to specific criteria.

Self-assessment and assessment by others. In self-assessment, the students make judgements about their own proficiency. In assessment by others, the judgements are made by an examiner.

In the following, the discussion will center on one of the above distinctions, namely assessment using global and factorial grades, with the focus on the use of factorial grades.

Factorial Assessments of Students’ Oral Proficiency
The combinations of factors that are used in factorial assessments of oral proficiency may vary (for the factors employed in the Gothenburg project, see 7.5. Rating Criteria and Appendix 8). Below follows a survey of such combinations used in research contexts or in actual assessments. The list of studies demonstrates the variability of the field.

Underhill (1987: 96) discusses possible factors in terms of ‘mark categories’ and distinguishes between traditional mark categories and the more modern style of mark categories. The former comprises, for instance, Grammar, Vocabulary, Style and Fluency, and Content. The latter has an emphasis on language as a tool for communication and the categories are described as ‘performance criteria’. They are, for instance, Flexibility, Hesitation, Repetition, and Independence. The purpose of the test decides which factors to use in the assessment.

Five factors were decided on by Haastrup (1977: 61) in a study of the
assessment of oral proficiency of English. The oral test was part of an investigation of the language proficiency and communicative competence of 45 teachers-to-be of English studying at the Royal Danish School of Educational Studies. The five factors used in this test were Pronunciation, Grammaticality, Vocabulary, Fluency, and Expressiveness comprising extra-linguistic features.

Higgs and Clifford (1982: 62) discuss three separate factors in the assessment of oral proficiency, which are referred to as the ‘functional trisection of speaking ability’. The three components are Function, Content and Accuracy.

Interactive skill is an example of another factor which may be included in assessments, and it is one of the factors of the performance to be taken into account in the assessment in the Canadian OTESL-test (St. John 1992: 310). Interactive skill includes the student’s ability to participate in the conversation in an active way with the help of responses such as backchannels, questions or comments.

Comprehensibility is another example of a factor. The American SPEAK-test added Comprehensibility to three common components, namely Pronunciation, Grammar and Fluency (Douglas and Selinker 1992: 319).

Six factors of oral proficiency were identified by Phillipson (1977: 74) in his analysis of oral compositions at the English Department of Copenhagen University:

1. Fluency (building sentences into more complex and larger units),
2. Accuracy (using grammar and vocabulary correctly),
3. Variety (extending the range of usage and idiom),
4. Intelligibility (using intonation with more expressiveness),
5. Pronunciation (standard British or American English), and
6. Speed (talking with ease and confidence).

Six criteria were also the basis for the global grades assigned in some experiments with assessment of oral proficiency in the Finnish Foreign Language Diploma for Professional Purposes. The experiments were carried out to create a nationwide system of certification for professional language competence. The six criteria were Pronunciation, Speech Flow, Grammatical Accuracy, Vocabulary, Appropriacy, and Discourse Management. The assessors were to use the approximate average of the factorial criteria when assigning the Global grade. Furthermore, six additional features were listed which were allowed to influence the Global grade. The six features were textural organization, dependence on the interlocutor for comprehension, relevance and adequacy of content, use of strategies to get around difficulties in communication and to enhance performance, mother tongue interference and interactiveness (Huhta, Sajavaara and Takala: 1993: 145–7).
Fulcher (1987) discusses the compromise between the “constructive interplay with unpredictable stimuli”, i.e. the ‘real life’ mirrored in the oral interview, and “the scientific measurement” with the assessment scales provided to the interviewer. In his description of the criteria of the Interview Assessment Scale of the English Language Testing Service (ELTS), five factors are distinguished: Repetition, Hesitation, Stumbling, Propositional Development and Grammatical Accuracy. These five factors and the assessment scale are compared with the language of a short extract of authentic conversation. The assessment scale has been criticized for laying too much emphasis on notional-functional considerations at the expense of discourse considerations.

In the Canadian OTESL-test, the examiners worked with two types of assessments that had different purposes. First they applied an assessment scale to show how well the students communicated. Then the assessors gave a diagnostic assessment of the students’ strengths and weaknesses. The second assessment was given to provide extra information about the student. Five factors were to be considered in this second diagnostic assessment: Listening, Fluency and Pronunciation, Interactive Skills, Content, and finally, Grammar and Vocabulary (St. John 1992: 310).

Walker (1983: 43) distinguished three factors worth taking into consideration in assessments:

1. linguistic factors (phonology, grammar and vocabulary),
2. paralinguistic factors (hesitations, voice quality and loudness, voice dynamics),
3. non-linguistic factors (the sex of the assessor and the student, the perceived personality of the student).

In a Dutch study on the assessment of oral proficiency in French as a foreign language, the conclusion was drawn that “attainments along three dimensions of oral proficiency should be reported separately: pronunciation, accuracy, and fluency”. These three dimensions were found to be positively correlated, but an overall score was considered to constitute “an unwarranted simplification” (De Jong and Van Ginkel 1992: 203).

McNamara (1996: 132) discusses ratings assigned to student performance as a function of the interaction of three factors. These three factors are the ability of the candidate, the difficulty of the task, and the characteristics of the assessor. In this model, information about each factor is taken from the data used; for instance, the characteristics of the assessor are determined with the help of other ratings given in a testing material, and the difficulty of a task is based on the performance of other candidates. The model, called the multi-faceted Rasch model, predicts the likely score for the combination of candidate, task and assessor, and also evaluates the ac-
The accuracy of the prediction. This model thus states that the likelihood of a particular grade on a task from a particular assessor and for a particular student can be predicted mathematically from the student’s ability, the difficulty of the task, and the severity of the assessor. In this way, the model seeks to move beyond raw scores and investigate the effect of the conditions of the assessment on the grades assigned.

The rating criteria in the assessments using factorial grades can be adapted so that they depend on the different test formats and the tasks. Chalboub-Deville (1995: 21) used different criteria for different tasks, for instance, “linguistic maturity” was used for interview and narration, “proper temporal shift” and “creativity” for narration only, and “ability to melodize the script to make reading meaningful” for read-aloud. Saleva (1997: 85) also used different criteria for different tasks in a study on 60 Finnish school-leavers given oral tests which contained five subtests. In the first subtest called “Reading a letter aloud”, the two criteria “pronunciation” and “fluency” were used, whereas, for instance, in the last one called “Reacting in situations and expressing opinions” only one was used, namely “appropriacy”. In the conclusion of this study, however, it was suggested that pronunciation and fluency are natural to use as criteria but “they need not be rated in every subtest” (ibid. 1997:145).

When factorial grades are used in the assessment of student production, one of the factors can be given special emphasis. Various methods have been used for the purpose of giving more emphasis to one or several of the factors in the overall score (Underhill 1987: 97). For instance, in the American FSI-test, a weighting is done with the help of a Conversion Table of the five factors: Accent, Grammar, Vocabulary, Fluency, and Comprehension. Grammar is considered to be the most important of the five factors in this test and is therefore given special emphasis by way of weighting (Wilds 1975:38–39; Oller 1979: 323). Another way of weighting factors in factorial assessments is by assigning a percentage to each subtest. In this way, the weight of each subtest can be decided in advance in proportion to the other subtests. In Saleva (1997: 86), there were five subtests, and the greatest weight was given to the most demanding subtest called “Presenting Finnish education” comprising the two factors “Transmitting information” and “Fluency”. A third method of weighing was used by McNamara (1996) in the Occupational English Test (OET), which is an Australian government

---

4 The Rasch model is a variant of methods which belong to what is known as Item Response Theory (IRT). According to the IRT, it is possible to separate the difficulty of the item from the ability of the student and thereby arrive at two separate measures: item difficulty and person ability (Alderson 1993: 3). Three main IRT models have been worked out: the one-parameter model which only takes account of candidate ability and item difficulty; the two-parameter model, which also takes account of item discrimination; and the three-parameter model, which in addition takes account of guessing (Alderson 1995: 91).
test of English for health professionals. The test is part of the registration for practice in Australia and has been the subject of considerable research. In the Speaking sub-test of the OET, six scales were used in the assessment: (1) Overall communicative effectiveness, (2) Intelligibility, (3) Fluency, (4) Comprehension, (5) Appropriateness of language and, (6) Resources of grammar and expression. The first category carried more weight than the others in the computation of the results by way of determining the student’s final score in the following way: the score of overall communicative effectiveness was added to the average of the scores of the five other categories (McNamara 1996: 112–113, 220). It was of interest to investigate what factor that was the most significant for the total score. The results showed that an interviewer’s “perception of the grammatical and lexical accuracy of a candidate’s performance is the most significant factor in the allocation of the candidate’s total score” (ibid. 1996: 218).

It is against this general background that factorial grades were used in the assessment of the Gothenburg project (see 5.2. Global and Factorial Grades, and 7.5. Rating Criteria).

The Relationship between Students’ Achievement in Different Skills

Previous studies of the relationship between students’ achievement in the four different skills, reading, writing, listening, and speaking are obviously of interest to a study like the present one.

When students’ achievement in different skills of English is compared, one can look at student performance in different tests which are designed to measure particular skills. In studies of students’ performance, there may be a distinction between mode and direction (De Jong and Van Ginkel 1992: 187). Listening and speaking both represent oral skills as opposed to reading and writing, representing written skills (mode). On the other hand, listening and reading are receptive skills, as opposed to the productive skills, writing and speaking (direction). Correlations in students’ results may thus be investigated of achievements in oral skills (listening and speaking), written skills (reading and writing), receptive skills (listening and reading) and productive skills (speaking and writing). According to Wood (1991: 231), evidence of substantial correlations between the oral skills (listening and speaking) “and other more general verbal abilities” comes from several sources, including studies of English as a foreign language.

Bachman and Palmer (1981) reported high correlations for foreign students between various measurements of speaking and reading. Five different methods of measuring reading were correlated with the corresponding methods of measuring speaking.

In a study from Australia, tests with separate scales of rating for the four skills (writing, reading, listening, and speaking) were used. Twenty stu-
dents, described as “moderately to well-educated” participated in these tests. Of these twenty students, only six were assigned the same grade in the four skills (Ingram 1984).

The results of students’ performance in reading skills and oral proficiency were compared in two other studies (Bachman and Palmer 1981, Reed 1992: 336–338). Reed used the written part in the TOEFL-test and compared its results with the results in an oral proficiency interview. He found considerable variation in the results. There was less variation, however, at the lower and upper ends of the rating scale. This would mean that there was a stronger correlation between the results in the tests of students who were at an advanced level or at a somewhat elementary level. The conclusion was that, generally speaking, there was a weak relationship between grades assigned in oral tests and those in written tests.

A similar tendency was shown in two additional studies: good results in a written test do not always predict good results in an oral test (St. John 1992: 312, Farhady 1982). These results would suggest that students may show varying results in tests of separate skills, especially when the students are at an average level in general.

However, the opposite tendency was shown in a study on immersion in Swedish schools. The results put forward by Washburn (1997: 229) gave support to the idea that a written test could predict the students’ oral proficiency. Washburn found that the National Test of English in Sweden was a reliable measure of the students’ oral proficiency. Tests of oral proficiency were not included in the score at the time. The National Test only comprised tests of vocabulary, reading comprehension, listening comprehension, grammar, and so-called “cloze” tests. Washburn’s study concerned Swedish students participating in immersion classes, and their oral skill was tested. The results were compared with the students’ results in the National Test. The results obtained by measuring their oral proficiency with the help of a limited number of linguistic features, such as grammatical correctness, syntax, and lexicon, showed no great differences compared to the results the students had obtained in the National Test.

Finally, regarding different results on different tasks included in oral proficiency tests, Pietilä’s study (1999) on the oral proficiency of Finnish students of English at the university level showed interesting results. In her study, the students did not display different degrees of skill in their grammatical competence, but did so in the “socio-pragmatic aspects of language use” that became apparent in the two tasks they were involved in (oral presentations and group discussions). Success in one task did not necessarily predict success in the other task. As for these advanced students, the

---

3 In cloze tests, words are regularly omitted and the student is asked to fill in the missing words.
conclusion was that they would benefit from increased awareness of various speaking conventions, such as conversational behaviour, and from increased opportunities to practise speaking.

Students' Results in Relation to the Assessor's Background
Chalhoub-Deville (1995) gives a summary of findings from a number of studies showing that, depending on their background, diverse assessor groups may differ in their assessment of students' oral proficiency. The background variables listed are, for instance, teaching and non-teaching assessors and whether assessors reside in the students' community or live in the target language community. Different tendencies could be observed in the various studies regarding the assessors' severity; whilst one category of assessors tended to be critical of students' linguistic abilities, others could be more tolerant. In Chalhoub-Deville's own study of six students of Arabic as a foreign language assessed by three assessor groups (82 assessors), the results showed that the three assessor groups emphasized three dimensions in different ways when assessing students' overall oral ability. The results were not consistent with results from previous studies. Galloway (1980) and Hadden (1991) have shown, for instance, that teachers tend to emphasize grammar, and non-teachers tend to be concerned with the more communicative aspects of language.

1.2. Oral Proficiency Testing
In this section, six issues are discussed with reference to the testing of oral proficiency:

- tasks in oral tests,
- the interviewer’s methods in oral tests,
- discourse structure and phases in oral proficiency tests,
- particular difficulties of oral proficiency testing,
- the length of oral tests, and
- students’ attitudes to oral proficiency testing.

Tasks in Oral Tests
The test method and the tasks are important in oral proficiency tests. Chalhoub-Deville (1995) emphasizes the influences of the tasks and the test methods on the results. Alderson, Clapham and Wall (1995: 62) also discuss the need of oral interviews to be carefully structured so that important

---

6 The three dimensions were (1) grammar and pronunciation, (2) creativity and adequacy of information in subject's narration, and (3) amount of detail provided (Chalhoub-Deville 1995).
aspects are covered with each student and each student is tested in a similar way. “A brief chat” is not enough to be able to assess a student’s level of proficiency.

Tasks which are to elicit talk may be compared in terms of difficulty. Brown, Anderson, Shillcock and Yule (1984: 64) make a classification of tasks of ascending difficulty. The easiest ones are static tasks, for instance describing a diagram, followed by dynamic tasks, such as telling a story. Abstract tasks are considered to be the most demanding ones, giving one’s own opinion, for instance. Obviously the tasks may also be easy or more difficult within one of the three categories.

Whether the tasks of oral tests are prepared or unprepared may be considered to be dependent on the students’ achieved level of proficiency. Prepared tasks may be the retelling of previously studied texts or various functional tasks which have been prepared. Unprepared tasks could be discussions of controversial issues and role-plays (Tornberg 2000: 157–158).

Weir (1990: 74–80) briefly lists some of the more useful tasks for testing speaking according to the criteria for communicative language testing. In what follows, eight types of tasks are described in order to give a picture of various tasks in communicative oral tests.

1. Verbal essay (the student is asked to speak on one or more specified general topics)
2. Oral presentation (the student gives a talk on a topic which he/she has prepared beforehand)
3. The free interview (a conversation which unfolds in an unstructured way; no procedures are decided in advance)
4. The controlled interview (a conversation in which there are procedures determined in advance for eliciting the student’s performance, for instance, a set of questions)
5. Information transfer: description of a picture sequence (the student sees a set of pictures depicting an ordered sequence of events and tells the story)
6. Information transfer: questions on a single picture (the student is asked about the content of the picture, the thoughts and attitudes of the people in the picture and future developments)
7. Interaction with information gap, interaction between two students or between a student and an examiner: completion of missing information (the student should complete the task by asking for missing information, for instance in a diagram or a set of notes)
8. Role play, between two students or between a student and an examiner (the student is expected to play one of the roles in an interaction)

The list and the terms are from Weir (1990: 74–80).
The ACTFL (American Council on the Teaching of Foreign Languages) oral proficiency interview is a widely used instrument to assess oral proficiency in a number of foreign languages. In this oral proficiency interview, the student should have the feeling of participating in a natural conversation, yet the ACTFL test is highly structured. This oral test includes three basic types of interaction (questions and answers, conversation, and role play) and four mandatory phases (the warm-up, the level checks, the probes, and the wind-down). The examiners are provided with extensive guidelines regarding strategies for eliciting speech, for instance the effect of the type of questions asked, and regarding the use of the package of role cards provided for the highest levels of the test. Furthermore, the ACTFL Tester Training Manual (Buck 1989) emphasises the use of natural conversational language and that “the transmission of meanings” is paramount. Experienced and trained examiners of the ACTFL interview are expected to think of the oral proficiency interview as a series of topics, rather than a series of questions. Conducting the ACTFL oral proficiency interview is described as highly demanding. The results from Saleva’s study on Finnish students in ACTFL oral interviews showed that “particularly in the case of advanced candidates, the linguistic demands combined with the situational complexity present so demanding a challenge to a non-native interviewer that both reliability and validity may suffer” (Saleva 1997: 143).

The tasks in oral tests could be expected to provide a representative sample of useful interactions in speech. In an Israeli study on 103 twelfth-grade-students, four experimental oral tests were used which represented four speech styles of different domains of oral proficiency. The four tests were the oral interview (a dialogue with the examiner), the role play (spontaneous speech behaviour within given roles eliciting speech functions), the reporting test (a monologue), and the group discussion (spontaneous discussion of a controversial issue). Each test was rated separately and with the same rating scale. The results of the study were, among other things, that there was a justification for the use of four different tests in order to tap overall oral proficiency and a wide range of speech functions, and to offer the student more than one opportunity to demonstrate his or her oral competence.

Tasks which are completed in language studios can be a means of ensuring that a range of situational contexts are covered. The Finnish Foreign Language Diploma for Professional Purposes contained an oral part (Huhta, Sajavaara and Takala 1993: 140–144). In this test, oral proficiency was tested with the help of a language studio test. The tasks in the studio tests were, for instance, reading aloud, reacting in simulated telephone con-

\[8\] The test is described in 1.1. Measuring Student Performance.
versations and various other contextual situations, and giving an oral presentation. In addition to the tasks of the language studio test, the students were engaged in a face-to-face interaction in two, or when necessary, three parts. The tasks were a formal presentation, which was followed by an interview or a discussion. The third task consisted of the student asking the interviewer about certain issues.

The Interviewer’s Method

As discussed above, oral tests may occur in an interactional context. The examiner in an oral test is not just the assessor but often also the conversational partner who is supposed to elicit a rich sample and support the student in his or her oral production. Since all students are to be given the same opportunities to show their competence, it may be argued that it is of importance that the examiners conduct themselves in similar ways (for the instructions for the interviewers of the Gothenburg project, see Appendix 5).

In a study on 58 Japanese language school students with varying levels of proficiency who were interviewed by ten native speaker examiners in a two-stage assessment of the CASE (Cambridge Assessment of Spoken English), the types and patterns of the interviewers’ support were investigated. The analysis was descriptive, focusing on the range of speech behaviours present in the material. Eight types of support were identified. They are described in what follows.

1. Priming topics. The interviewer introduced the topic to the candidate before the question was asked.
2. Supplying vocabulary or other collaborative completions. The interviewer supplied words or completed turns during a student’s answer.
3. Giving evaluative responses. The interviewer gave evaluative responses, such as “Good!” after a student’s answer.
4. Echoing and correcting responses. The interviewer repeated, as confirmation, or corrected the student’s answer.
5. Repeating questions. When the student had difficulties answering, the interviewer repeated the questions with slowed speech, more pausing, and over-articulation.
6. Questions followed by statements that only require the student’s confirmation. After the student had answered the interviewer’s question, the interviewer gave statements that merely required confirmation.
7. Drawing conclusions for the student. After the student had answered a question, the interviewer reacted with a conclusion based on that answer.
8. Rephrasing questions. The interviewer rephrased questions to check or facilitate comprehension.
The conclusions of the study were that processes of and practices in conversation obviously were present in the oral interviews, and that all the examiners supported the students to some degree. Differences could be observed in the types and frequencies of support employed by the ten examiners (Lazaraton 1996).

Young and Milanovic (1992) studied the discourse structure in the oral part of the Cambridge First Certificate Examinations. Their investigation of 30 examinations showed that students talked more than the interviewers did, a finding which was not surprising since the purpose of the exam was to make the candidates speak. The study also showed that only a minor part of the variation in the student’s oral production could be accounted for by the interviewers’ behaviour.

**Particular Difficulties of Oral Proficiency Testing**

Oral proficiency testing is today considered to be a valuable and useful means of obtaining a complete evaluation of a student’s speaking skill in English. Many teachers regard oral proficiency tests as natural parts in the assessment of students’ overall performance of a foreign language. The introduction of oral proficiency tests of English, to complement tests of essay writing and tests of reading and listening comprehension, has been a natural way of developing language testing at school.

Nevertheless, testing the ability to speak English has been, and is still regarded, to a certain extent, to be the most problematic area both in terms of practical arrangements and in terms of reliable assessments (cf., for instance, Weir 1990: 73–74; Shohamy, Reves and Bejarano 1986, Ericsson 1989: 399). Furthermore, testing the oral skill entails some particular difficulties that should be mentioned.

The oral proficiency interview is often carried out with the students’ language teacher at school, and there is thus a communication in a foreign language with individuals who have the same mother tongue. Particularly at elementary levels, the test may be constructed so that the examiner knows the answers to the questions which are posed, and this makes the conversation far from authentic. Furthermore, the conversation can be limited to deal with the students’ personal matters, such as his/her family and hobby. The student may be unwilling to talk about these issues but at the same time may lack appropriate means of expressions to discuss other topics. Ducroquet (1986) discusses these issues as practical problems in communicative oral language testing.

As to practical arrangements, oral interview tests are claimed to be highly time-consuming, whether they are carried out individually or in groups of students. The fact that the oral interviews require the presence of an assessor, often the students’ teacher, during the completion of the test
can make it a cumbersome means of evaluating students’ proficiency in English. In addition, oral tests require special practical arrangements in terms of rooms at schools; oral tests cannot, as other types of tests, take place in the same room in which other students are studying. Moreover, the fact that the oral test may be a new way of testing for many teachers involving the use of new methods and equipment could be problematic in some contexts and put new demands both on the school administration and on the teachers (von Elek 1995; Erickson 1996: 40; Olsson 1996).

The Length of Oral Tests
The length of the oral proficiency interview of the ACTFL (see above) is described as varying with the proficiency level of the student; the higher the level, the longer it takes to elicit an adequate sample of speech due to the extensive rating criteria of the higher levels. Furthermore, the instructions state that “unusual cases of various kinds, such as particularly shy or reserved students, need longer interviews”. In general, the range of the length of the tests varies from under ten minutes at the most elementary level up to thirty minutes in the most unusual cases (Buck 1989: 4-1). According to Underhill (1987: 40), oral tests are most often between eight and twelve minutes long. Factors that influence the length of the tests are the level of the student’s proficiency, the aims of the test, the personality and background of the student, and practical considerations such as time available.

It is worth observing that the length of tests may be expected to be of significance in the reliability of the test scores. The longer the test is, the more adequate the sample of the ability is expected to be (Bachman 1990: 220) (for the presentation of the length of the oral tests in the Gothenburg project, see 6.4. Issues to Take into Account: the Students’ Speech Rate and the Length of the Oral Test, and Appendix 10).

Students’ Attitudes to Oral Proficiency Testing
Students’ attitudes to certain types of tests or testing in general should not be neglected. The students’ attitudes and expectations could influence their test results. In a study of Finnish students, the results showed that the attitudes towards both learning and testing the spoken language were “strongly positive” (Saleva 1997: 139). Erickson and Börjesson (2001) state that students seem to appreciate tasks which are felt to be authentic conversation, and that generally and irrespective of age or level of language proficiency, they like tasks which include free oral and written production in the Swedish National Language Tests. In a report from the introduction of oral tests at a Swedish comprehensive school, the great majority of the students expressed in the questionnaire used that “oral tests are important or very important” (Olsson 1996). As regards the students’ attitudes to direct and indirect tests,
it seems that students have a positive attitude towards the former (Shohamy 1982). (For the students’ attitudes in the Gothenburg project, see the questionnaire in Appendix 7 and 8.9. Students’ Opinions about the Test).

1.3. Linguistic Analyses of Students’ Proficiency

An important part of the present study includes the investigation of the extent to which the assessment of oral proficiency can be influenced by students’ use of certain linguistic features (cf. 6.2. Linguistic Features). For that reason, this section discusses results from previous analyses of student proficiency, particularly with respect to the assessment of the linguistic features listed below and students’ uses of these features. The features are as follows:

- the verbal group
- vocabulary and compensatory strategies
- discourse markers: well and you know
- aspects of pronunciation

The reasons for including these particular linguistic features are partly that they are well-known difficulties for Swedish learners or foreign learners of English more or less generally, and partly that other researchers have found that they seem to influence the assessment of oral proficiency. (For a further discussion of the selection of the linguistic features, see 6.2. Linguistic Features.)

The Verbal Group

The third person singular marker. The third person singular marker is taught at an elementary level in Swedish schools, and teachers know that Swedish students’ use of it differs from the usage prescribed in grammars. Furthermore, accuracy in verb morphology is generally recognised as an index of development in second language acquisition (Bartning and Schlyter 1986: 9, Crookes 1989: 371).

The third person singular morpheme is also of interest due to the variation found in some varieties of English. In African-American Vernacular English and in dialects of the East Anglian area, there is evidence of native speakers omitting the third person singular marker with all persons (Labov 1972: 33). In western and northern dialects of England, on the other hand, the ending is added to all persons of the present tense on the verb (Trudgill 1990: 94).

Students’ spoken language often contains a higher proportion of grammatical errors than their written language. When students write, they have time to correct their production and consequently can be more accurate
than in spontaneous speech. A lower level of accuracy can therefore be expected in oral than in written data. In a study of grammatical morphemes, Tarone (1985) found lower levels of accuracy in speaking than in writing in a study of style shifting, and with a focus on three grammatical morphemes (the third person singular marker, the article, the plural -s) and one grammatical structure (direct object pronouns) in three different tasks. Students were least accurate with the third person singular marker in an oral narrative. Similar findings are reported in a study on Arabic-speaking students learning English (Scott and Tucker 1974: 79). Those results showed that verb errors had different frequency patterns in written and oral production. The third person singular verb was unmarked three times more often in speaking than in writing.

The grammatical function of a morpheme is also of significance for its occurrence and distribution in student production. In a study of Japanese students of English, the students were more inclined to omit the third person singular marker than the plural -s when the latter morpheme contained information. This suggests that the redundancy of the third person singular marker contributed to its omission (Saunders 1987: 261). The results are in line with what Young defines as “the functional hypothesis” (Young 1993). Furthermore, Larsen-Freeman (1976) found that language background (Arabic, Persian, Spanish and Japanese) did not have a significant effect on the way learners order English morphemes.

The fact that Swedish students, in contrast to native speakers, have difficulties with the prescriptive use of the third person singular marker is documented in several studies. In the production by native speakers of English, omissions of the third person singular marker can be regarded as, for instance, slips of the tongue, whereas the same phenomenon in a non-native speaker’s production makes a different impression on the interlocutor (Thagg-Fisher 1985: 42; Fromkin 1973: 266).

Although non-standard instances of the third person singular marker are regarded as serious errors (see below), they are frequent in Swedish student production. In a study carried out on Swedish university students of English, the category of singular pronouns combined with unmarked verb forms (*he go) was the largest category of “contiguous” concord errors. Plural pronouns with marked forms (*they wants) yielded far fewer instances. The uses with noun subjects were distributed evenly between unmarked and marked verb forms (Thagg-Fisher 1985: 39).

Even though some varieties of English contain non-standard instances of the third person singular marker, there is evidence that assessors are critical of students who do not follow the norm of the standard language. Native

---

informants showed this attitude in the study carried out by Thagg-Fisher (1985) on Swedish university students of English. The students participated in oral tests with two tasks: storytelling and description of pictures. In these tests, instances of concord error according to the pattern “singular pronoun /singular noun + plural verb”, i.e. the absence of the third person singular marker, were assessed on a scale whose polar terms were “completely natural and normal” to “extremely unnatural and abnormal”. Structures of the type “plural noun + singular verb”, i.e. where a non-standard third person singular marker was added, obtained the same assessment as those with the omission of a standard third person singular marker. The results led to the structures with non-standard addition or omission of the third person singular marker being categorised in the group “Extremely objectionable items” (Thagg-Fisher 1985: 182).

The progressive form. The acquisition of the progressive form in English may pose difficulties for a native speaker of Swedish. One of the reasons is probably that the construction does not exist in Swedish. The correct use of the progressive form, sometimes situationally bound, presupposes knowledge of aspect, which can be a very demanding concept to internalise even for the advanced language student. The considerable difficulties Swedish students have with the progressive aspect are discussed by e.g. Edström (1973: 128) and Axelsson and Hahn (2001).

In a school grammar of English used in Sweden, it is claimed that Swedes tend to use the progressive form too often (Hargevik and Hargevik 1993: 180). Hultfors (1986: 106) also points out foreigners’ tendency to exaggerate the use of the -ing participle. In Hultfors’ study, native speakers’ attitudes to the improper use of the -ing participle with the usually stative verb mean were investigated. The sentence “He is meaning this book, not that one” made a “rather strong foreign impression on the informants but did not cause much intelligibility difficulty” (Hultfors 1986: 186). The sentence was classified as more serious than most other errors.

Vocabulary

It may be assumed that students’ vocabulary is of significance for successful communication; the more words they know, the more likely they are to get their messages across. In spite of this assumption, students’ use of vocabulary seems to have been a neglected area of linguistic research (Jaatinen and Mankkinen 1993: 148). Vocabulary, as an open-ended system, has attracted less attention than more closed systems, such as grammar (Takala 1984: 1). Furthermore, vocabulary has not received much attention in teaching either. Syntactic structures to be learnt seem to be more easily specified in a syllabus than size, types or range of vocabulary (Carter 1987:
The importance of vocabulary in the assessment of student production cannot be neglected, however. Students’ lexical errors may be less generously tolerated than mistakes in syntax (Carter 1987: 145).

It is to be expected that an incorrect use of words could be an important negative factor in the assessment of oral proficiency. In a study of native reactions to errors in Swedish students’ English, Johansson (1978: 71) found that lexical errors caused a higher degree of irritation than grammatical errors. Johansson states that this is not surprising since “grammar deals with the rule-governed predictable aspects of language, the lexicon with the idiosyncratic, unpredictable aspects”. Similar results were found in a study of Arabic students of English and the assessments of their written production. Native speakers tended to judge semantic errors as more serious than syntactic errors (Khalil 1985). These tendencies were confirmed in a study of Swedish university students and their errors in written and spoken English: lexical errors were categorised as serious errors (Petti 1979). In that study, three categories of lexical errors were identified. The first category included “very common words” such as the distinction between teach and learn. The second category consisted of words that are almost identical in spelling in Swedish and English but have completely different meanings, for instance, novel in English and novell in Swedish (=short story). The third category included words that are or should be studied systematically owing to the differences in their uses and meanings in Swedish and English, for example the Swedish word skada, which can be translated in a number of ways, such as injury, damage, harm, mischief and pity. A scale of 0 to 3/4 was used to describe the seriousness of errors. Most of the lexical errors which could be categorised in the first category, “very common words”, were assessed as “unacceptable usage and felt to be a serious error”. They often caused irritation and/or misunderstanding.

In another Swedish study the conclusions were slightly different, however. Five sentences were used in order to test native reactions to lexical errors, for instance the inappropriate choice of the verbs make or do. The effect of the lexical errors was described to be of medium gravity (Hultfors 1986: 202).

The lexical errors described above are evidence of interference from Swedish. According to Corder (1981: 24), interference is defined as the procedure when “the student carries over habits of the mother tongue into the second language”. Lexical errors occur when students use patterns of vocabulary from their mother tongue in the foreign language. Evidently, interference may be either positive or negative. Positive interference leads to grammatical and appropriate use since there are similar patterns and structures in the student’s mother tongue and the target language. Negative interference leads to ungrammatical and inappropriate production.
Students’ vocabulary can be investigated not only in terms of correct and incorrect uses, but also from another perspective, namely whether the chosen words contribute to an impression of fluency in the assessment of students’ speech. In a 1994 study, the oral performance of Dutch students aged 11 to 12 years was assessed in terms of fluency and delivery in narrative discourse. Two variables were thought to be relevant for the prediction of the Global grades assigned. They were “lexical means for reinforcing or weakening narrative register” and “phonetic means”. A subcategory of the former variable was “reinforcers of register”. This term was based on a classification by Quirk et al. (1985) and included the subcategories adjectives and adverbs as modifiers (emphasizers, amplifiers, downtoners), verbs with a reinforcing function, reinforcing repeats, theatrical utterances, original metaphors, and lexical anaphors. The reinforcers of register were measured by counting all instances of the subcategories above; they emerged as one of the most powerful predictors of the ratings of delivery (van Gelderen 1994: 300). In this connection it is interesting to note that modifiers occur more often in the spoken than in the written language. In a study of two corpora,¹⁰ the distribution of the types and tokens of degree modifiers of adjectives was investigated. The degree modifiers included in the investigation were more common in the spoken than in the written data (Paradis 1997: 40).

Compensatory strategies. One important aspect of students’ lexical difficulties to be studied is their ability to cope with their inadequate knowledge of words. Strategy is the term that frequently emerges in discussions of these techniques. In what follows, definitions and descriptions of categories of communication / compensatory strategies are given in order to provide a background to the twelve categories of compensatory strategies distinguished in the present study (cf. 9.2.3. Students’ Lexical Errors and Compensatory Strategies).

Selinker (1972) was the first to discuss communication strategies: he used the term to account for certain errors made by students of a foreign language. In the students’ attempts to express themselves in spontaneous speech, these errors are consequences of inadequacies in their use of the foreign language. Corder (1981: 102) defines a communication strategy as:

\[\text{a systematic technique employed by a speaker to express his / her meaning when faced with some difficulty.}\]

Bialystok (1990: 1) defines communication strategies as attempts to overcome gaps between the students’ needs and their knowledge of English.

¹⁰ The London-Lund Corpus and the Lancaster-Oslo/Bergen Corpus.
These gaps can be of various types: a word, a structure, a phrase, a tense marker, or an idiom.

The term compensatory strategy is used by Lennon (1990: 396) to describe the methods students employ when they discover that their knowledge is not sufficient to communicate an intended message. Compensatory strategies are claimed to be distinguishing features of advanced students’ spoken performance. A compensatory strategy is defined as:

the strategies that learners employ, consciously or unconsciously, to communicate under circumstances of impending breakdown when their linguistic resources threaten to be inadequate for communicating the intended message.

Students may be either conscious or unconscious of their uses of strategies. Bialystok (1990: 3) discusses this characteristic trait, i.e. conscious or unconscious use of strategies and adds two other characteristics common to all communication strategies. Firstly, the speaker has a problem that may interrupt the communication. Secondly, the student also masters a number of strategies. With this perspective, the student may deliberately select one strategy before another.

A communication strategy is not only a characteristic of non-native speakers but may also emerge in communication between native speakers. Researchers have tried to detect differences between strategies used by native speakers and those used by non-native speakers. The results from several studies indicate that there are in fact few qualitative differences when native and non-native speakers try to solve problems in communication. Parallels can be drawn between strategies employed by learners of a language and for instance, mechanisms developed in pidgins to extend a restricted vocabulary (Poulisse 1989: 8).

Tarone (1977) classifies strategies into three categories: paraphrases, borrowing, and avoidance. Paraphrases include approximation, word coinage, and circumlocution. Borrowing comprises the subcategories of literal translation, language switch, appeals for assistance, and mime. Finally, avoidance includes either topic or message avoidance.

Varadi (1983) has a slightly different approach and bases his classification on the presumption that communication strategies involve message adjustment. He distinguishes message adjustment by way of reduction or replacement. Corder (1981: 104) also includes message adjustment in his classification, stating that students choose message adjustment strategies or resource expansion strategies. In the former category, message adjustment including risk-avoidance strategies, students tailor their messages to the

11 Bygate (1987: 115) uses the term of compensation strategies to describe “ways of communicating by improvising temporary substitutes when the speaker lacks normal language”.

49
linguistic resources that are available, for instance by topic or message avoidance or message reduction; owing to the lack of linguistic resources students choose to say something different than what was planned. The latter category, resource expansion strategies, which could be risk-running strategies, includes strategies where students attempt to increase their resources with the risk of failure in the form of misunderstanding or communication breakdown.

Faerch and Kasper (1980) describe a taxonomy of three categories of strategies with regard to whether students avoid difficulties or compensate in some way. The three general categories are formal reduction strategies, functional reduction strategies, and achievement strategies. In the first category, students avoid a linguistic item, for instance a word. In the second category, they avoid an entire speech act, for instance an apology. Finally, in the third category, they do not avoid or omit linguistic elements but compensate in the form of transfer, paraphrase, or word coinage.

Bialystok (1983) categorises strategies according to the language used. The first category is language 1 based strategies, which includes strategies where students switch to their mother tongue. The second category is language 2 based strategies, where strategies such as circumlocution are used. The third category includes strategies of paralinguistic features.

Strategies of interaction with the interlocutor may be classified in a separate category as types of achievement strategy. In the classification by Tarone (1977: 199), appeal for assistance is a category, and Faerch and Kasper (1980: 98) distinguishes cooperative strategies in their taxonomy. Interactional strategies are, for instance, appeal for help, and strategies when students try to talk about their difficulties in finding a word or arrive at solutions with the help of the interlocutor. Students’ talk about their language use was investigated in a study of Swedish students in immersion classes (Washburn 1997), mentioned above (cf. 1.1. Measuring Student Performance). The primary reason for using such meta-linguistic comments was vocabulary difficulties. Seven categories of compensatory strategies were identified, some of them involving meta-language: (1) the use of a Swedish word, (2) the abandonment of the message, (3) a request for help in Swedish, (4) a meta-linguistic comment in Swedish, (5) the use of an English paraphrase, (6) a meta-linguistic comment followed by an English paraphrase and finally, (7) a request for help in English.

Evidently, it would be useful to investigate the types of strategies used in relation to the grades given to students’ performance. In a study on school children in New York who studied French, the results showed that the number of hesitations, self-corrections, false starts, incoherences, and word-creations did not affect the teacher’s global rating of oral proficiency (Snow and Dolbear 1989). On the other hand, the possible positive effect of
the use of strategies for communication is pointed out by Seliger (1980: 90). Seliger states that some strategies act as clarification devices and are important for the understanding since the language produced by a student may contain elements that are strange to a native speaker. Hesitations, repetitions, false starts, and word searching could make the message clearer and facilitate understanding.

The next linguistic feature that is investigated in this study on the assessment of students’ oral proficiency concerns the use of certain discourse markers that are frequent in oral production.

**Discourse Markers: Well and you know**

Written language is organised so that meaning is made clear, for instance, by breaking up the text into manageable chunks such as chapters, subsections and paragraphs. Spoken language may be organised, for instance in adjacency pairs or with the help of various conversational continuatives or discourse markers. These items may indicate that the speaker intends to continue or that the speaker wishes to involve the hearer in the communicative process (Collins and Hollo 2000: 165). In what follows, the use of two discourse markers in spoken language is discussed.

**Well.** Well is a frequent discourse marker in spoken English. In the London-Lund Corpus (casual conversation) there was an average of one well for every 150 words (Svartvik 1980: 169). Stenström (1984: 143) found that well was the most common initiator of any move in a Question-Response exchange. Aijmer and Olsson (1990: 41) claim that well is the most common English “conversational phrase”, and that it occurs almost as frequently as the words is and was in English. Quirk et al. (1985: 444) emphasise the significance of discourse markers because of their high frequency as initiator in spoken English.

Bygate discusses four main ways in which speakers can facilitate production of speech. They are simplification, ellipsis, the use of formulaic expressions and the use of fillers and hesitation devices. The fillers and hesitation devices give speakers the time they need for the planning of the production. Well and you know (cf. below) are listed as frequent types of fillers (Bygate 1987: 18).

**Well** is attributed a range of characteristics and discoursal functions. Leech and Svartvik (1975: 114) describe well as an attention signal, a response to a statement to express surprise, pleasure, and/or regret. Stenström (1984: 139) classifies well as an initiator which has the main pragmatic function to reflect attitude. Schourup (1985) introduces the concept of evincive in a discussion of common discourse particles. He defines the evincive item as an item which indicates some form of unexpressed thinking on the part of the speaker. According to his definition:
a linguistic item that indicates that at the moment at which it is said, the speaker is engaged in or has just been engaged in thinking, the evincive item indicates that this thinking is now occurring or has just occurred but does not completely specify its content (Schourup 1985: 18).

One of the functions of evincives is backgrounding: the speakers show that they are or have just been thinking without exhibiting unnecessary details (Schourup 1985: 26). Well is described as an evincive used as an indicator of some internal consultation speakers wish to bring up (Schourup 1985: 90–91). The speaker has something in mind but does not say exactly what it is. Well may thus indicate that what is said has been subject to prior consideration. Other possible reasons for the occurrence of well in speech are, for instance, topic shifts, closings, or changes from introductory to topic talk (Schourup 1985: 73). Lakoff (1972) discusses the use of well when conversational rules are violated. This assumption is based on the theory that in conversation we comply with certain rules. These so-called conversational maxims are described, among others by Grice (1968) and Gordon and Lakoff (1971). Well, along with other items, is used if these rules are violated and the participants in the conversation discover, for instance, contradictions. Lakoff states that such particles serve as:

warnings to participants in the discourse that one or more of these rules is about to be or has been violated (Lakoff 1972: 916–917).

Well may also tell us something about the speakers’ attitudes either to their audience or to what they are saying. According to Leech and Svartvik (1975: 23) well can indicate such attitudes as agreement, reinforcement, or a positive reaction.

The survey above provides a picture of the fact that well is used in different contexts and for various purposes. It may be assumed that well is of significance for the impression of fluency in spoken production. This idea is further discussed in what follows.

In Riggenbach’s study from 1991, the results suggest that the use of filled pauses – and well could be a filled pause – tends to be one of the factors that contribute to an impression of fluency (Riggenbach 1991). In this study, English instructors rated six non-native speakers of English as either fluent or non-fluent. Differences were found in the kinds of filled pauses that were used by the two groups. Although the differences were not statistically significant, it was noteworthy that “fluent” non-native speakers used more so called higher-level fillers, i.e. lexical fillers, than did “non-fluent” non-native speakers, who tended to use non-lexical fillers and unfilled pauses (ibid. p. 431). The filled or unfilled pauses were considered the most important components for fluency in non-native speech.
In the light of the high frequencies and various functions of *well* listed above, it may be expected that instances of *well* can influence the interviewers’ impression of the students’ performance. For that reason, *well* is a highly useful linguistic feature to include in the present survey.

*You know.* *You know* is the second discourse marker of interest in the present study. When speakers use *you know*, they want hearers to adjust their orientation toward the reception of the message. Speakers use *you know* with an emphasis on the interaction. Östman (1981: 20) discusses this role of *you know* and describes it in terms of switching from deference to camaraderie. *You know* could be a signal of a more informal level of conversation and indicate to the interlocutor that the speaker wants to be taken as an equal partner in the conversation.

Another characteristic adhered to *you know* is its use for implicit transmission. The speaker communicates his attitudes and feelings to the propositional content of what is being said but does not add anything new to the meaning of the sentence by using *you know*. Östman (1981: 17) defines the core meaning of *you know* in the following way:

The speaker strives towards getting the addressee to co-operate and/or to accept the propositional content of his utterance as mutual background knowledge.

Using *you know* is a plea for co-operation in understanding or interpreting what is being said.

Quirk et al. (1985: 1481) list *you know* among the comment clauses that are used as direct allusions and appeals to the addressee. They are claimed to be characteristic of speech. Another interesting tendency worth mentioning is that *you know* is found to be used more by women than by men (Östman 1981: 72). *You know* is one of the hedges (other examples are *sort of, I think* and *perhaps*) which express the speaker’s certainty or uncertainty about the proposition under discussion. Previous research has shown that differences in men’s and women’s use of *you know* not only is a matter of frequencies but also of functions (Coates 1993: 116–118).

The position of *you know* in the utterance has also been investigated. The occurrence of *you know* as turn-initiator and in turn-final position is described by for instance Schiffrin (1987) and Östman (1981). Erman (1987: 50) found that a great majority (84.6%) of the instances of *you know* emerged in medial position. Finally, the contexts where *you know* occurs have been investigated. Östman found that *you know* was used “in particular in narrative parts of conversations” (Östman 1981: 16).

Considering the functions of *you know* listed by these researchers, it seems reasonable to assume that the occurrence of *you know* in the student production may be of significance for the assessment.
The fourth and final linguistic feature investigated in the present study concerns the students’ pronunciation.

Aspects of Pronunciation

In discussions of non-native speakers’ pronunciation of English, three issues should be taken into consideration. Firstly, pronunciation is geographically and socially less uniform than many other variables, such as syntax or lexis (Saleva 1997: 43). With the many speakers of English throughout the world, the question of norm is a matter of controversy (see for instance, Kachru 1992; Newbrook 1991). Secondly, the objective of the learning of pronunciation is of significance for testing, i.e., what is the standard required? Native-like fluency can be regarded to be the traditional aim. Van Els and De Bot (1987: 147) however discuss intelligibility and acceptability as desirable aims. If the two come into conflict, assessors’ decisions should be based on considerations of the aims of the learning. Thirdly, we do not know for sure about the role played by different errors of pronunciation for the assessment, for instance, errors in the pronunciation of individual sounds in comparison with errors in pausing or stressing (Lehtonen 1977: 41). Though the importance of different types of error is not clear, there is no doubt about the significance of pronunciation for the assessment in oral proficiency interviews. Mistakes of vocabulary and grammar can, at least partly, be avoided by efficient communication strategies, whereas pronunciation is present and is assessed throughout an oral test.

Intonation. Intonation is obviously important for meaning in many different ways since it has many different functions. At the same time however, intonation patterns are notoriously difficult to present concisely and accurately, partly because they are so deeply interwoven with information about the speaker’s attitude to the interlocutors. The number of possible correct intonation patterns is quite large (Minugh 1991: 66). Intonation can modify, or even contradict, the literal content of an utterance. In contexts where intonation and other elements come into conflict, the interlocutors’ inferences on the meaning are based on the intonation used (Hurley 1992). In investigations of non-native speakers’ intonation of English, there is evidence of how meaning is inferred from intonation. Seliger (1980: 89) discusses the use of rising intonation by students of English in one particular context. This intonation is used to indicate some form of hypothesis-testing which allows the student some additional time to understand what the partner in the conversation has said, as in the following example:

“– How do you feel as a foreign student?”
“– Foreign student? (rising intonation)”
The repetition of a part of the question with rising intonation is a form of check of what has been said. Intonation is used as a means of confirming the correctness of the student’s own hypothesis of his/her use of a word or structure. The word or structure is first uttered, as it were, for inspection and then uttered a second time to indicate its acceptability, the first time with rising intonation and the second time with falling intonation (Seliger 1980: 93). The rising intonation can be interpreted as a request for confirmation from the listener, similar to a question.

Although descriptions of intonation emerge in rating criteria, intonation is seldom tested by itself. Saleva (1997: 49) discusses a number of explanations for this fact, which can be summarised as follows. Firstly, given the prevalent communicative approach in teaching and testing at school today, the testing of single features, such as intonation, is rare. Secondly, the complexity of the phenomenon itself and thereby the insufficient capacity of the tester contribute to other approaches in testing. Thirdly, in spite of various attempts at systematization, no standard has emerged. Furthermore, formerly established theories change, for instance, the rule about *wh*-questions having a falling tone, which is now partly disputed (Gutknecht 1991: 266). Fourthly, it is difficult to assess a phenomenon whose effect is dependent on the context and the information that other elements, such as vocabulary or syntax, bring to the utterance.

*Speech rate.* Fluency is associated with a temporal element, and speech may be expected to have a regular pace and an even tempo. The Sajavaara-Lehtonen 1978 experiment, however, showed experimentally that speed was not constant across different discourse genres but varied and depended on the task (Sajavaara and Lehtonen 1978). In that study, students and native speakers were asked to read aloud simple and complex English texts and describe two sets of cartoons in English. In the reading tasks, the speech rate was faster than in the cognitively more demanding narration task.

Tauroza and Allison (1990) also showed that speech rate is different for different types of discourse in British English. Their results showed higher speed in, for instance, conversation than in lectures to audiences consisting mainly of non-native speakers of English. Their conclusion was that there is no correct rate of speech applicable to all situations.

Studies of non-native speech and ratings of fluency indicate that students’ speech rate shows a clear pattern of correlation to the grades assigned. Riggenbach (1991) found significant differences in a study of six non-native speakers and pointed out that the “chunking together” of disfluencies was important. A chunk of disfluency was defined as, for instance, “word search or ways of analysing, processing and observing what will come next”. A high number of disfluency chunks leading to a low speech
rate may indicate low fluency, whereas a low number of disfluency chunks leading to a high speech rate may indicate high fluency.

The results of a study of German students of English point in the same direction (Lennon 1990). The participants were four female German university students aged 20 to 24 years who were spending six months at the University of Reading. On their arrival in Reading, they had a multiple choice test which comprised grammar, vocabulary and reading comprehension. The four students got high scores on the test. Within two weeks of their arrival they were also asked to retell a story with the help of some pictures. Shortly before their departure they were asked to re-tell the six-picture sequence again. The assessments of fluency were carried out by nine native-speaker teachers of English. They were asked to assign fluency ratings to each recording. The results of Lennon’s study show that the assessors had no difficulty in deciding for each subject which was the earlier and which was the later recording. A majority agreed that for each subject the second rendering was more fluent than the first one. To investigate fluency closely, Lennon used twelve measures, and in three of them all four students changed their linguistic behaviour in the same direction. The three measures, namely speed, filled pauses and T-units followed by a pause are described below. First, Lennon distinguishes “pruned words” per minute. Lennon defines “unpruned words” as all the words produced by the speaker whereas “pruned words” exclude self-corrected words, repetitions, comments on the narrating task itself and words addressed to the interviewer. The second time there was an increase in speed. Speed was thus measured by counting pruned words in the production. Secondly, Lennon counted filled pauses per T-unit. In the second rendering there were fewer filled pauses. Thirdly, attention was paid to the percentage of T-units followed by a pause. Three students had a decrease the second time and one of them had the same score.

These aspects of students’ pronunciation, together with the other linguistic features discussed in this chapter, are components of students’ oral proficiency in English. In the next chapter, certain central concepts, such as oral proficiency, are discussed.
CHAPTER 2
Methodological Considerations

2.1. Introduction
The aim of the present chapter is to introduce five central concepts applicable to any kind of foreign language testing and necessary for the understanding both of the test format employed in the Gothenburg project and of the analysis of the results. As has been described above in the Introduction, oral proficiency is one of the main goals of foreign language syllabi and needs to be analysed in some detail. Secondly, the basic statistical and psychometric concepts of reliability and validity should be discussed with special reference to the testing of oral proficiency. Finally, the two concepts of acceptability and intelligibility, which are crucial to the assessment of any kind of language production, will be analysed and defined.

2.2. Oral Proficiency
In this subsection, proficiency, and in particular oral proficiency, is discussed. In investigations of students’ mastery of a foreign language, certain terms are used. Besides proficiency, these terms are, among others, students’ ‘competence’, ‘capability’ and ‘capacity’. They sometimes refer to the same phenomenon and they may also occur in similar contexts. Their definitions are often more or less overlapping, as, for instance, Widdowson’s (1983: 8) definition of ‘capacity’ which could very well be used to define ‘proficiency’:

The ability to use, produce and understand utterances by using the resources of the grammar in association with features of context to make meaning.

However, these concepts are often treated separately, with their own different definitions, as is the case in the following example where Ellis (1994:
720), as is quite common in foreign language literature, makes a distinction between proficiency and competence\(^1\):

Proficiency is the student's skill in using the second language. Competence is the knowledge of the language the student has internalised. Proficiency is thus the ability to use the knowledge in various tasks.

Proficiency, as well as the other notions mentioned above, is a broad term that may have many implications. Jones (1985: 7), for example, maintains that proficiency covers the entire spectrum from beginners to those who have a command of the language equivalent to that of an educated native speaker. Proficiency is also widely used in contexts of language testing but not always defined in detail. Farhady (1982: 44), for instance, states that language proficiency is one of the most poorly defined concepts in the field of language testing.

Two types of proficiency are distinguished by Cummins (1983). The two types are abbreviated BICS and CALP. BICS (=Basic interpersonal communication skills) is the proficiency used for oral fluency and in social contexts. CALP (= Cognitive academic language proficiency) is the linguistic knowledge and proficiency required for academic work. Proficiency can be regarded as two interacting continua comprising BICS and CALP (Cummins and Swain 1986).

Proficiency can also be defined with regard to the student’s ability to use his/her knowledge in specific contexts. According to this view, the student’s proficiency is a variable factor as opposed to the student’s competence (Ellis 1990; Taylor 1988). In a study of Finnish-speaking students’ oral proficiency in Swedish, oral proficiency was seen theoretically as the strategic use of the language user’s competencies in communicative language activities (Hildén 2000).

The modern overall concept used in describing students’ command of, or proficiency in a foreign language in, for example, syllabi and descriptions of tests and what to test, is communicative competence. In its earliest definition, communicative competence referred to the native speaker’s ability to produce and understand sentences which are appropriate to the context in which they occur (Hymes 1972). The notion, with its definition, spread to second language teaching and testing in the 1980s. In the definition most commonly used in the last ten years, communicative competence is considered to consist of several kinds of competence, namely

\(^1\) Cf. The important distinction between 'competence' and 'performance' by Chomsky (1965).
(1) linguistic competence (e.g. morphology and syntax),
(2) sociolinguistic competence (e.g. sociocultural rules), and
(3) strategic competence (e.g. the ability to make up for lack of vocabulary) (Canale & Swain 1980; Canale 1983).

In a discussion of the objectives for foreign language learning in a study from the Council of Europe, van Ek also defines communicative competence with the help of certain sub-competences (cf. Introduction and The Threshold Level). Van Ek lists six components of communicative competence: linguistic competence, sociolinguistic competence, discourse competence, strategic competence, sociocultural competence and social competence (van Ek 1986: 36).

Svartvik (1990) describes communicative competence as one of three levels in students’ mastery of a foreign language. The first level is the “survival kit-level”, the second level includes the skill to use language correctly, and the third level implies not only correct but also adequate language use.

Communicative competence has become the term most commonly used during the last 15 to 20 years both in syllabi and in research on language teaching and testing. Terms such as communicative language tests are often used (see e.g. Morrow 1979 and Weir 1988). The term Communicative Language Competence appears in Common European framework of reference for languages: learning, teaching, assessment (2001: 108–130) and in this context has three components: linguistic, sociolinguistic and pragmatic competence. Each competence is subdivided into specific categories:

– Linguistic competence: lexical, grammatical, semantic, phonological, orthographic and orthoepic.²
– Sociolinguistic competence: the mastering of linguistic markers of social relations, politeness conventions, expressions of folk wisdom and register differences.
– Pragmatic competence: discoursal and functional.

Besides communicative competence, Bachman’s (1990: 84) definition of the concept “communicative language ability” (CLA) is useful to include in discussions of oral proficiency since it is sometimes used in studies of students’ performance in English as a foreign language in at least two ways: CLA is either regarded as being equivalent to language proficiency (see e.g. Sasaki 1996), or it is used as the basic framework in analyses of the nature of oral proficiency (see e.g. Saleva 1997). Bachman also includes three components, only slightly divergent from those used in the above two definitions of communicative competence, in his CLA:

² Orthographic competence concerns the perception and production of the symbols of the written text. Orthoepic competence regards the correct pronunciation of the written form.
(1) language competence
   – organizational competence
   – grammatical competence
   – textual competence
   – pragmatic competence
   – illocutionary competence
   – sociolinguistic competence
(2) strategic competence, and
(3) psychophysiological mechanisms.

Each one of these has its own subcategories. Thus language competence (1) is divided into organizational competence and pragmatic competence. The former, i.e. organizational competence, includes grammatical competence (for instance morphology), and textual competence (for instance cohesion). The latter, i.e. pragmatic competence, is concerned with matters such as the ability to use language appropriately according to given contexts. Two categories of pragmatic competence are identified, namely illocutionary competence and sociolinguistic competence.

Strategic competence (2) is not only concerned with the ability to avoid threatening communication breakdowns but also includes the entire solving process. Finally, psychophysiological mechanisms (3) include the processes involved in the actual execution of the language, for instance to make sounds.

A central issue is whether the concept of proficiency in a foreign language can be defined or tested in terms of several abilities or whether it is impossible to isolate integrated skills, for instance, in a testing situation. As for oral proficiency, it is closely related to the concept of fluency. Speaking a foreign language fluently may take years to achieve. The struggle to think of what to say, to find the right words and to express oneself are clearly reflected in the way we speak. Brown (1977: 108) discusses the complexity of spoken production and states that even very few native speakers are entirely fluent when they speak spontaneously. Tornberg (2000: 137) also emphasizes that oral proficiency in a foreign language is very complex because of the fact that all aspects of the speaker’s competence (pronunciation, intonation, listening comprehension, grammar, vocabulary, sociolinguistic and sociocultural knowledge) must be activated simultaneously and without much time for planning.

Oral proficiency as a sub-skill of proficiency has been investigated by Higgs and Clifford (1982: 69). In their Relative Contribution Model of Speaking Proficiency, they hypothesize how the sub-skills of proficiency are proportioned at different levels of learning and using a foreign language. In this model, the five basic sub-skills are vocabulary, grammar,
pronunciation, fluency and sociolinguistic skills. Different percentages refer to the contributions of the five sub-skills to total language proficiency. Vocabulary has the highest percentage at a low level. Fluency and sociolinguistic skills show an obvious increase from the lowest to the highest level. At advanced levels, not only sociolinguistic but also paralinguistic competences can be included. This was evident in the Australian Second Language Proficiency Ratings (ASLPR). Proficiency in a foreign language was described not only as including the ability to communicate in particular situational contexts, but also to have the knowledge of the ways in which communication was performed. Paralinguistic and sociolinguistic aspects of the performance were to be considered in the testing situation (Ingram and Wylie 1993: 221).³

By way of conclusion, concepts such as proficiency and communicative competence occur in descriptions of students’ use of a foreign language. Oral proficiency was the aspect of the students’ overall English proficiency or communicative competence that was made the central concept in the tests of the Gothenburg project and was thus to be assessed.

2.3. Reliability

When a test is described as being reliable, this implies that the test measures whatever it measures in a reliable and non-random way. The student who knows the answer gives a correct reply, the one who does not know the answer gives an incorrect reply. If the test results depend on factors such as, students’ skill at guessing or teachers’ methods in correcting the answers, reliability can be described as low. Furthermore, the reliability of a test does not tell us anything of what is being measured, but only indicates whether this is done systematically or not (Lindblad 1991: 36).

The reliability of a test is thus a matter of the quality of test scores. Bachman (1990:161) suggests the following question as a way of identifying test reliability:

How much of an individual’s test performance is due to measurement error or to factors other than the language ability we want to measure?

Alderson, Clapham and Wall (1995: 294) present the following definition of test reliability:

Reliability is the extent to which test scores are consistent.

³ For discussions of students’ sociolinguistic competence and their social background as factors of their overall language proficiency, see for instance Janicki (1985) and Mays (1982).
Reliability concerns questions such as to what extent an individual student’s test results are due, for instance, to measurement errors, random variables or the structure of the test rather than the student’s language proficiency which is what we want to measure. The following definition of reliability appears in the Common European framework of reference for languages: learning, teaching, assessment (2001: 177):

It is basically the extent to which the same rank order of candidates is replicated in two separate (real or simulated) administrations of the same assessment.

Discussions of the reliability of a test thus concern the consistency of the results. Low reliability may be caused by the context of testing, the rating scales, time pressure, and factors such as the lack of clear instructions, or insufficient experience of the assessors, or weaknesses in the test itself (Bachman 1990:24; Alderson, Clapham and Wall 1995: 128). These factors are unrelated to the abilities to be tested, but they can nevertheless affect the test results. To achieve high reliability in the assessment of student production, the test constructor must be aware of these risks and do whatever can be done to minimize errors of measurement. In addition to the factors mentioned above, Bachman lists a number of factors causing errors of measurement (for instance, poor health, fatigue, lack of interest or motivation, and test wisdom) and groups them in three categories: test method facets, personal attributes and random factors. The smaller the number of such factors affecting test scores, the greater is the reliability of the test scores (Bachman 1990: 160–164).

There are different kinds of reliability depending on the kind of test used. So-called itemized tests, i.e. tests that consist of individual tasks that are marked as correct or incorrect are tested for reliability usually using a so-called test of homogeneity (with a formula known as Kuder-Richardson 21, KR21). The individual items are correlated with the test or subtest total, and items that show low correlations are excluded from the test after try-out runs. This results in values of .90 or, for certain tests, even higher correlations.

When it comes to tests that demand student production, such as oral proficiency tests, this kind of reliability coefficient can not be calculated. Instead one can either have the same interviewer grade the recorded interviews on two separate occasions. The correlation between these two gradings is referred to as the intra-rater reliability of the test. If instead two or more interviewers grade the same interview, the correlation between these can be calculated and is referred to as the inter-rater reliability of the test scores.

4 Facets of a test could be, for instance, characteristics of the testing environment and of the input.
A further possible source of measurement error is thus evident in test scores that are given subjectively. This pertains to inconsistencies in the raters’ assessment of oral interviews (Bachman 1990: 178).

As described above, two types of reliability can be described: intra-rater reliability and inter-rater reliability. Intra-rater reliability is a measure of the degree to which one single assessor assign two independent grades to a given sample of language production and inter-rater reliability pertains to an analysis of the grades assigned by several assessors of the same test (Alderson, Clapham and Wall 1995: 129). The consistency is thus checked across assessors. Effects due to sequencing, i.e. the order in which students’ performance is assessed, may influence intra-rater reliability, and this could possibly explain differences in grades assigned. For instance, if there are two low-achievers in a row, the next student after these two may be assessed more favourably simply because of the contrast to the two preceding students. Bachman (1990: 179) gives many examples of the effect of sequencing.

Factors other than sequencing that may influence reliability in the grades assigned to students’ oral performance are the difficulty and the variability of the tasks, the characteristics of the interviewers (personality, age or sex), and the interaction between the interviewers and the students (Brindley 1991: 156; Porter 1991). Topic choice may be another factor that could influence student production and thereby the grades assigned (Shohamy 1988). These factors may be regarded as being of minor importance for the reliability of the assessor’s subjective judgement of students’ oral performance, however, and oral proficiency interviews are also described as reliable assessments of students’ skill of speaking English (Magnan 1987).

Bachman (1990: 225) discusses so-called “performance” tests, such as oral interviews, with respect to reliability and asks whether grades can be interpreted as indicators of language competence or just regarded as the individual’s ability to perform well in a specific test situation. The variance component in oral interviews has been found to be substantial and in some cases even of greater importance than the components of the language ability measured (Bachman and Palmer 1981, 1982).

Reliability is considered in the present study by looking at the variation in the Global and Factorial grades assigned to the individual student’s production in three tests. It is then assumed that the individual student’s oral proficiency did not vary substantially during the day the tests were carried out.

2.4. Validity

It is not enough, however, for a test to have reliability as discussed above. A test must also give a correct picture of what one wanted to test or measure;
the test must have validity. The validity of a test is the extent to which it measures what one wants to measure and it is therefore related to the purposes of the test. There are different kinds of validity. If a test of English, for instance, seems to give an adequate picture of students’ proficiency in English, the test can be claimed to have high validity. This validity is used to be known as face validity.\textsuperscript{5}

A more scientific term is concept validity which means that the test gives a correct measure of a concept, for instance, the concept of oral proficiency, as in the present study. Various methods can be used to ascertain whether the testing procedure actually fulfils the demands of concept validity. Two other kinds of validity should be mentioned: prognostic validity, i.e. the ability of the test and the test results to predict the future progress of the students, and empirical validity, which is established through correlational studies between the test which is to be validated and some other, often well-established and generally accepted, measure of the same skill (Lindblad 1990: 278; 1991: 37).

Besides reliability, validity is thus an essential concept in measurement and testing theory. Reliability is closely related to validity: in order for a test score to be valid, it must be reliable. It should be noticed, however, that the opposite does not apply. The validity of a test is concerned with issues such as whether the test measures what it is supposed to measure, what abilities contribute to the variance in the results, and how much of a student’s test performance is due to the language abilities which are to be measured. Validity is always relative to the purpose of the test because it relates to the uses made of test scores and the ways in which test scores are interpreted (Alderson 1995: 296).

Henning (1987: 89) offers the following definition of validity:

Validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is purported to measure. A test is said to be valid to the extent that it measures what it is supposed to measure. It follows that the term valid when used to describe a test should usually be accompanied by the preposition for.

Bachman (1990: 25) also discusses the significance of the test interpretation:

While reliability is a quality of the test scores themselves, validity is a quality of test interpretation and use.

The following definition of validity appears in the Common European framework of reference for languages: learning, teaching, assessment (2001: 177):

\textsuperscript{5}Face validity may be regarded to be “a dead concept” in discussions of language testing. For a survey of a critical discussion of face validity, see Bachman (1990: 285–289).
A test or assessment procedure can be said to have validity to the degree that it can be demonstrated that what is actually assessed is what, in the context concerned, should be assessed, and that the information gained is an accurate representation of the proficiency of the candidate concerned.

In examining the validity of tests, the focus is thus on the tests or tasks that involve the ability tested. Furthermore, the content, the format of the testing procedure, and how the students perform are considered. Other matters of interest in the validation of a test are identifying the specific abilities the test is designed to measure and how the test is intended to be used (Bachman 1990: 236–238).

A number of types of validity have already been described above. These different types can be considered to be different methods of assessing validity (Alderson, Clapham and Wall 1995: 171). For interesting further discussions of the types introduced above and definitions of other types of validity, see among others, Alderson, Clapham and Wall (1995), Bachman (1996), Lindblad (1990), McNamara (2000) and Wood (1991).

In the present study, validity is considered from two angles. Firstly, the focus is on the relationship between the students’ test performance and their performance in other contexts (for instance, comparisons between the Global grades assigned in the oral tests of the Gothenburg project and the students’ results from the National Test of English, although their oral proficiency was not tested in this test). It could be expected that if the grades of the oral tests correspond to the students’ results from other contexts, such as their school grades in English and the results from the written parts of the National Test of English, it seems reasonable to assume that the grades assigned to their performance in the oral tests are based on their language proficiency. Secondly, the Global grades assigned to the student performance are compared with the occurrence of certain linguistic features in their production (see 6.2. Linguistic Features). If student production is assigned high grades, and their performance is grammatically correct and contains varied and appropriate vocabulary, along with idiomatic intonation and stress, this would indicate that the tests measure what they are supposed to measure and thus show reasonable empirical validity.

2.5. Acceptability and Intelligibility

In the assessment of student production of a foreign language, one of the central questions must be whether it is possible or not to describe acceptability and intelligibility in terms of quantity or quality. Intelligibility is closely related to acceptability: how many errors can appear in a student’s production before the interlocutor, or the assessor in an oral test, claims that
it is unintelligible? Discussions of acceptability and intelligibility concern matters investigated in studies of error evaluation and issues in research carried out on the effect of errors on the interlocutor, in this case the interviewer. This effect can be a matter of whether the addressee understands the meaning, his/her affective response to the errors, or, as in oral tests, the assessor’s evaluation and the grades which are assigned to the student production. Several studies of error evaluation were made in the 1970s and 1980s (see for instance, Quirk and Svartvik 1966; Greenbaum and Quirk 1970; Olsson 1973 and 1978), and some of the main tendencies in the results of these studies are worth mentioning in this context. The presentation that follows is based on Ellis (1994: 63–67).

Firstly, several studies have shown that native speakers tend to judge lexical errors as more serious than grammatical errors. Furthermore, errors that affect overall sentence organization are considered to affect comprehension more than errors that affect single elements in a sentence.

Secondly, non-native speakers, and thus teachers of English, tend to be more severe in their judgements, in particular regarding morphological errors. A Danish study of native speakers’ assessment of oral interviews came to the conclusion that the frequency of errors was decisive, irrespective of the types of errors made:

all errors are equally irritating ... irritation is directly predictable from the number of errors regardless of the error type or other linguistic aspects (Albrechtsen, Henriksson and Faerch 1980: 394).

Thirdly, native speakers and non-native speakers seem to have different criteria for evaluating students’ errors. Native speakers tend to be focused on the effect that an error has on their understanding, whereas non-native speakers tend to be influenced by their background knowledge of the syllabus, the textbook, the students’ use of grammatical structures studied and the students’ mother tongue. However, it is difficult to draw conclusions from studies of error analysis and error evaluation since, according to Ellis (1994: 67), it is not always clear what criteria assessors use when asked to assess the seriousness, intelligibility or acceptability of an error.

Svartvik (1985) provides teachers with guidelines for the assessment of errors and their acceptability in student production of English in upper secondary schools. When teachers correct and assess student production, they are advised to have a descriptive view of errors, as opposed to a prescriptive one in the teaching situation. The descriptive view implies that they are aware of the variation that may exist in English, for instance regarding the use of there’s and there are followed by a plural noun. Besides this descriptive view recommended in assessments, three criteria for the assessment of student production are set forth by Svartvik (1985). These criteria
are intelligibility (what the student says is understood), usage (what the student says could have been produced by a native speaker of English) and appropriateness (what the student says is appropriate for the situation and in the linguistic context).

2.6. Concluding Remarks

In this chapter, concepts and issues have been presented which are of importance in discussions of assessing oral proficiency. Proficiency is the broad concept which is in focus in the assessment of student performance. Although often used rather loosely, proficiency is a multi-faceted term that must be carefully defined if it is to be tested and graded. Reliability and validity are central concepts in testing and necessary terms in discussions of the interpretation of students' test results. Finally, the notions of acceptability and intelligibility need to be considered, especially when students' linguistic competence is studied and when, in that context, errors are analysed and discussed.
CHAPTER 3

Formats: Tasks and Organisation

In this chapter, the definition of the oral test will be scrutinized and three ways of organizing oral proficiency testing will be described. The formats of three oral tests in varying degree different from the one used in the Gothenburg project will be described and compared with each other and with the test used in the Gothenburg project. The three oral tests are the oral part of the CAE, i.e. the Cambridge Examination of Advanced English, an oral test from Holland and an oral test from Lithuania.

3.1. The Oral Test: Definition

Underhill (1987:7) defines oral tests in the following way:

An oral test is a repeatable procedure in which a learner speaks and is assessed on the basis of what he says.

The definition above distinguishes oral tests from everyday conversation by way of introducing the procedure of assessment. In everyday conversation the interlocutors may certainly assess one another on the basis of the language produced and thereby draw conclusions about the interlocutor’s background and social status, but in oral tests, one of the participants is instructed to assess the other participant’s production. It is this process of systematic assessment of speech that makes a conversation a test. The systematic assessment can be carried out by the interlocutor or by an assessor who does not participate in the conversation but only assesses the student’s performance.

The purpose of an ideal oral test is to elicit the richest possible speech sample from the student in the shortest period of time (Lowe 1976). Furthermore, the ideal oral test would sample a variety of speech styles, and within each style, sample the functions that are needed the most by the student in future authentic communicative situations.
3.2. Three International Oral Tests: Descriptions and Comparisons

In this section the oral tests and the assessment procedures of three exams are described. The three tests are the Cambridge Examination of Advanced English (CAE), a Dutch National Test of English and a Lithuanian National Test of English. The focus here is on the parts of the tests that are designed to test oral production. The description of these three tests is made with the aim of pointing out not only differences but also similarities in tests used in different situational and geographical contexts. The purpose is thus to provide a background to an evaluation of the test used in the Gothenburg project by comparing it to three other oral tests. The descriptions below are based on test samples, handbooks or brief guidelines of these three tests and focus on similarities and differences with regard to the

- number of parts in a test,
- types of prompts in the tasks,
- instructions to students,
- preparation time for students,
- assessment procedure, and
- number of examiners.

The Certificate of Advanced English (CAE). The CAE is one of the highest language-only qualifications in the examinations of the Local Examinations Syndicate of the University of Cambridge. Numerous handbooks and study guides are available to prepare students for the Cambridge exams. In these handbooks, detailed descriptions of the tests are provided. Students can prepare themselves and decide to take the exam whenever they find appropriate, for instance, when they estimate their proficiency to be at the level required. The test consists of a theme-directed conversation and other oral tasks based on a variety of stimuli. The test is organised in four parts: (1) general social interaction, (2) using transactional language in individual long turns, e.g. comparing and contrasting, (3) negotiating and collaborating, and (4) explaining, summarising and developing the discussion. In (1), students are asked to respond to questions about themselves; in (2), they are given visual prompts which they are to talk about, in (3), they are given visual or written prompts of two-way collaborative

1 These illustrative examples were selected for the following reasons. The CAE represents a widely used test of oral proficiency in English. The Dutch test was in use (1994) when the material of the Gothenburg project was collected. The Lithuanian test provides an illustrative example of a test of English as a Foreign Language from a different school context.

2 The examinations of the University of Cambridge Local Examinations Syndicate are of similar types and span five levels. The CAE is at Cambridge Level Four.

3 Examinations are generally organised on two occasions every year.
problem-solving tasks such as comparing or selecting; and, finally, in (4), the topic area from (3) is developed in a discussion of wider issues. The standard format of the oral tests is two candidates and two examiners and the test is expected to last for 15 minutes. The assessment is carried out with the help of the rating criteria within the overall context of the Cambridge Common Scale for Speaking where the CAE is at level 4 on a 5-level scale. The two examiners assess the candidates with marks on a scale of 0–5. The interviewer uses a Global scale, and the assessor applies detailed Analytical Scales of the following criteria: Grammar and Vocabulary (Accuracy and Appropriacy), Discourse Management, Pronunciation (Individual Sounds and Prosodic Features) and Interactive Communication (Turn-taking, Initiating and Responding).

The Dutch National Test of English. The Test of Oral Proficiency in English, level 4 (the version of 1994 of the 16+ Examination, D-programme), of the Dutch School system comprises six tasks/parts: (1) warm up, (2) pronunciation, (3) structure, (4) description of events, (5) giving an opinion, and (6) engaging in a conversation. Task 1 is simply intended to put the candidate at ease and is not evaluated. In Task 2, the student reads a short written text. Separate sounds and sound combinations are to be judged as correct or incorrect, and the focus seems to be on sounds that may cause difficulties for Dutch students. In Task 3, sentences are to be completed (the first few words are given in the instructions to the student) and these sentences are answers to questions posed by the interviewer. Pictures along with three to five of the first words in the sentences are provided as guidance for the students. In Task 4, the students are asked to retell three stories with the help of series of pictures. In Task 5, the students are to give an opinion or an explanation, and finally in Task 6, the students participate in a conversation in which they are to take turns by either asking for information or answering a question. Each task is assessed separately using separate rating criteria. In Task 2 the criteria are Correct or Incorrect in Task 3 the students’ answers are to be rated as “No answer or illogical”, “Meaningful, but disturbing errors” or “Meaningful and almost error-free”. In Tasks 4 to 6 there is a scale of four levels of different criteria, for instance, Intelligible, Correct, Complete and Fluent are used in Task 4. The assessor is to give a total number of points to the performance in each task. The responses in the individual test items are assessed and scores calculated. These points are summed up as the test score. Pronunciation is regarded as

---

4 The topics of the test materials introduced for the June 2001 session comprised for instance “Carrying things” (compare, contrast and hypothesize, Part 2) and “Competition” (discuss, evaluate and select, Part 3 and 4).

5 The test was produced by CITO, the National Institute for Educational Measurement, Arnhem, Holland.
a separate skill within oral proficiency, and two separate test scores are therefore reported: one for Pronunciation and one for general Oral Proficiency.

The Lithuanian National Test of English. The Speaking Part of the Lithuanian National Test in English of the students’ 12th year at school (the version of 2000, levels A and B) comprises three tasks: (1) discussion expressing an opinion on a given topic, (2) an information-gap activity and, finally, (3) a role-play with a role to act out in a given situation. The test is preferably carried out with two students, one interviewer and one assessor. The interviewer is to conduct the test, the assessor to evaluate the performance. The students are given time to prepare the tasks just before the test is carried out. Furthermore, detailed instructions are given to the students how they can prepare for the test in the best possible way. The students’ performance is rated with the help of a scale from 1 to 10 with rating criteria for Linguistic ability (Pronunciation, Fluency, Grammatical Accuracy, and Vocabulary Range or Relevance of Content). Additional rating criteria are provided for every task considering “transaction”, “interaction” or “discussion”.

3.3. Discussion

The survey above displayed similarities and differences in three oral tests of English. The Cambridge Test of Advanced English has a rich diversity of prompts and different themes in two parts of the conversation. Two examiners (interlocutor and assessor) assess the students’ performance but do so in different ways; global grades are given by one and analytical grades are given by the other. The Dutch oral test has more focus on the pronunciation of individual phonemes. The student is also given more detailed instructions before the tasks and the responses are intended to follow certain structures. Scores are calculated on the basis of the students’ results. A separate score is given for pronunciation, besides oral proficiency. The Lithuanian test uses an elaborated scale in the assessment, from 1 to 10, and the test format is close to the CAE; the participation of two students is recommended, an interviewer conducts the test and an assessor grades the students’ performance.

In sum, the three tests provide examples of oral proficiency tests as background to the test format and the assessment procedure used in the Gothenburg project (cf. Chapters 5 and 7). Differences between the three tests discussed worth noticing concern

– the number of examiners and their tasks in the tests (eliciting or/and assessing),
– the types of prompts used (for instance, whether tasks are highly structured, such as reading aloud or answering with the help of words of the expected answers, or organised around a common theme)
– the students’ preparations (for instance, whether advice how to prepare for the test is provided in instructions or in handbooks)
– the scales used in the assessment procedure (for instance, pronunciation may be regarded as a separate skill, as an integrated skill or as two types of abilities to be rated, namely individual sounds and prosody)
– the assessment procedure (the method used in the final scoring shows a variety of possibilities, from a detailed scoring of individual tasks to grades assigned to separate skills)

In the comparison between the three oral tests described above and the test used in the Gothenburg project, the following observations can be made. In the test of the Gothenburg project

– the procedure with one interviewer/assessor was used,
– the tasks showed some similarities with the tasks described above, for instance, the students were instructed to tell stories with the help of some pictures or express their own opinion,
– the students were given ten minutes to prepare two of the three tasks, no time for preparation was given before the last task,
– one global and three factorial grades were given,
– grades were given for the students’ whole performance, i.e. the production in the three tasks.

The presentations of the three illustrative examples of oral tests above show that there is no obvious consensus on what methods are best to use in oral proficiency testing. There is thus a need for empirical studies in order to find efficient and useful ways of measuring students’ ability to speak English. This provides a justification for the present investigation of the testing and grading of Swedish students’ oral production.
Part 2:
The Aims of the Study and Hypotheses to Be Tested
4.1. Aims

The present study has three main aims to be introduced in this chapter. As explained above, the study is based on the material of the Gothenburg project. The aim of that project was:

- to study to what extent different examiners tend to agree or disagree in grading advanced Swedish students’ oral proficiency in English.

In the collection of the data for the project, nine English teachers, representing three different categories of teachers, were engaged to interview ten students each (for the description of the interviewers, see 7.3.). Each of the 29 students was thus interviewed by one teacher from each category, i.e. three times. In their evaluation of the students’ performance, the interviewers were instructed to give each student four separate grades: one Global for the overall oral proficiency impression and one for each of three different Factors, namely Communication and Fluency, Grammar and Vocabulary, and Pronunciation (for a complete description of the material, see Chapter 7). The research design was geared to investigate methods of assessing oral proficiency in view of the compulsory oral tests that were to be introduced as part of the National English Tests in Sweden a year later.

By way of an elaboration of the above aim of the Gothenburg project, the first aim of the present study is two-fold:

- to study the relationships between the Global grades and the three Factorial grades given by each of the nine interviewers individually;
- to study the relationships between the Global grades and the three Factorial grades for each of the three interviewer categories.

Along the lines of the Gothenburg project, this two-fold aim thus focuses on the work done by the interviewers, not on the students’ proficiency as such.

\(^1\) 29 students turned up out of the 30 who had been selected for the project.
In connection with the follow-up work based on the tape- and video-recorded interviews, a second aim was formulated for the purposes of the present study. This further aim makes this study an inter-disciplinary product that brings together the frameworks of linguistic and pedagogical study. The second aim is then

– to study the students’ linguistic competence, and to some extent their strategic competence, in terms of their standard English or non-standard English use of certain “difficult features” in English, their vocabulary, their use of certain discourse markers, three aspects of their pronunciation, and their ability to use efficient strategies (for the linguistic features studied, see 6.1. and 6.2.).

This part of the investigation is a study of grammatical accuracy as well as of various discoursal and collocational features, and of the strategies that students use to carry on the conversation without resorting to Swedish, even when in want of an appropriate English word or expression.

Finally, as a combination of the above first and second aims, the third aim of the present study is

– to investigate to what extent the students’ linguistic and strategic competence is graded differently by the three categories of interviewers, and to study the role played by the linguistic features in the Global grades assigned to the students.

The results are discussed in six chapters focusing, on the one hand, on the grades assigned (Chapter 8), and, on the other, on the students’ use of the selected linguistic features in their production (Chapters 9 to 13). The results are then interpreted in Chapters 14 and 15.

4.2. Hypotheses

In pedagogical research it is customary to start with a ‘null hypothesis’, i.e. no differences between the groups that are compared are expected. This is the principle even if the researcher may have some ideas of what differences to expect. The null hypothesis is important, among other things, to ensure the use of proper statistical methods and interpretations.

The null hypothesis of the Gothenburg project was that no differences would be found between the grades given by the different individual interviewers to the same student, nor between the grades given by the different categories of interviewers. This implies that the project was planned and carried out in such a way that no other hypotheses would be favoured.

For the linguistic analyses involved in the second aim of the study, differ-
ent principles could be adopted. These analyses pertain to the grammatical correctness and communicative efficiency of the students’ performance in English and therefore the following hypothesis could be formulated:

– Students who are at the lower end of the grading scale, both as far as their previous English grades from school, and the grades given by the interviewers are concerned, tend to make more elementary grammatical mistakes, have larger lacunae in their vocabulary and a poorer strategic competence, which means that they may have a tendency to resort to Swedish words, sometimes pronounced in an ‘English-like’ way to make them sound more correct.

In addition to the hypothesis formulated above, a further hypothesis concerning the third aim could be formulated as follows:

– Differences might be found in the attention paid by the members of the three categories of interviewers to certain linguistic features (i.e. the students’ linguistic competence) when assigning the Global grades.²

These two main hypotheses and their extensions will be further dealt with in Chapter 6 and Chapter 8.

² Owing to space and time limitations, the discussion of differences in the Factorial grades assigned (for instance Factorial grades for Grammar and Vocabulary) in comparison with the occurrence of certain linguistic features will not be dealt with in the present study.
Part 3:
Grading Students’ Oral Performance
and Analysing Their Competence

In Part 3, some basic principles will be presented regarding both the assessment of the students’ performance and the analysis of their linguistic and strategic competence. These principles concern the methods employed, the criteria used in the classification of certain categories, such as test rating categories, sources of variation in the grades assigned, and sources of error in the linguistic analyses, which should be taken into account in the discussions of the results.
The following three issues are discussed in this chapter:

- principles in the assessment,
- global and factorial grades, and
- sources of variation.

5.1. Principles in the Assessment

Three important principles were applied in the assessment of the student production in the Gothenburg project.¹

(a) First of all, student production was graded by the interviewers during or immediately after the interview. The interviewers assigned a Global grade and three Factorial grades to each student (cf. the instructions for the Interviewers and the Oral Mark Sheet, Appendices 5 and 9). Furthermore, the interviewers were asked to make relevant comments regarding other matters, such as body language.

(b) Secondly, the grading was based on the interviewer’s interpretation of a set of rating criteria. The interviewers (with a few exceptions) did not know the students before they met for the interview and had no knowledge about the students’ grades or test results in English from school. The interviewers were given the criteria about a week before the day of the testing session, but they did not discuss the interpretation of the rating criteria with each other or with the research team of the project as a part of the organisation before the assessment was carried out (cf. the Rating Criteria in Appendix 8).

¹ The design of the procedures and principles in the assessment of the Gothenburg project was decided by Docent Lindblad and his colleagues at the Language Teaching and Testing Research Unit, Department of Education and Educational Research, Gothenburg University.
(c) Thirdly, the tasks were constructed in order to give all the students, independently of personal characteristics, interests and background, the possibility of showing as much as possible of their oral skill.

The grades assigned are thus based on the interviewers’ assessment of the students’ performance during the interviews and their interpretations of the rating criteria provided. The interviews were organised in order to encourage the students to speak as much as possible in spite of the somewhat unfamiliar and stressful surrounding, at least for some students.

In the discussion of the grades assigned to the student production, the following two conditions should be kept in mind. Firstly, the grades were assigned to the quality of the students’ proficiency such as it was on the day the oral tests were carried out. Secondly, the grades assigned were probably to some extent dependent on how well the interviewers succeeded in making the students speak, and whether the interviewers thus managed to make the students do themselves justice and show their real oral skill.

5.2. Global and Factorial Grades

As described in the Introduction, the student production was graded by assigning each student’s test one Global grade and three Factorial grades. The three Factorial grades were based on the interviewers’ assessment of the students’ Communication and Fluency (2a), Grammar and Vocabulary (2b), and Pronunciation (2c). The three Factorial grades were formulated in conformity with those that were used during the five-year experiment with oral tests as part of the Swedish National Test of English used in Swedish schools at the time of the Gothenburg project (i.e. 1993, cf. Introduction).

In the assessments, specified sets of rating criteria for the Global grade and for the three Factorial grades were used (for these rating criteria; see 7.5. Rating Criteria and Appendix 8). In the present study, the grades awarded will be studied and compared in the following two respects. Firstly, the grades (Global, Factorial 2a, 2b, or 2c) assigned by the three interviewers to the individual student’s three tests will be studied (cf. Tables 8.1. to 8.4. and Tables 8.8. and 8.9.). Secondly, the Global grade and the Factorial grades assigned to each student’s production by her/his three interviewers will be presented (cf. Tables 8.5. to 8.7.).

The grades used by the interviewers were A (the highest), B, C, D, and E (the lowest). What these grades stood for is found in the lists of rating criteria for the Global grade and the three Factorial grades (cf. Appendix 8). In the tables and the discussions in Chapter 8, these ‘letter grades’ are transformed into ‘figure grades’ (an A corresponds to 5, and then the fol-
lowing letters to 4, 3, 2, and 1). The original ‘letter grades’ represent a so-called *ordinal scale*, which means that an A is better than a B, and a B better than a C, etc. but these ‘steps’ cannot be guaranteed to be equal. When the letters are transformed into figures, the result is an *interval scale*, which means that the ‘steps’ of the scale are equally large. In accordance with what has become common practice,² the tables and the discussion thus contain means, standard deviations and correlations rather than medians and other non-parametric statistics.

5.3. Sources of Variation

The grades that were assigned in the Gothenburg project were based on the rating criteria provided and the students’ oral proficiency in English. Nevertheless, other factors could influence the grades assigned. These possible factors might have influenced the interviewers’ assessments, and they should therefore be mentioned. Four such potential sources of variation in the assessment of student production will be discussed. They are:

(a) the students’ and the assessors’ previous experiences and expectations,
(b) unknown background variables,
(c) the so-called halo-effect, and
(d) a tendency to go for the middle of the grading scale.

(a) The conversations in oral tests may seem to be similar to other conversational situations, but in particular, the process of systematic assessment makes them different (cf. 3.1. The Oral Test: Definition). In other respects, oral proficiency interviews and everyday conversation can appear to be similar. When people take part in conversations, they bring assumptions and expectations about what is to be said, how the conversation will develop and what kind of contributions they are expected to make (Richards and Schmidt 1983: 119–120). This is also the case for participants in oral proficiency interviews. These expectations, both the students’ and the interviewers’, may influence the students’ output and the grades assigned. For instance, the students could have ideas about what the interviewers expect from them. As regards the interviewers, their previous experience may influence their interpretations of the rating criteria and their views on the

² Similar principles are used, for instance, by the Swedish National Agency of Higher Education (*Högskoleverket*) in the calculation of students’ grades for entrance into higher education in Sweden. Individual students’ letter grades G, VG and MVG are transformed into the figures 10, 15 and 20 respectively (cf. www.vhs.se and the principles for admission to higher education by Verket för Högskoleservice).
students' performance, in particular when the interviewers knew the students from school.3

From the students’ perspective, previous experience of oral testing may play an important role. The students may have had previous experiences of oral proficiency interviews with their language teacher as the assessor. Such experiences of successful or unsuccessful oral tests may, at least partly, explain why certain students felt nervous during the oral proficiency interviews. In this context it is worth pointing out that encounters between teachers and students in teaching situations are often quite different from testing situations. Teaching is co-operative. Tests do not necessarily have to be, but can perhaps sometimes be regarded by some students as too challenging. There are possible outcomes which may be regarded as unfairly negative to the student (Cheepen and Monaghan 1990: 91). Such students’ performance may perhaps be influenced by the experiences they have of teachers’ assessment of their oral proficiency, and their performance might be better in other contexts. Certain students are probably nervous and anxious whereas other students may feel quite relaxed and calm. Whether the students have experiences of being recorded may influence their oral production. The presence of a microphone can contribute to a more hesitant delivery (Skolverket 1993: 121), and thus negatively influence the students’ performance.

The students’ general knowledge concerning controversial issues and whether they are used to debating in general are factors also worth mentioning in the discussion of the possible significance of advanced students’ background and previous experiences. Students who know about and who are used to discussions in different contexts may have many ideas of what to say and how to say it in an oral test of English where there is a discussion, as for instance in Task 3 of the Gothenburg project (cf. 7.6. Tasks). It is thus evident that, in one way or another, the students’ general knowledge contributes to their ability to communicate. The students’ general competences, such as knowledge of the world and sociocultural knowledge, may be regarded as aspects of their communicative competence (Common European Framework of Reference for Languages 2001: 101). The students’ personal characteristics may also be of influence. In a study of Finnish students in oral proficiency tests of Swedish, the results showed the tendency that extrovert and lively students scored higher than shy and silent, and in the same way, initiating speakers scored higher than passive speakers (Hildén 2000: abstract). Finally, the possibility that some students know the interviewers whereas for others it may be the first time they meet them is also a factor to take into account. The Gothenburg project collected infor-

---

3 School Teacher 2 knew two of the nine students he interviewed, and School Teacher 3 knew seven of the ten students she interviewed.
mation regarding several aspects concerning the students’ background (cf. 7.2. Students). Nevertheless, there may be background variables which are not known and which could account for some students’ spoken performance and the variation in the grades assigned.

The background and experiences of the interviewers are of great interest to the present investigation (cf. 4.1. The Aims of the Study, and 7.3. Interviewers). The different background and experiences of the interviewers were the factors behind the three different interviewer categories of the Gothenburg project. Whether, for instance, school teachers and native speakers make different interpretations of the rating criteria in the assessment of student production, and whether they put emphasis on different aspects in the students’ performance, for instance, on accuracy or on pronunciation, are among the issues investigated in the present study.

The so-called halo-effect is the term which may be employed to describe the possibility that assessors are positively influenced by, for instance, his/her previous knowledge of a student, and thereby this knowledge may influence the grades which are given to the student’s production. Furthermore, when assessors assign several Factorial grades, they may also be influenced by the grades assigned to the other factors. For instance, having just given a high grade for Vocabulary and Grammar, the assessor may be influenced by this high grade when he/she assigns the grade for Pronunciation. This phenomenon could be explained by the demands put on the assessor to assess several aspects of student performance simultaneously. It may be hard for the assessors to distinguish components in the course of their listening task (Yorozuya and Oller 1980; Oller and Hinofotis 1980). It may also be claimed that judging several factors at the same time, such as vocabulary, grammar, and pronunciation, puts unreasonable demands on the assessors (Walker 1984: 43).

A factor that may influence the grades assigned is a tendency for the assessors to go for the middle of the grading scale. Marks may, for instance, cluster round the median 3 on a 5-point scale (Heaton 1975: 99). This tendency may thus particularly but not only occur in assessments when scales with, for instance, three or five grades are used. In these assessments, the assessors tend to use the grades which are in the centre of the scale, such as the grade 3 on a scale of five grades. An even-numbered rating scale may be preferred to avoid this phenomenon (ibid. 1975: 99)

The discussion above provides illustrative examples of factors that could be of significance for the grades assigned to the students’ performance in the tests of the Gothenburg project.
CHAPTER 6
Analysing the Students’ Linguistic and Strategic Competence

In this chapter, the principles and method used in the linguistic analyses will be described. The discussion will include

– the criteria and principles for the selection of the linguistic features,
– the ten linguistic features selected,
– the classification of the student production into test rating categories,
– the influence of speech rate and the length of the oral tests on the results,
– the classification of standard English and non-standard English instances, and
– possible sources of error in the linguistic analyses.

6.1. Criteria and Principles

Learner language, i.e. English produced by non-native speakers, may be regarded as a variety of the English language. This variety has characteristics in common even though there may be great differences of quality when the production by two individual learners is compared. The language of non-native speakers is described as a “simplified language” by Corder (1981: 79–81) who claims that “simplified languages” share structural similarities. Corder exemplifies this term with reference, for instance, to the interlanguage of non-native speakers, pidgins and baby-talk and gives several examples of structural similarities which are considered to be the most salient:

– a simple or virtually non-existent morphological system,
– a more or less fixed word order,
– a simple personal pronoun system,
– a small number of grammatical function words and grammatical categories,
– little or no use of the copula,
– the absence of an article system, and
– tense and aspect typically performed by lexical means, for instance by using adverbs.

Even though the students of the Gothenburg project were at an advanced level of proficiency in English, Corder’s discussion of structural features of “simplified languages” was one of the influential works which were consulted at the early stages of the present study. His ideas were of significance in the first investigations of the language produced by the students and in the selection of the ten linguistic features included in the present study, for instance the third person singular marker.

The ten linguistic features selected comprise four areas: syntax and morphology (the verbal group), vocabulary and compensatory strategies, discourse markers and aspects of pronunciation. For each feature, the selection is based on at least one of the following three criteria:

1. The features are referred to in school grammars of English for Swedish students.
2. They represent different aspects of student production (morphology, syntax, vocabulary, strategies, discourse phenomena and pronunciation).
3. The features could be hypothesised as revealing differences in relation to the Global grades assigned by the members of the interviewer categories.

The interviewers’ comments in the questionnaires submitted directly after the completion of the testing sessions were also valuable for the choice of features as well as two additional interviews carried out after the testing session (cf. 7.1. The Material of the Gothenburg Project). Some interviewers’ comments on students’ production on the Oral Mark Sheet also contributed to the choice. Finally, my personal experiences from teaching English in Swedish schools were of help in the selection of the features.1

The occurrence of, and the students’ mastery of, the selected linguistic features are compared with the Global grades assigned to the students’ production. In the preliminary stages of the present investigation, both the Global grades and the three Factorial grades were taken into account. However, since this approach with four grades assigned to every student’s production tended to complicate matters in an unnecessary way, the outline was

---

1 At an early stage of the present investigation more linguistic features were included. However, it was not within the scope of this study to describe the occurrence of all these features in student production. The linguistic features analysed at an early stage were the following: auxiliary do; relative clauses; the passive voice; adjectives, irregular verbs, adverbs with -ly; certain hedges: sort of, kind of, type of; the phrases I think so, I don’t really know; the discourse markers I mean, you see, oh, and, but, or, or-tags; topic-comment structure and meta-talk.
changed. Only the Global grades assigned are now used when the grades are compared with the occurrence and the students’ mastery of the linguistic features.

Obviously, the Global grades assigned could be expected, at least partly, to be based on the adequate production or the occurrence of the linguistic features investigated. It is evident that grammatical accuracy in a student’s production can influence not only the Factorial grades of Grammar and Vocabulary but probably also the Global grades: we may notice the obvious interdependence between student production assigned a high grade and the occurrence of certain of the linguistic features, for instance, the standard English use of the third person singular marker (cf. 6.5. Defining Standard English and Non-standard Instances). It may be argued that there is a predictable relationship between these two variables. One of the interesting issues in the linguistic analysis is to clarify the possible relationship between the Global grades assigned and grammatical accuracy, i.e. the occurrence of standard English instances of certain of the linguistic features in the students’ production. In the same way, the investigations of the students’ use of vocabulary, their use of two discourse markers and their idiomatic, or unidiomatic intonation and stress may yield useful information regarding the Global grades assigned and aspects of lexis, discourse and pronunciation. The linguistic investigation thus provides useful information regarding the assessment of 29 students’ performance, and thereby the relationship between the Global grades assigned and the students’ linguistic and, to a certain extent, strategic competence in English.

Because of the interdependence of the variables described above and the expected linguistic results, the interest is focused on linguistic features that show great distributional differences in student production assigned different grades by the members of the three interviewer categories. Since few oral tests were assigned Low grades, the potential of the survey of the students’ use of the ten linguistic features lies in identifying features that distinguish student production assigned a High grade from student production assigned an Average grade (cf. the description of the three test rating categories High, Average and Low in 6.3. Classifying Student Performance into Test Rating Categories).

Finally, two caveats should be mentioned in the context of the analysis of the students’ language. Firstly, the present study describes the language produced in a specific setting and with a particular purpose. Secondly, understandably enough, it is not possible to give a full description of the linguistic production of each student’s three interviews; only a survey of their use of certain linguistic features is provided.
6.2. Linguistic Features

In the surveys of the ten linguistic features under scrutiny, raw, average and normalised figures are used to describe the occurrence and the distribution across the test rating categories. Furthermore, correlation coefficients, mean values and standard deviations are presented in order to show not only the relationship between the Global grades assigned and the occurrence of the linguistic features, but also the distribution of the instances across and within the three test rating categories High, Average and Low. To give a full account of the distribution within a test rating category, the figures for the occurrence rates in all student production rated High, Average or Low are given in the presentations. The limited amount and the character of the material necessitate an approach with a discussion of the performance of groups of students and the Global grades assigned to their performance.

As stated above, the ten features included in the analysis cover four aspects in student production. The ten features in question are:

Morphology and syntax: The third person singular marker
The progressive

Vocabulary: Abstract nouns
Modifiers of adjectives
Lexical errors and compensatory strategies

Discourse phenomena: Well ...
You know ...

Aspects of pronunciation: Intonation
Stress
Speech rate

There were various reasons for the selection of these linguistic features. In what follows, some of the most evident reasons are presented. (The reasons for the selection of each linguistic feature are further discussed in Chapter 9.)

The two features included in the category of Morphology and syntax (the third person singular marker and the progressive) were of special interest since they can both be regarded as accuracy markers and their uses are described in the normative grammars of English used in Swedish schools. Difficulties with the third person singular marker are mentioned as typical of learner language by Corder (1981: 79–81) and the use of the progressive is described as a contrastive difficulty in English school grammars for

\[2\] Standard deviations are generally calculated when the number of observations is large. In the present study, the numbers are small, in particular regarding certain linguistic features and student production assigned a Low grade. Standard deviations are nevertheless valuable in order to show differences in distribution and between assessors.
Swedish students (see, for instance Hargevik and Hargevik 1993: 180). Abstract nouns were included because their presence may be correlated with a general range of vocabulary or they might indicate the degree of abstract thinking shown in students’ speech. Modifiers of adjectives were also worth paying attention to, since they seemed to be used with highly varying frequencies by the students and differences in tokens and types of modifiers used could provide evidence of the significance of certain aspects of vocabulary for the assessment. Lexical errors were important to investigate since it seemed obvious that they could disturb communication. The occurrence of lexical errors and the students’ compensatory strategies, i.e. the strategies the students employed when their knowledge of vocabulary was not sufficient, were mentioned by the interviewers as being of significance for the assessment in the answers to the questions in the questionnaire (cf. 9.2.3. Students’ Lexical Errors and Compensatory Strategies). The students’ use of the discourse markers well and you know was relevant to analyse since these items are frequent in spoken English and because they were used with remarkable frequency by some students. Furthermore, it could be assumed that the students had been exposed to these discourse markers and that they used them for different purposes, for instance, well as a sign of thinking and you know as an indication of involving the interlocutor in the communicative process. The differences in frequency and use may be expected to influence the assessments. It was natural to include the three aspects of pronunciation (intonation, stress and speech rate) in an investigation of students’ spoken performance. It was evident that the students’ pronunciation was of significance, not only for the Factorial grades for Pronunciation, but also for the overall impression of the production and the Global grades assigned.

The investigation of the ten linguistic features of the students’ linguistic competence may be summarised with the help of ten sub-hypotheses, each one of them with reference to a linguistic feature. The ten sub-hypotheses are presented in Table 6.1.

A number of comments are in order concerning the sub-hypotheses in Table 6.1. As for the two features of the Verbal Group, the question largely boils down to the proportions of standard English and non-standard English instances in the student production. In the Vocabulary Group, the sub-hypotheses pertain not only to mere frequencies but also to the types of occurrences, what kinds of words are used, errors in words selected, and the solutions found to lexical lacunae. The sub-hypotheses concerning the two Discourse Markers investigated deal with frequencies and discoursal functions of well and you know. Finally, the sub-hypotheses concerning the three features in the ‘Aspects of Pronunciation Group’ all pertain to the following expected results:
Table 6.1. *Sub-hypotheses of the ten linguistic features*

<table>
<thead>
<tr>
<th>Linguistic feature</th>
<th>Sub-hypothesis</th>
<th>Sections of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Group: The third person singular marker</td>
<td>There is a higher proportion of standard English instances in student production assigned a High grade than in other student production.</td>
<td>9.1.1., 10.1. and 15.1.</td>
</tr>
<tr>
<td>Verbal Group: The progressive</td>
<td>There are fewer non-standard English instances in student production assigned a High grade compared with student production assigned an Average or Low grade.</td>
<td>9.1.2., 10.2. and 15.1.</td>
</tr>
<tr>
<td>Vocabulary: Abstract nouns</td>
<td>Abstract nouns characterise student production assigned a High grade. Frequent instances and types of abstract nouns, in particular nouns with no similar Swedish equivalent, are evidence of a more developed and advanced argumentation.</td>
<td>9.2.1., 11.1. and 15.2.</td>
</tr>
<tr>
<td>Vocabulary: Modifiers of adjectives</td>
<td>More instances of modifiers emerge in student production assigned a High grade than in other student production.</td>
<td>9.2.2., 11.2. and 15.2.</td>
</tr>
<tr>
<td>Vocabulary: Students’ lexical errors and compensatory strategies</td>
<td>Lexical errors are more frequent in student production assigned an Average grade, and, in particular, a Low grade. Compensatory strategies are more frequent in student production rated Low. Differences in occurrences of types of strategy can be observed across the test rating categories.</td>
<td>9.2.3., 11.3. and 15.2.</td>
</tr>
<tr>
<td>Discourse Marker: Well…</td>
<td>Well is more frequent and is used more idiomatically in student production assigned a High grade than in other student production.</td>
<td>9.3.1., 12.1. and 15.3.</td>
</tr>
<tr>
<td>Discourse Marker: You know</td>
<td>You know is more frequent in student production assigned a High grade than in other student production. You know is used as a pause-filling device and as one of the means to establish interaction between the interviewers and the students.</td>
<td>9.3.2., 12.2. and 15.3.</td>
</tr>
<tr>
<td>Aspects of Pronunciation: Intonation</td>
<td>“Unidiomatic use” of intonation in declarative sentences is more frequent in student production assigned an Average or Low grade than in students production assigned a High grade.</td>
<td>9.4.1., 13.1. and 15.4.</td>
</tr>
<tr>
<td>Aspects of Pronunciation: Stress</td>
<td>Stress, both on functional as opposed to lexical words and of the phrase and so on, is more often found to be used according to English usage in student production assigned a High grade than in other student production.</td>
<td>9.4.2., 13.2. and 15.4.</td>
</tr>
<tr>
<td>Aspects of Pronunciation: Speech rate</td>
<td>Student production assigned a High grade contain samples with higher speech rate than the samples of other student production.</td>
<td>9.4.3., 13.3. and 15.4.</td>
</tr>
</tbody>
</table>
more instances of idiomatic pronunciation emerge in speech by the students assigned a High grade than by other students in the material.

A majority of the sub-hypotheses in Table 6.1. follow what can be expected on the basis of common sense thinking: the students who were assigned a top grade produced a higher proportion of standard English instances, used more varied vocabulary, made fewer mistakes and had more idiomatic pronunciation. Other sub-hypotheses, for instance those stating that high-achievers speak faster than others and that their production contains high frequencies of discourse markers, can be regarded to be more controversial.

6.3. Classifying Student Performance into Test Rating Categories

The 82 oral tests were categorised into three groups: those that received a High Global grade, an Average Global grade and a Low Global grade. These three categories were based on the Global grades assigned to the 82 oral tests which were successfully recorded (cf. 7.1. The Material of the Gothenburg Project). The Global grades were A, B, C, D and E, A being the top grade (cf. the Rating Criteria in Appendix 8 for the description of the Global grades A, B, C, D and E). Student production graded A or B was classified as “High”, student production graded C was classified as “Average” and student production graded D or E as “Low”). Table 6.2. shows the distribution of the 82 oral tests across the three test rating categories.

As Table 6.2. shows, the material yielded only 13 oral tests assigned a Low Global grade. The small proportion of student production in the Low category and the limited number of oral tests labelled Low is taken into

<table>
<thead>
<tr>
<th>Test rating category</th>
<th>Number of oral tests</th>
<th>Percentage of oral tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>32 oral tests</td>
<td>39%</td>
</tr>
<tr>
<td>Average</td>
<td>37 oral tests</td>
<td>45%</td>
</tr>
<tr>
<td>Low</td>
<td>13 oral tests</td>
<td>16%</td>
</tr>
<tr>
<td>Total</td>
<td>82 oral tests</td>
<td>100%</td>
</tr>
</tbody>
</table>

3 In the discussions in the present study, the terms “High”, “Average” and “Low” refer to the three test rating categories.
account in the discussion, for instance, when comparing normalised figures in the discussion of frequencies (cf. 6.4. Issues to Take into Account: The Students’ Speech Rate and the Length of the Oral Tests).

6.4. Issues to Take into Account:
The Amount of Student Speech, the Length of the Oral Tests and the Correlation Coefficients

The precise length of the oral tests could be a factor influencing the occurrence of the linguistic features and the grades awarded and should therefore be taken into account. This applies both to time (total number of minutes, thus including the time used by the interviewer) and the amount of spoken student production (number of student words). For this reason, the length of the oral tests was measured in minutes, starting from the first greeting until the moment when the students left the interviewers. The difference in length between the shortest and the longest oral test was approximately 12 minutes; the shortest oral test comprised 8.20 minutes, and the longest 20.25 minutes (cf. Appendix 10. The Length of the Tests).

To take into account the amount of student speech, the words produced by the students were also counted (cf. Appendix 11. The Number of Students’ Words in the Tests). The material comprises 76,760 student words (tests assigned a High grade: 32,040 words, tests assigned an Average grade: 35,920 words, and tests assigned a Low grade: 8,800 words).

The frequencies of certain linguistic features (the progressive, abstract nouns, modifiers of adjectives, lexical errors, compensatory strategies, well, you know and and so on) in the student production were normalised per 1,000 words (see Biber 1988: 14). Frequencies were divided by the number of words of the respective test rating category, and this figure was multiplied by 1,000. The following example illustrates the procedure: 30 abstract nouns appeared in tests assigned a Low grade comprising 8,800 words. 30 / 8,800 = 0.003409, and then, 0.003409 x 1,000 = 3.409. The normalised figure is thus 3.4 per 1,000 words (cf. Table 11.1).

Finally, the correlation (r) given between the occurrence of certain linguistic features and the grades assigned should be commented on (cf. for instance Table 10.9). A correlation coefficient is a figure that indicates to what extent two sets of measures are related and to what extent variations in one of them tally with variations in the other (in the present study, for instance, for the occurrence of the progressive and the Global grades assigned, see Table 10.9, and, for the Global grades and the Factorial grades

4 Seconds were rounded upwards to the nearest interval of 5.
assigned, see Table 8.5). Correlation can be tested for significance and results may be expressed in terms of a 5% or a 1% level of significance. In the present study, significance at these two levels will be indicated in the tables with one asterisk (significance at the 5% level) or two asterisks (significance at the 1% level) to indicate whether the results are significant or not and if so, at which significance level (the significance levels are from Woods, Fletcher and Hughes 1986: 302).

6.5. Defining Standard English and Non-standard Instances

Since accuracy is one of the factors assumed to influence the assessment in the present study, linguistic features in the student production which do not follow standard usage as defined in normative school grammars are examined. In a school context, these instances would be judged as “incorrect”. To avoid too prescriptive an approach in the discussion, the terms “standard English” and “non-standard English” instances were used to distinguish the two types of uses. The distinction is applied in the analysis of the students' use of the third person singular marker, the progressive and uncountable abstract nouns.

The category of “standard English” instances includes structures which are described in at least one of three normative school grammars used in Swedish schools, namely, Svartvik and Sager (1971), Hargevik and Hargevik (1993) and Ljung and Ohlander (1992). Instances which were classified as “non-standard English” are not referred to in these school grammars. Certain “non-standard English” instances may of course occur in varieties of the English language (cf. for instance 1.3. and the discussion of the third person singular marker), but in this specific testing situation it is of interest to focus on structures not considered to be appropriate in a school context and to see to what extent these features occurred in the student performance across the three test rating categories.5

6.6. Sources of Error in the Linguistic Analysis

Researchers who analyse spoken language are fully aware of the difficulties and laborious work which may be involved in the collection of material. For

5 The notions of error and mistake are discussed in an investigation of grammatical errors in Swedish 16-year-old learners' written production in English. In that study, the notion of error stands for a form that is regarded as incorrect in relation to standard English grammar (Köhlmyr 2003: 17).
a researcher to analyse spoken production the material must in many cases be transcribed. The quality of the transcription of the oral script is extremely important for the subsequent linguistic analysis. In the present study, a great majority of the oral tests were transcribed by a secretary at the Department of Education and Educational Research, Gothenburg University. The written material was then compared with the tape recordings and corrected by the researcher himself. In a few oral tests, the language had to be transcribed with the help of the video-recorded material. Evidently, there may sometimes be a discrepancy between what the transcriber wrote and what the students said. In some instances, due to bad sound quality or misunderstandings, there may be divergences. These sequences of bad sound quality should not be neglected but they are unlikely to have influenced the results of the linguistic investigations to a great extent.

In the linguistic analyses of the linguistic features, various factors are taken into account in order to describe the students’ performance. They are, for instance, discoursal factors, collocational factors, types of standard English and non-standard English instances and the position of certain linguistic items in the utterance. In the presentations of the students’ uses of the linguistic features, more factors could be included. They are factors such as the way discourse markers or certain nouns attested were pronounced with regard to stress and rhythm, and the absence of certain verb forms where they would be expected to appear, for instance the progressive. It is not probable that these factors would have influenced the results. Nevertheless, they may be regarded as sources of error for the interpretations of the linguistic analyses.
Part 4:
The Practical Organisation of the Testing Session

The present study describes and interprets the results of the Gothenburg project, i.e. the grades assigned by the interviewers and the language produced by the students (cf. the Introduction and 4.1. Aims of the Study). Docent Lindblad, head of the Foreign Language Teaching and Testing Research Unit of the Department of Education and Educational Research, Gothenburg University, had the full responsibility of organising the testing and collecting the material. The principles of the organisation and the procedures in the collection of the material will be described in this part of the study.¹

¹ The Gothenburg project also contains assessments of some of the students’ performance by the interviewers in a second round, and by a group of international assessors. These assessments are not included in this study.
CHAPTER 7
Selecting, Testing and Assessing: Organising the Testing and Collecting the Material

The present chapter provides a survey of the organisation of the testing and the collection of the material of the Gothenburg project. This survey is structured in the following way:

- the material of the Gothenburg project,
- the 29 students and their background,
- the nine interviewers and their background,
- the three tasks and the topics, and
- the rating criteria.

7.1. The Material of the Gothenburg Project

The material of the Gothenburg project, briefly described in the Introduction, was collected on May 8, 1993, with the participation of nine interviewers and 29 students from the Swedish gymnasieskola, i.e. upper secondary school.

The testing day yielded three kinds of material: the recordings of the tests, the grades assigned (four grades for each performance, cf. 5.2. Global and Factorial Grades) and the answers to the questionnaires given by the students and the interviewers (cf. Appendices 6 and 7). The 82 successfully recorded oral tests are available on audio-tapes and video-tapes, distributed as follows (the total number of graded tests is given in brackets):²

- Oral tests assessed by School Teachers: (29) 29
- Oral tests assessed by University Teachers: (29) 24
- Oral tests assessed by Native Speakers: (29) 29

Total of oral tests: (87) 82

² Five oral tests were not recorded for technical reasons: one interviewer had trouble with the tape-recorder and the video-camera.
Once the oral tests had been completed, an assistant at Gothenburg University transcribed orthographically all that was said in the course of the tests. (Normal orthography was used.)

Below follows a short extract of the beginning of one student’s production given by way of illustration (cf. the pictures of Task 1 in Appendix 4C):

Interviewer: So right now my name is X okay and you are going to talk about some pictures.
Student: Yeah.
Interviewer: Right have a good look so then right off we go.
Student: Okay there are two boys eh who are going to the beach and they are lying there on the beach and eating and so then they see a airplane a small airplane a model in the mountains and they climb up to catch it.
Interviewer: Did you say the mountains? Cliff, I'll give you the word.
Student: Aha, cliff okay and they climbs up to catch it and eh then they can't climb down and eh and an old pair sees them and calls the police and then a police helicopter comes and catch them and they eh thank the police and the old lady.
Interviewer: Right good (laugh) you have you ever been in a situation like that or so?
Student: No.
Interviewer: Never had to be rescued by the police?
Student: No (laugh).
Interviewer: (laugh) No, boys will be boys. (Task 1, Test S19)

As illustrated in the extract above, non-verbal utterances, for instance, the students’ laughs or sighs were included in the transcription. All counting of students’ words needed for the linguistic analysis was performed manually by me from the written transcripts.

The grades assigned comprise the assessments of all the 87 oral tests (82 of which are available on tapes, cf. above). As described above (cf. 5.2. Global and Factorial Grades), the tests were assigned the following four types of grades: a Global grade and three Factorial grades for Communication and Fluency, Grammar and Vocabulary, and Pronunciation respectively. The grades were given with the help of a set of rating criteria (cf. 7.5. Rating Criteria).

Two questionnaires, one for the students and one for the interviewers, were distributed immediately after the oral tests had been carried out. The questions posed to the students dealt with, for instance, their achievement profile at school, their spare time occupations and their attitudes to the oral tests they had just taken part in. The questions in the interviewers’ questionnaire concerned issues such as the method used in the oral proficiency interviews and their attitudes to the student language they had just heard (see Appendices 6 and 7 for the questionnaires used).  

3 I transcribed a few oral tests myself.  

100
7.2. Students

The group of students taking the tests consisted of 29 students from one and the same school in Gothenburg who were about to finish school a few weeks later. The students were asked by their teachers if they would be willing to participate in oral tests at the university (no random sampling took place). They were invited by letter to come to Gothenburg University on a Saturday and were remunerated with 200 Swedish crowns for their participation in three oral tests. They were informed about the research project and told that the grades assigned would not affect their school grades and that their production would be evaluated purely for research purposes. The background variables collected are given in Table 7.1.

The group of students consisted of 17 girls and 12 boys. The great majority, 24 of them, were 19 years old, four were 20 years old and one student was 21 years old.

4 Two additional tape-recorded meetings were arranged by myself, one with a native speaker interviewer participating in the Gothenburg project and one with a test constructor in Holland. In these two interviews, certain students’ production was listened to and discussed. These discussions provided useful information on what linguistic features could be of significance for a top grade. They were also valuable in connection with the selection of the linguistic features (cf. 6.1. Criteria and Principles).

5 No students were of immigrant background. The small-scale experiment of the Gothenburg project did not give priority to collecting material produced by students of different backgrounds since the focus of interest was on the interviewers’ grading.

6 These background variables were collected in due course. In the collection of the material for the Gothenburg project, the focus of interest was on the grades awarded, not on the students’ performance as such.
At the time of the testing, the Swedish *gymnasieskola* was structured in “study programmes” with different profiles. The students all followed so-called theoretical study programmes. The 29 students were not selected to be representative of Swedish school-leaving students in general or a group of average students of their age; instead, they belonged to the group of students in the population who had chosen theoretical studies at upper secondary school level, and with English as one of their subjects of interest. They could therefore be expected to be more interested in language studies than many other 19-year-old Swedish people. They represented three course programmes as follows:

- **Economics (*Ekonomisk linje*)**: 13 students,
- **Social Science (*Samhällsvetenskaplig linje*)**: 12 students, and
- **Natural Sciences (*Naturvetenskaplig linje*)**: 4 students.

The students had probably all studied English for nine years in the Swedish school system.

At the time of the testing, the norm-referenced grading system with grades ranging from 1 to 5 was still in use. A 5 was the highest grade, and a 3 was laid down as the average grade.\(^7\) The grades of the 29 students taking the tests are given in Table 7.2.

On the basis of the school grades above, the students represented both students who could be considered to be successful in their English studies at school, and some students whose proficiency was lower than average in the group of students in the 3- and 4-year programmes preparing for higher academic education. More students belonged to the former than to the latter category.

One possible reason for success in language studies could be the time spent abroad in English-speaking countries. According to the information given by the students in the questionnaires, seven students had never visited

\(^7\) For a discussion of norm-referenced assessment, see 1.1. Measuring Student Performance. At a national level the five grades of the norm-referenced assessment were to follow a normal distribution of 7%, 24%, 38%, 24% and 7% (cf. the Introduction).

---

Table 7.2. *Students’ school grades in English*

<table>
<thead>
<tr>
<th>School grade</th>
<th>Number of students</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

---

102
countries where English is spoken, eighteen students had spent between one and five weeks in such countries and four students had spent more than five weeks (and up to 16 weeks) in countries where English is spoken. Most of the students who participated in the collection of the material had thus spent only short periods, probably holidays, in English-speaking countries. Eight of the students had participated in a summer language course of English arranged by a Swedish organisation.

Since English is the language of communication in many international contexts, the students were asked whether they had spent periods abroad and then had been exposed to English in other countries than those where English is the primary language. The students’ answers showed that all of them had spent at least three weeks abroad. Nine students had spent between three weeks and three months abroad. Eleven of them had been abroad between three and six months, and seven had spent more than six months abroad. (One student had been abroad for 72 weeks.) Generally, the students thus had experiences from time spent abroad and, it may be assumed, also from contexts where they had used the English language as a means of communication.

The students were also asked about their reading habits and whether they had access to international TV channels. Of the 29 students, 19 answered that they often read books in English in their spare time, five students that they had done so only once, and five students stated that they had never read English books outside school. Fifteen students had access to English or American TV-channels such as BBC, CNN or MTV at home.

To sum up, the 29 Swedish students had slightly better grades in English than the average student of English. They were enrolled in programmes with a theoretical profile in the gymnasieskola. Furthermore, the majority of the students had experiences of stays abroad, many of them from countries where English is the main native language. A majority of them often read books in English in their spare time and had access to international TV-channels at home.

7.3. Interviewers

The interviewers were six men and three women, representing three interviewer categories: Teachers of English at the Swedish gymnasieskola, University Teachers of English and Native Speakers of English (two of them specially trained examiners, see below). In every category both male and female interviewers were represented. There were reasons for the choice of each one of the three categories of interviewers, and these reasons are described in what follows.
As for the first category, it was natural to include school teachers since oral tests rated by school teachers form the most common setting for oral proficiency testing of English in Sweden. In Swedish schools, the student’s own teacher of English is usually the interviewer and the assessor in the oral test even though teachers, in the last few years, have been encouraged to cooperate in different ways. Furthermore, school teachers are familiar with the students’ expected level of proficiency in English. It was thus natural that this category of interviewers should be included in the Gothenburg project. The second category, the university teachers of English, can be expected to have a thorough theoretical knowledge and to be used to testing university students of English. With this background, their assessment of the students’ performance is of interest. The third category, the native speakers, may pay attention to different features, both positive and negative, than do non-native speakers. Overall, the interest in the Gothenburg project was not so much in the individual student’s grades as in the grades assigned by the interviewers and particularly in the differences between the members of the three categories of interviewers.8

All the nine interviewers had many years’ experience of Swedish students of English and the teaching of English in Sweden as well as of oral proficiency testing with Swedish students. The three School Teachers knew some of the students from school, but only in a few cases were they their teachers of English. Two of them had worked for more than twenty years on language testing and the construction of the Swedish National Test in English. As for the University Teachers, one of them was a native speaker but had studied English at universities in Sweden and specialised in analysing young Swedish students’ written production. The fact that she was a native speaker of English was not regarded to be an influential factor because of her academic studies in Sweden. Two of the Native Speakers had been examiners for the Cambridge exams in Sweden for more than twenty years. One of them also worked in the group that was responsible for the construction of the National Test of English in Sweden at the time. The third Native Speaker had worked with the English language at university level both in Sweden and in Great Britain. The three Native Speakers were thus well acquainted with Sweden and could be expected to be familiar with the performance of Swedish students.

The nine interviewers in the three categories were all about fifty years old or older. This fact might, to a certain extent, have contributed to the interviewers having similar ideas regarding the assessment of foreign students’ proficiency in English.

The fact that the interviewers represented three different categories of

8 In the present study, the interviewer categories are discussed in terms of School Teachers (ST), University Teachers (UT) and Native Speakers (NS).
“English teachers” was part of the project plan and will be followed in at least three ways. In the first place, the investigation of three interviewers’ grades of the same student’s performance in three interviews provides interesting information as regards assessment of oral proficiency in relation to the background and experiences of the interviewers. Secondly, it is of interest to investigate the grades assigned by the members of a category of interviewers in relation to the occurrence of the linguistic features investigated. For that reason, the interviewer categories are taken into account in the discussions of the results of the linguistic analyses (Chapters 10 to 13). Thirdly, the interviewers’ ways of carrying out the tests could be investigated, for instance, regarding the number and types of questions posed and the interviewers’ use of feed-back items. The interviewers’ methods are, at least partly, considered in the discussion of the results (Chapter 14. Grading Swedish Students’ Oral Proficiency in English).9

7.4. Tasks

The tasks were constructed to elicit language as close as possible to authentic and as much as possible from all kinds of student face-to-face conversation but with a certain control of content. This control was achieved with the help of a series of pictures in Task 1 and with a limited topic choice in Tasks 2 and 3. The tasks could be expected to make the students primarily concern themselves with the content of the message, the conveying of information, and the interaction and therefore not overtly concentrate on the way they formulated the message.

The idea of having a test consisting of three tasks was in line with the oral part of a Swedish National Test of English of the Swedish gymnasieskola at the time of the collection of the material. The topics were also, to a certain extent, the same as in that test. The three categories of interviewers had similar kinds of task but with different topics. Task 1 was a narrative task and the students were given different stories to tell the three interviewers they met. Task 2 contained three topics to talk about and the students chose one of the three topics. Each new interviewer that the student met had a new set of topics for the student to choose among, but the character of the topics provided by the interviewers was similar. Task 3 provided three alternative topics for discussion: the students chose whether to argue for or against a position on a controversial issue (for a full description of the instructions and the tasks, see Appendices 4A, 4B, 4C and 5).

9 It was not within the scope of the present study to provide a complete description of the interviewers’ methods in the oral proficiency interviews.
The three tasks represented a rising degree of difficulty in order to provide challenges for all students and to give them the opportunity to use the English language in different contexts and for various purposes. Task 1 was considered to be an easy task. The purpose of this task was to make the students talk with ease at the beginning of the test. Task 1 was labelled “to entertain”: the students were to relate a simple story with the help of some pictures or to compare two pictures. Task 2 was given the label “to inform”: the students were to choose one out of three topics and speak about their topic for a few minutes. The topics were entitled, for instance, “An important meeting in my life” and “A book I would like to read again”. The purpose of Task 2 was to make the students speak freely about their own experiences or memories. In Task 3 the students were expected to discuss a controversial issue and show their ability to argue for or against a specific position. Task 3 was labelled “to argue a case”. The students were to choose one of three topics. The topics were entitled, for instance, “The age for getting a driving license should be lowered” and “Scientific experiments on animals should be prohibited”. The students were given ten minutes of preparation for Task 1 and Task 2. Instructions for Task 3 were given during the test (cf. the discussion of prepared and unprepared tasks in 1.2. Oral Proficiency Testing).

For the purpose of illustration, the topics most often selected within each interviewer category are listed below:

Interviews with School Teachers: “It is necessary / not necessary to have marks at school” and “The age for getting a driving license should be lowered / raised” (These two topics were equally popular.)

Interviews with University Teachers: “It is better to start a family early in life / when you are older”.

Interviews with Native Speakers: “Keeping up with new trends in clothing is important / not important”.

It is evident that if a student tells exactly the same story twice or three times, his or her performance may improve in the second and third rounds. Differences in grades could then be explained by the student’s practice in telling a story or talking about an identical topic. This would have been the situation if the topics and the tasks had been identical in the Gothenburg students’ three tests. For that reason, similar tasks were used but with different topics in the three tests assessed by the three groups of interviewers.
7.5. Rating Criteria

Rating criteria in general are used to facilitate assessments and to give the assessors a common ground for their grades. Rating scales may thus be used to describe levels of language proficiency in guided criterion-referenced assessment (cf. 1.1. Measuring Student Performance). The scales describe different levels of student performance, and assessors are to decide what descriptions the student production corresponds best to. Qualifiers such as "sometimes", "most", "often" and "occasionally" are characteristic of the language of rating scales (Underhill 1987: 100). Rating scales typically have between three and nine levels (McNamara 2000: 41).

The rating criteria were more or less the same as those being formulated at the time by, among others, Docent Lindblad, for the new upper secondary school national syllabus, known as Lpf94 (cf. Introduction) and close to those used in the National Test of English for the Swedish gymniasieskola (Centralprov 1993, åk 2:3). There were some differences, however. More grading levels were used in the assessment in the Gothenburg project: five as against the three used in the National Test. This applies both to the Global and the Factorial grades. The intention was that all the five levels of the rating criteria should represent levels that might all be applicable in the assessment of the students who participated in the Gothenburg project.

The rating criteria of the Global Evaluation include descriptions on a five-step scale and range from

(A) – Distinction: Fluent and overall correct use of English
to
(E) – Poor: Poor production of English in many respects.

Parts of the wordings that occur in the global criteria of the other three grades (grades given in brackets) are
(B) production with minor inaccuracies,
(C) acceptable language, and
(D) often inaccurate.

The rating criteria of Factor 2a (Communication and Fluency) contain descriptions of the students’ initiatives. The criteria of four out of the five grades include descriptions of the students’ initiatives:

(A) frequently takes the initiative,
(B) often takes the initiative,
(C) sometimes takes the initiative, and
(D) rarely takes the initiative.

10 2:3 = the second year of the three- and four-year study programmes.
The rating criteria of Factor 2b (Grammar and Vocabulary) focus not only on the occurrence of errors but also on the use of vocabulary and structures. Accuracy is of certain importance of this factorial grade since the one of the two words error or mistake appears in the criteria of all the five grades:

(A) errors never disturbing,
(B) few errors when using complex structures,
(B) basic errors very rare,
(C) … a few noticeable basic errors,
(D) basic errors are noticeable,
(E) makes many kinds of mistake, and
(E) frequent basic errors.

Factor 2c (Pronunciation) emphasises on intonation and the pronunciation of individual sounds. The concepts “intonation” and “pronunciation of individual sounds” are used in the criteria of four out of the five grades. The criteria which describe intonation include:

(A) very good intonation,
(B) good intonation,
(C) acceptable intonation, and
(E) first language intonation.

Interference from Swedish is described in the rating criteria for Factor 2b (Grammar and Vocabulary) and, in particular, for Factor 2c (Pronunciation). Regarding Factor 2b, the criteria describe interference from Swedish in the use of vocabulary:

(E) Sometimes uses Swedish words.

Descriptions of the influence of the mother tongue appear in the criteria of three grades for the assessment of Pronunciation:

(B) some influence of native language clearly detectable
(C) foreign accent evident, and
(E) first language intonation.

The acceptability of errors is pointed out in the top criteria of Grammar and Vocabulary

(A) errors never disturbing
whereas the intelligibility of the student’s production is given prominence in the criteria of the grades at the bottom of the scale of Factor 2a:

(E) speech can be difficult to follow

and of Factor 2c:
(D) some utterances difficult to understand, putting strain on the listener, and
(E) difficult to understand due to pronunciation.

The rating criteria employed in the Gothenburg project thus reflect the levels of proficiency in English that the students taking part in the project could be expected to represent, i.e. students who were about to leave the Swedish gymnasieskola after studies in programmes preparing for academic education in 1993 (cf. the Introduction). They also focus on, for instance, the students’ degree of initiative and on interference from the mother tongue.

It might be of interest to compare the rating criteria of the Gothenburg project with other rating criteria developed to assess language proficiency. For comparison, the rating criteria developed in the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (2001; note: eight years after the Gothenburg project) will be scrutinized. The Framework document (CEF)\(^\text{11}\) contains several criteria and an outline of six broad levels is provided to give an adequate coverage of the learning space relevant to European language learners (ibid. 2001: 23). The six levels are higher and lower interpretations, respectively, of the common division into basic, intermediate and advanced. The six levels are

A: Basic User (A1 Breakthrough; A2 Waystage),
B: Independent User (B1 Threshold; B2 Vantage), and
C: Proficient User (C1 Effective Operational Proficiency; C2 Mastery).

The descriptions of the Common Reference Levels regarding qualitative aspects of spoken language use were selected for the comparison with the rating criteria employed in the Gothenburg project. The CEF rating scale contains criteria regarding five factors and rating criteria of the five factors – Range, Accuracy, Fluency, Interaction and Coherence – are thus provided for the six levels described above.

In what follows below (see Table 7.3.), the rating criteria of the Gothenburg project for Communication and Fluency, and Grammar and Vocabulary are compared with the rating criteria of the five factors of the Common Reference Levels in CEF.\(^\text{12}\) Formulations which I consider to be close in meaning are matched.

\(^\text{11}\) The abbreviation CEF is used, as it is for the Common European Framework, for instance, by Andered (2001).
\(^\text{12}\) The rating criteria for Pronunciation of the Gothenburg project were excluded in the comparison since there are no descriptions of intonation, stress, accent and the articulation of individual sounds in the Framework document.
Table 7.3. The rating criteria of Communication and Fluency versus the rating criteria of CEF

<table>
<thead>
<tr>
<th>The Rating Criteria of Communication and Fluency of the Gothenburg project</th>
<th>The Rating Criteria of CEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade A: “Performs with ease”</td>
<td>Level C1 of Range: “Without having to restrict what he/she wants to say”</td>
</tr>
<tr>
<td>Grade A: “Comfortable and completely fluent”</td>
<td>Level C1 of Fluency: “Can express himself/herself fluently and spontaneously, almost effortlessly”</td>
</tr>
<tr>
<td>Grade B: “Occasional hesitations and pauses for language search”</td>
<td>Level B2 of Fluency: “…he/she can be hesitant as he/she searches for patterns and expressions”</td>
</tr>
<tr>
<td>Grade C: “Sometimes produces longer and fairly coherent flow of language”</td>
<td>Level B2 of Fluency: “Can produce stretches of language with a fairly even tempo …”</td>
</tr>
<tr>
<td>Grade D: “Can generally only produce short and less complex utterances”</td>
<td>Level B1 of Coherence: “Can link a series of shorter, discrete simple elements into a connected linear sequence of points”</td>
</tr>
<tr>
<td>Grade E: “Constantly needs help”</td>
<td>Level A2 of Interaction: “Is rarely able to understand enough to keep conversation going of his/her own accord”</td>
</tr>
</tbody>
</table>

The criteria of the levels C1 – B2 – B1 – A2 of CEF at least partly match the criteria of the grades A – B – C – D – E of Communication and Fluency of the Gothenburg project. The rating criteria for Grammar and Vocabulary of the Gothenburg project are compared with the criteria of the five aspects of CEF in Table 7.4.

The criteria of the levels C1 – B2 – B1 – A2 of CEF at least partly also match the criteria of the grades A – B – C – D – E of Grammar and Vocabulary of the Gothenburg project. On the basis of the comparisons above, in spite of the fact that the rating criteria were constructed in different contexts, it may be concluded that the descriptions in the rating criteria of the Gothenburg project can be described in the following way according to the levels of CEF: from “Effective Operational Proficiency of the Proficient User (C1)” to “Waystage of the Basic User (A2)”.

7.6. The Organisation of the Testing Session

The testing was arranged so that the students would meet as many different interviewers as possible and that few students would be interviewed by the same three members of the three interviewer categories. The aim was to
provide as many combinations as possible of the nine members of the three interviewer categories.

The order in which the tests were carried out was also taken into account. One third of the students started with interviews assessed by School Teachers, one third with interviews assessed by University Teachers and, finally, one third had their first performance assessed by Native Speakers. Moreover, there was a break of at least 20 minutes between two oral proficiency interviews for each individual student.

On their arrival, the students were given a code: a letter (= the student) and three numbers (= the tests) (cf. Appendix 1). There were several reasons for this arrangement. Firstly, the aim was to facilitate both the identification of the individual student and his/her three interviews in further investigations. Secondly, the letters employed (instead of the students' names) made it possible to use the material in additional assessments without hav-
ing to reveal the students’ identity. In these additional rounds of assessments (for instance, a second assessment by the nine interviewers of the project) the students could be identified with the help of the numbers only. The numbers given to the students’ tests were arranged so that an assessor in an additional round of assessment would not have been able to identify the individual student with the help of the number. In the present study, a student’s performance is referred to as, for instance, A76 (A = the student; 76 = the number of the test; information on the interviewer categories is added where relevant). (The schedule of the testing on May 8, 1993, is provided in Appendix 3.)

7.7. Summary

The material of the Gothenburg project used in the present study consists of recorded oral proficiency interviews, the four grades assigned to each test, the interviewers’ comments on the tests and the answers to the questions of questionnaires. Twenty-nine Swedish students who were about to complete their studies at the Swedish gymnasieskola and who represented three different theoretical study programmes were interviewed. The participants who were asked to interview the students in the tests and assess their performance were three Swedish school teachers of English at the gymnasieskola, three university teachers of English and three native speakers of English. Three tasks were used which represented increasing levels of difficulty and were expected to give the students the opportunity to show their oral proficiency in English. Rating criteria were employed to grade the students and the interviewers were instructed not only to give an overall Global grade, but also to grade the students for Communication and Fluency, Grammar and Vocabulary, and Pronunciation. The overall focus was not on the grades given to the individual student but rather on the grades assigned by the members of the categories of interviewers.

The Gothenburg project was initiated, planned (the selection of students, the construction of tasks, the selection of interviewers, the organisation of the testing sessions, the principles in the assessments of the student performance and the formulation of the rating criteria) and carried out by Docent Lindblad of the Foreign Language Teaching and Testing Research Unit of the Department of Education, Gothenburg University.
Part 5:
The Results of the Project
CHAPTER 8
Grades Awarded to the Students’ Oral Proficiency

The assessment of the 87 oral tests carried out by the nine interviewers will be the topic of the present chapter. In the questionnaires and in spontaneous discussions after the completion of the interviews, the interviewers expressed their opinion that the students were good at speaking English. We will now see whether this generally positive attitude is confirmed by the grades. The results will be reported and briefly discussed in the present chapter, but will be subject to further discussion and interpretation in Chapter 14.

The assessment resulted in Global grades for all 87 oral tests. Factorial grades were given to almost all oral tests (Factor 2a to 85, Factor 2b to 84, and Factor 2c to 85 oral tests). Every test was thus assigned four different grades (Global grades and Factors 2a, 2b and 2c) with the help of the rating scale A-B-C-D-E (cf. Appendix 8. The Rating Criteria). For the Global grades, the following short descriptions were provided: A = Distinction; B = Very good; C = Pass; D = Not enough; E = Poor. (For all the Global and Factorial rating criteria, see Appendix 8).

In what follows the grades awarded will be presented, first the students’ Global grades and then the three Factorial grades. The three Factorial grades comprise Communication and Fluency (Factor 2a), Vocabulary and Grammar (Factor 2b), and Pronunciation (Factor 2c). Then the four grades assigned to each test will be compared. The focus will be not only on the grades awarded by the interviewer categories but also on the individual interviewer’s grades. Furthermore, the grades that each individual student obtained for his/her production will be presented.

1 School Teacher 1 did not assign Factorial grade 2a to two tests, Factorial grade 2b to three tests, and Factorial grade 2c to two tests.
8.1. Global Performance Grades

In Table 8.1., all Global grades are presented together with the mean values and the standard deviation across the three interviewer categories. Table 8.1. and Figure 8.1. show that the students are not a homogeneous group of high-flyers. This increases the interest in the other tables.

Table 8.1. shows that the highest mean grade (4.7) was given to students M and WW, and the lowest (2.0) to ZZ. Furthermore, it shows that the overall mean of the 29 students is 3.3, and that 13 students have means of 3 or 3.3. The rating criteria of the Global Evaluation are as follows:

<table>
<thead>
<tr>
<th>A = 5</th>
<th>B = 4</th>
<th>C = 3</th>
<th>D = 2</th>
<th>E = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluent and overall correct use of English.</td>
<td>Very good production with minor inaccuracies.</td>
<td>Acceptable language in spite of errors.</td>
<td>Not showing enough to be acceptable.</td>
<td>Poor production of English in many respects.</td>
</tr>
</tbody>
</table>

In terms of the rating criteria, it is possible to describe this mean level of the students roughly in the following way: the students’ language is acceptable in spite of some errors, and their production is sometimes even considered to be very good and only to have minor inaccuracies. As can be seen from Table 8.1., the three interviewers’ Global grades do not always agree. Most students received two different grades, and some students even three. This is interesting and will be discussed at some length (cf. 14.1. Global Performance Grades). At this stage we can observe that the University Teachers and the Native Speakers gave higher Global grades to the students than did the School Teachers. The grades (means and standard deviations) given by the individual interviewers within each of these categories are also quite different. The standard deviations of the grades assigned to the individual student in Table 8.1. point to differences between the interviewers. A standard deviation of 0 indicates that the three interviewers gave the same grades to a student. This is the case for eight out of the 29 students. It is worth observing that these eight students’ production was given either 3 or 4 (cf. Table 8.1.).

The grades assigned represent the middle standard of the list of criteria. It is also notable that four students were given three different grades by the three interviewers. Three of these four students were assigned the

---

2 For clarity, the grades in the Tables are presented in the following way: A=5; B=4; C=3; D=2 and E=1. For a discussion of the implications of transforming the original ‘letter grades’ awarded by the interviewers into ‘figure grades’, see 5.2. Global and Factorial Grades.

3 Cf. note 33 in Chapter 6.
grades 4, 3 and 2. Their performances were thus also rated in the middle of the scale.

Two students (R and S) were rated 4 and 2; student R was assigned the grade 4 by two interviewers and 2 by the third interviewer, student S the grade 2 by two interviewers and 4 by the third interviewer (standard deviation: 1.2). Both students were assigned low grades, i.e. 2, by the School Teachers and high grades, i.e. 4, by the University Teachers. Fifteen students were assigned two identical grades by two interviewers and a differ-

<table>
<thead>
<tr>
<th></th>
<th>ST 1</th>
<th>ST 2</th>
<th>ST 3</th>
<th>UT 1</th>
<th>UT 2</th>
<th>UT 3</th>
<th>NS 1</th>
<th>NS 2</th>
<th>NS 3</th>
<th>m</th>
<th>sd</th>
<th>m / sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>3.3</td>
<td>0.7</td>
<td>3.1/0.7</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>2.9</td>
<td>0.7</td>
<td>2.9/0.7</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>6.6</td>
<td>1.2</td>
<td>6.6/1.2</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.2</td>
<td>1.0</td>
<td>5.2/1.0</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.2</td>
<td>1.2</td>
<td>7.2/1.2</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.4</td>
<td>1.2</td>
<td>5.4/1.2</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td>1.2</td>
<td>6.0/1.2</td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.3</td>
<td>1.2</td>
<td>6.3/1.2</td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.4</td>
<td>1.2</td>
<td>5.4/1.2</td>
</tr>
<tr>
<td>J</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.4</td>
<td>1.2</td>
<td>7.4/1.2</td>
</tr>
<tr>
<td>K</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.4</td>
<td>1.2</td>
<td>7.4/1.2</td>
</tr>
<tr>
<td>L</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>5.5</td>
<td>1.2</td>
<td>5.5/1.2</td>
</tr>
<tr>
<td>M</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.4</td>
<td>1.2</td>
<td>9.4/1.2</td>
</tr>
<tr>
<td>N</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.6</td>
<td>1.2</td>
<td>6.6/1.2</td>
</tr>
<tr>
<td>O</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.6</td>
<td>1.2</td>
<td>6.6/1.2</td>
</tr>
<tr>
<td>P</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td>1.2</td>
<td>6.0/1.2</td>
</tr>
<tr>
<td>Q</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3.3</td>
<td>1.2</td>
<td>3.3/0.8</td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td>1.2</td>
<td>6.0/1.2</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>2.7</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.5</td>
<td>1.2</td>
<td>5.5/1.2</td>
</tr>
<tr>
<td>S</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.4</td>
<td>1.2</td>
<td>7.4/1.2</td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td>1.2</td>
<td>6.0/1.2</td>
</tr>
<tr>
<td>U</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td>1.2</td>
<td>6.0/1.2</td>
</tr>
<tr>
<td>V</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.6</td>
<td>1.2</td>
<td>6.6/1.2</td>
</tr>
<tr>
<td>W</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.4</td>
<td>1.2</td>
<td>5.4/1.2</td>
</tr>
<tr>
<td>X</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.4</td>
<td>1.2</td>
<td>5.4/1.2</td>
</tr>
<tr>
<td>Y</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td>1.2</td>
<td>6.0/1.2</td>
</tr>
<tr>
<td>Z</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.0</td>
<td>1.2</td>
<td>8.0/1.2</td>
</tr>
<tr>
<td>XX</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.0</td>
<td>1.2</td>
<td>8.0/1.2</td>
</tr>
<tr>
<td>YY</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.0</td>
<td>1.2</td>
<td>8.0/1.2</td>
</tr>
<tr>
<td>ZZ</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1.2/0.7</td>
<td></td>
<td></td>
<td></td>
<td>3.3</td>
<td>1.2</td>
<td>3.3/1.2</td>
</tr>
<tr>
<td>WW</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.4</td>
<td>1.2</td>
<td>5.4/1.2</td>
</tr>
<tr>
<td>m</td>
<td>3.3</td>
<td>2.9</td>
<td>3.2</td>
<td>3.6</td>
<td>3.5</td>
<td>3.3</td>
<td>3.7</td>
<td>3.1</td>
<td>3.3</td>
<td>3.3</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>sd</td>
<td>0.7</td>
<td>0.8</td>
<td>0.8</td>
<td>0.9</td>
<td>0.5</td>
<td>1.1</td>
<td>0.7</td>
<td>0.6</td>
<td>1.1</td>
<td>0.6</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>m / sd</td>
<td>3.1/0.7</td>
<td>3.4/0.8</td>
<td>3.3/0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ent one by the third interviewer (standard deviation: 0.6). In more than half of the instances, the School Teachers assigned these different grades (for eight of the 15 students). Six of these eight grades were lower than those assigned by the other two interviewers.

The means of the Global grades (as well as the standard deviations) of the individual interviewers are shown at the bottom of Table 8.1. For the interpretation of the figures, it should be borne in mind that the interviewers assessed different students. Nevertheless, it is worth observing that it is a School Teacher (School Teacher 2) that has the lowest mean. The School Teachers (as a category) also show the lowest mean. This finding is commented on in Chapter 14. The distribution across the different grades in terms of the students’ means in the three oral tests is illustrated in Figure 8.1. The figure shows that most students have means of 2.7, 3.0 or 3.3, i.e. in the middle of the scale. More students have their means on the upper half than on the lower half of the scale. No students had means of 2.3, 2.5, 3.5, 4.3, or 5.

8.2. Communication and Fluency Grades

Table 8.2 shows the distribution of the grades assigned for Communication and Fluency (Factorial grade 2a). The rating criteria are as follows

---

4 The University Teachers gave different grades to five of the 15 students and the Native Speakers assigned different grades to two of them.
The highest mean grade for Communication and Fluency was given to student M (5.0) and the lowest to ZZ (2.0). As can be seen in Table 8.2, eight of the 29 students were given the same grades by the three interviewers (standard deviation: 0). Seven of these eight students were assigned grades on the upper half of the rating scale, either 4 or 5. They represent the best performances judged by the interviewers according to the rating criteria. Five students (students F, Q, S, T and ZZ) were assigned three different grades (standard deviation: 1.0). Four of these five students were assigned the highest grades by the Native Speakers or by the University Teachers. One student (R) was assigned grade 2 by the School Teacher and grade 4 by the other two interviewers (standard deviation: 1.2). Thirteen students were assigned two identical and one different grade (standard deviation: 0.6). In the case of eight of these thirteen students it was the University Teachers who disagreed with the other interviewers, either by awarding a lower (four students) or a higher (four students) grade. The means of the interviewer categories reveal that the School Teachers gave the lowest grades and the Native Speakers the highest. The standard deviations obtained for the individual interviewers (within interviewer categories) vary from 0.5 to 1.1.

The results show that the students were given high grades for Communication and Fluency. These grades were even higher than those assigned in the Global evaluation, i.e. the overall impression of the student production. This is obvious in the grades given by the Native Speakers, in particular.

The distribution of the students’ means obtained for Communication and Fluency is illustrated in Figure 8.2. As shown in this figure, the majority of the students’ means are between 3 and 4. In terms of the rating criteria, students are thus “easy (or very easy) to understand”. Moreover, “they take

<table>
<thead>
<tr>
<th>A = 5</th>
<th>B = 4</th>
<th>C = 3</th>
<th>D = 2</th>
<th>E = 1</th>
</tr>
</thead>
</table>

\footnote{Students A and D were assigned two grades only, and these were different. The standard deviation shows 1 for two different grades assigned.}
the initiative in the conversation and contribute at natural speed and
tempo”. Figure 8.2. also reveals that the students’ means for Communication and Fluency tend to spread more than the means for the Global grades (cf. Figure 8.1.).

### Table 8.2. Factorial Grade 2a (Communication and Fluency) of the 29 students (A-WW) by School Teachers (ST), University Teachers (UT) and Native Speakers (NS)

<table>
<thead>
<tr>
<th></th>
<th>ST 1</th>
<th>ST 2</th>
<th>ST 3</th>
<th>UT 1</th>
<th>UT 2</th>
<th>UT 3</th>
<th>NS 1</th>
<th>NS 2</th>
<th>NS 3</th>
<th>m</th>
<th>sd</th>
<th>m / sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.5</td>
<td>1</td>
<td>3.4/0.9</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3.7</td>
<td>0.6</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZZ</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| m  | 3.4  | 3.3  | 3.6  | 3.7  | 3.4  | 3.5  | 3.8  | 3.6  | 3.6  | 3.5  | 0.6 |
| sd | 0.7  | 1.2  | 0.9  | 0.9  | 0.5  | 1.1  | 0.7  | 0.5  | 1.1  | 0.7  | 0.4 |

| m / sd | 3.4/0.9 | 3.5/0.8 | 3.7/0.8 |
| m / sd | 3.5/0.8 |
8.3. Grammar and Vocabulary Grades

The second Factorial grade (2b) pertains to Grammar and Vocabulary; the rating criteria are as follows:

<table>
<thead>
<tr>
<th>A = 5</th>
<th>B = 4</th>
<th>C = 3</th>
<th>D = 2</th>
<th>E = 1</th>
</tr>
</thead>
</table>

Table 8.3. shows that the mean is lower for this competence (3.1) than for Communication and Fluency (3.5) in Table 8.2. or the Global grade in Table 8.1. (3.3). A comparison of the means of the interviewer categories shows that the School Teachers have the lowest mean. This is in line with the tendency observed in the grades for Communication and Fluency. The grades presented in Table 8.3. show that the interviewers gave lower grades for Grammar and Vocabulary and that the grades varied more than the Global grades and the grades for Communication and Fluency (the standard deviations are slightly higher in Table 8.3. than in Table 8.1. or 8.2.). When
the grades of the individual student’s mastery of grammar and vocabulary are compared, Table 8.3. shows that the highest mean grade was given to student M (5.0). The three interviewers agreed on her top performance (standard deviation: 0). The lowest mean grade was assigned to student ZZ (2.0). In his case, the interviewers had very different opinions about his performance and gave him three different grades (i.e. 1, 2 and 3; standard deviation: 1). Seven of the 29 students were given identical grades by the

<table>
<thead>
<tr>
<th></th>
<th>ST 1</th>
<th>ST 2</th>
<th>ST 3</th>
<th>UT 1</th>
<th>UT 2</th>
<th>UT 3</th>
<th>NS 1</th>
<th>NS 2</th>
<th>NS 3</th>
<th>M</th>
<th>SD</th>
<th>M / SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>–</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2.5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3.3</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZZ</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>4.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>2.7</td>
<td>3.2</td>
<td>3.1</td>
<td>3.4</td>
<td>3.0</td>
<td>3.3</td>
<td>3.0</td>
<td>2.9</td>
<td>3.1</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.7</td>
<td>1.1</td>
<td>0.8</td>
<td>1.1</td>
<td>0.5</td>
<td>1.1</td>
<td>0.9</td>
<td>0.7</td>
<td>1.1</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m / SD</td>
<td>3.0/0.9</td>
<td>3.2/0.9</td>
<td>3.1/0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
three interviewers (standard deviation: 0; students A and D were only assigned two grades, cf. note 52). These seven students represent either a middle or a high standard in relation to the rating criteria (grades 3, 4 and 5). For the majority of the students, however, the interviewers did not give the same grades. Seventeen students were given two identical and one different grade, 3, 3 and 4, for instance (standard deviation: 0.6). The deviant grades were lower for ten students and higher for seven students. The deviant grades assigned spread across the three interviewer categories.6 One interviewer gave an altogether different grade to one student (T) (standard deviation: 1.2). The University Teacher and the Native Speaker agreed on grade 4, but the School Teacher gave a different grade (grade 2).

Figure 8.3. presents the distribution of the students’ means. Many students were assigned grades in the middle of the rating scale: 16 of the 29 students have means of 2.7, 3 or 3.3. With the help of the rating criteria their performances may be described in the following way: “The students have a good command of structure and vocabulary, but their production contains basic errors. They are sometimes dependent on paraphrase or simplification.”

---

6 For the ten students whose deviant grade was lower, the low grade was assigned by the School Teachers to three students, by the University Teachers to four students, and by the Native Speakers to three students. For the seven students whose deviant grade was higher, the high grade was assigned by the School Teachers to three students, by the University Teachers to two students, and by the Native Speakers also to two students.
8.4. Pronunciation Grades

Table 8.4. presents the grades for Pronunciation. The rating criteria for Pronunciation include descriptions of intonation, rhythm and individual sounds. They are as follows:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A = 5</td>
<td>Correct pronunciation of individual sounds and very good intonation and rhythm. A foreign accent may be discernible.</td>
</tr>
<tr>
<td>C = 3</td>
<td>Good pronunciation of most of the individual sounds and acceptable intonation. Foreign accent evident.</td>
</tr>
<tr>
<td>D = 2</td>
<td>Individual sounds poorly articulated. Some utterances difficult to understand, putting strain on the listener.</td>
</tr>
<tr>
<td>E = 1</td>
<td>Difficult to understand due to pronunciation. First language intonation. Poor articulation of individual sounds.</td>
</tr>
</tbody>
</table>

Again, student M was assigned top grades and has the highest mean (5.0). Student A has the lowest mean, of the two grades assigned (2.0)\(^7\). Eight of the 29 students were assigned the same grade for their pronunciation by the three interviewers (standard deviation: 0; student A was only assigned two grades). Five of these eight students were assigned grade 3. The remaining 21 students were thus assigned different grades. Six of them were given three different grades by the interviewers (standard deviation: 1.0)\(^8\). For five of these six students, the grades are centred in the middle of the scale, from 2 to 4. The University Teachers assigned the highest grades to four of the six students and the Native Speakers gave the middle grade 3 to five of them. One student (Q) was given a remarkably different grade by the School Teacher (grade 5), but identical grades by the University Teacher and the Native Speaker (grade 3) (standard deviation: 1.2). Twelve students were assigned two identical and one different grade (standard deviation: 0.6). For nine of them, the differing grade was lower; five of these low grades were given by the School Teachers.

The means of the interviewer categories are the same for School Teachers and Native Speakers (3.2), but there are slight differences in the spread of the grades (standard deviations: 0.9 and 0.7). This implies that grades given by the Native Speakers to Pronunciation tend to spread less than is the case with the two other interviewer categories. This finding is further investigated in 8.5. Comparative Discussion of the Four Grades Given, and discussed in 14.3. Interviewer Categories.

\(^7\) No Factorial grade was assigned by one interviewer (cf. note 52).

\(^8\) Students B and D were assigned only two grades. The two grades were different for both (standard deviation: 1.0).
Figure 8.4 shows the distribution of the student means. Ten students’ means were 3.0, and these students’ pronunciation can be summarized in the following way with the help of the rating criteria: “Good pronunciation with evident foreign accent.”

Figure 8.5 gives a survey of the four grades assigned to the student production discussed above. The figure shows that grade 4/B was given to approximately 40 oral tests for Communication and Fluency (Factor 2a). The students’ pronunciation was assigned grade 3 in approximately 45 cases. If grade 2 was given, it tended to be given for Grammar and Vocabulary (20 oral tests).

Table 8.4. Factorial Grade 2c (Pronunciation) of the 29 students (A-WW) by School Teachers (ST), University Teachers (UT) and Native Speakers (NS)

<table>
<thead>
<tr>
<th></th>
<th>ST 1</th>
<th>ST 2</th>
<th>ST 3</th>
<th>UT 1</th>
<th>UT 2</th>
<th>UT 3</th>
<th>NS 1</th>
<th>NS 2</th>
<th>NS 3</th>
<th>m</th>
<th>sd</th>
<th>m / sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>3</td>
<td>3.5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>3</td>
<td>2.5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>3</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>3</td>
<td>4</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3.7</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>5</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td>3.7</td>
<td>0.6</td>
<td>3.3/0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XX</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YY</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZZ</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WW</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4.7</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>3.6</td>
<td>3.0</td>
<td>3.1</td>
<td>3.2</td>
<td>3.6</td>
<td>3.3</td>
<td>3.3</td>
<td>3.2</td>
<td>3.0</td>
<td>3.3</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>sd</td>
<td>0.8</td>
<td>1.2</td>
<td>0.6</td>
<td>1.0</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>0.9</td>
<td>0.6</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m / sd</td>
<td>3.2/0.9</td>
<td>3.4/0.9</td>
<td>3.2/0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

125
Questions regarding the relationship between the Global grades and the Factorial grades are of great interest. If the Global grades and one of the Factorial grades, for instance Grammar and Vocabulary, are identical for many students, it could indicate that this aspect of the student production is the most important for the overall impression in the assessment by the inter-

---

**Figure 8.4.** Number of students receiving various mean Pronunciation grades.

**Figure 8.5.** The distribution of Global and Factorial grades to the tests. (A/5 = Distinction; B/4 = Very good; C/3 = Pass; D/2 = Not enough; E/1 = Poor).

8.5. Comparative Discussion of the Four Grades Given

Questions regarding the relationship between the Global grades and the Factorial grades are of great interest. If the Global grades and one of the Factorial grades, for instance Grammar and Vocabulary, are identical for many students, it could indicate that this aspect of the student production is the most important for the overall impression in the assessment by the inter-
The overall impression could also influence the Factorial grades assigned by the interviewers (cf. the halo-effect in 7.3. Sources of Variation). Table 8.5. shows the number of tests that were assigned the same or different Global and Factorial grades.

The three Factorial grades show the same pattern when compared with the Global grades assigned to the individual student’s production. The Factorial grades were different from the Global grades in roughly the same number of tests (24 or 25). This means that all three Factorial grades correlate equally with the Global grades.

The background and the experiences of the interviewers may no doubt have influenced the grades they assigned to the students’ production, but observed differences will mainly be followed up on the group level. Interesting tendencies may be observed in the distribution of the Factorial grades for Communication and Fluency, and for Pronunciation. Table 8.6. gives the three Factorial grades assigned to the student production by the three categories of interviewers: School Teachers (ST), University Teachers (UT) and Native Speakers (NS). These grades show two tendencies worth commenting on.

Firstly, the Native Speakers did not assign any student’s production the grade D/2 (= Not enough) for Factor 2a (Communication and Fluency) whereas the other interviewers did so. (One student however was assigned the lowest grade 1 for Communication and Fluency by a Native Speaker.) When commenting on the grades, the Native Speakers also emphasised aspects of oral production such as “actually getting the message across”. Secondly, the grades for Factor 2c (Pronunciation) tend to be in the middle of the scale. Most student production was rated C/3 (= Pass), particularly when assessed by the Native Speakers. The Native Speakers were slightly more reluctant to award the grades A/5 or B/4 for Pronunciation than were the School Teachers or University Teachers. The correlation coefficients

<table>
<thead>
<tr>
<th></th>
<th>Factor 2a Communication and Fluency</th>
<th>Factor 2b Grammar and Vocabulary</th>
<th>Factor 2c Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global grade the same</td>
<td>60</td>
<td>60</td>
<td>61</td>
</tr>
<tr>
<td>Global grade different</td>
<td>25</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>84</td>
<td>85</td>
</tr>
<tr>
<td>Correlations with Global grades</td>
<td>0.78**</td>
<td>0.81**</td>
<td>0.78**</td>
</tr>
</tbody>
</table>
between the categories of interviewers’ Global and Factorial grades are provided in Table 8.7.

Table 8.7. shows three tendencies worth pointing out. Firstly, the School Teachers’ and the Native Speakers’ grades for Grammar and Vocabulary correlated more (0.77 and 0.89) with the Global grades than did their grades for Communication and Fluency, and for Pronunciation. Secondly, the University Teachers’ grades for Communication and Fluency showed the highest correlation coefficient (0.90). Thirdly, the School Teachers’ grades for Communication and Fluency showed the lowest correlation (0.65). As a whole, it is interesting to observe the generally high correlation coefficients and the low correlation figure for Communication and Fluency in the grades assigned by the School Teachers.

The disagreement in the grades awarded to a student’s three oral tests by the three interviewers is worth discussing at some length (cf. Chapter 14), with the focus on the grades given by the individual interviewer. In Tables
8.1. to 8.4., we could observe that the interviewers assigned two or even three different Global or Factorial grades to certain students. The differences were described in terms of standard deviation (0.6, 1.0 and 1.2). It is of interest to look more closely at the students whose three tests were assigned three different grades (= standard deviation: 1.0), and the students’ tests which were given one divergent grade besides two identical ones (with one grade apart), for instance 4 by one interviewer, and 2 by two interviewers (= standard deviation: 1.2). Students whose production was assigned highly different grades are discussed in 8.6. Grades Assigned to the Individual Students, and in Chapter 14 (cf. 14.4. Individual Students’ Grades). In what follows, the focus is on the interviewers.

As for the Global grades, six students (C, L, R, S, XX and ZZ) were assigned very different grades (standard deviation: 1.0 or 1.2; cf. Table 8.1.). For two of them (L and R), the School Teacher, and more precisely School Teacher 2, gave the lowest grade (of the three grades assigned). The grades for Communication and Fluency, and Pronunciation show the same tendency. School Teacher 2 gave the lowest grade for Communication and Fluency to three (Q, R and S) of the six students (F, Q, R, S, T and ZZ) who were assigned very different grades (standard deviation: 1.0 or 1.2; cf. Table 8.2.). The grades for Pronunciation showed great differences for seven students (C, J, L, O, Q, T and XX). School Teacher 2 graded four of them (L, O, Q and T), and to three of them he gave the lowest grade of the three grades assigned, for instance 2 when the other interviewers gave 4 and 3. We may conclude that when differences in grades awarded to a student were found, School Teacher 2 gave the lowest grade to half of these students.

An investigation of the grades assigned by the individual interviewer can provide information regarding whether members of one of the interviewer categories tended to show more variation, or consensus with the other two interviewer categories. For that reason, it is also of interest to study the individual interviewer’s assessment when the interviewers agreed on the grades (cf. Table 8.8.). In Tables 8.1. to 8.4. we could see that some students were assigned the same grade in their three tests (standard deviation: 0). The interviewers agreed on the Global grades for eight students. The grades awarded by Native Speaker 2 are found in the grades assigned to five of them and the grades assigned by School Teacher 3 appear in four of them. This tendency can also be seen in the grades for Grammar and Vocabulary. School Teacher 3 and Native Speaker 2 gave the same grades as other inter-

9 School Teacher 2 gave the lowest Global grade to three of six students, the lowest grade for Communication and Fluency to three of six students, the lowest grade for Grammar and Vocabulary to two of four students and the lowest grade for Pronunciation to four of seven students.
viewers to three of the five students who were given the same grades in their three interview.

As for Communication and Fluency, University Teacher 1 gave the same grades as the other interviewers to five of the eight students who were assigned the same grades.

We may conclude that School Teacher 3 and Native Speaker 2 tended to give the same grades as other interviewers did in their Global grade and in their grade for Grammar and Vocabulary. University Teacher 1 did so in his grades for Communication and Fluency. These three interviewers thus tended to show more consensus with other interviewers in their assessments. Further comments are given to these findings concerning the grades assigned by the individual interviewers in Chapter 14 (cf. 14.3. Interviewer Categories).

8.6. Grades Assigned to the Individual Students

As described in Chapter 7, all the students were tested three times by three interviewers. In this section, the differences in the grades assigned by the three interviewers to the individual student will be discussed in more detail.

Table 8.8. displays how many students were assigned the same grades by the three interviewers and the number of students who were assigned different grades by their three interviewers (the Global and the three Factorial grades are presented in separate columns).

Table 8.8. shows that for a majority of the students, 21 or 22 students out

<table>
<thead>
<tr>
<th>Global grades</th>
<th>Factorial grades (2a) for Communication and Fluency</th>
<th>Factorial grades (2b) for Grammar and Vocabulary</th>
<th>Factorial grades (2c) for Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The same grade in the student’s three tests</td>
<td>8</td>
<td>8</td>
<td>7a</td>
</tr>
<tr>
<td>Different grades in the student’s three tests</td>
<td>21</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
</tbody>
</table>

a Two of these seven students were not assigned a Factorial grade in one of the three tests. (cf. Table 8.3.)
b One of these eight students was not assigned a Factorial grade in one of the three tests. (cf. Table 8.4.)
of the 29, the assessors disagreed in their assignment of Global and/or Factorial grades. The same tendency is seen in all Global and Factorial grades.

It is interesting to investigate the degree of divergence in the grades assigned to the individual student’s three oral tests, e.g. 5 – 5 – 4 (Student M) and 4 – 3 – 2 (Student C), (cf. Table 8.1.). Table 8.9. gives the number of students who were assigned different Global and Factorial grades for their production (“Two” stands for two different grades, e.g. 5 – 4 – 4; “Three” for three different grades, e.g. 5 – 4 – 3). Only students whose production was assigned different grades are included in Table 8.9. (cf. Table 8.8.).

Table 8.9. shows that most of the students who were assigned different grades in their three tests were rated differently in only two of the total of three tests (e.g. 4 in two tests, 3 in one). A few students, between three and six, were assigned different grades in all three of their tests (e.g. 4 in one, 3 in the second, and 2 in the third test). Very few of the students, i.e. only one or two students, were assigned two identical grades and one remarkably different grade (in their three tests (e.g. 4 in two tests, 2 in one).

Notice that the grades for Communication and Fluency and for Pronunciation show a slightly greater diversity in the individual student’s three tests than do the grades for Grammar and Vocabulary. Figure 8.6. illustrates these findings.

The two types of non-agreement of grades assigned to the individual students’ three tests (e.g. 4-4-3 or 4-4-2, cf. Table 8.9.) are presented together in the category “Two different grades”.

---

**Table 8.9. Number of students who were assigned the same, two different or three different Global or Factorial grades in the three tests by the three different interviewers**

<table>
<thead>
<tr>
<th></th>
<th>Global grade</th>
<th>Factor 2a Communication and Fluency</th>
<th>Factor 2b Grammar and Vocabulary</th>
<th>Factor 2c Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The same grade, e.g.</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>4-4-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two different grades,</td>
<td>15</td>
<td>15</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>one-step differential,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. 4-4-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two different grades,</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>two-step differential,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. 4-4-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three different grades,</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>e.g. 4-3-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
</tbody>
</table>

*The students who were not assigned one Factorial grades in one of the tests are included in Table 8.9. (cf. Table 8.8.).*
The students who were given three different grades in their three tests are of interest. Four students were given three different Global grades by the three interviewers. Three of them (students C, L and XX) were assigned grades from the middle of the scale, namely 4, 3 and 2. The fourth (student ZZ) was given the grades 3, 2 and 1. This tendency of assigning grades in the middle of the scale can also be observed in the grades for Communication and Fluency. Five students were given three different grades (students F, Q, S, T and ZZ). The grades given to three of them (students F, Q and S) were 4, 3 and 2. Only three students were given three different grades for Grammar and Vocabulary (cf. Figure 8.6.). For two of them, the grades were 4, 3 and 2. The grades for Pronunciation show slightly more differences: six students were given three different grades. Five of them were given grades in the middle of the scale (4, 3 and 2). Thus the differences in the grades assigned seem to concern the assessment of “average” students and the interpretation of the rating criteria in the middle of the scale. This finding is further developed in Chapter 14 and in Reflections in the Conclusion.

8.7. Grades in Relation to the Length and Order of the Interviews

The grades assigned might have been influenced by the length measured in terms of number of words or minutes of the student performance, whether the interview was the individual student’s first, second or third one, and whether the student was interviewed at the beginning or the end of the testing session by the interviewers (cf. Appendix 3. The Organisation of the Tests on May 8, 1993).
In this section, the following relationships are therefore presented:
(a) the Global grades and the number of words produced by the students,
(b) the Global grades and the length in time of the production, and finally,
(c) the Global grades and the order of the production, i.e. the individual student’s first, second and third interview of the testing session (cf. 7.6. The Organisation of the Testing Session).

(a) Number of words in student production and Global grades assigned. As described in 6.3., the student performances vary in length. In the shortest interview a student spoke 480 student words and in the longest one another student used 2,010 words. Figure 8.7. illustrates the relationship between the Global grades assigned and the number of student words in each interview. The results show that interviews which were given low grades, i.e. D (2) and 1 (E), consist of 500 to 700 words or 701 to 900 words respectively (the latter only grade D).

The interviewers did not assign low grades to performance which was longer than 900 words. It is thus evident that the students to whom the interviewers gave low grades did not speak at length. Long performance was assigned grades in the middle of the scale, i.e. the grades 4 and 3. Many of the tests which were assigned grade 4 were long. The tests assigned the highest grade were of average length (701 to 1100 words). The overall correlation between the Global grades assigned and the number of students’ words was 0.31.\textsuperscript{11} Thus, it seems that little student talk, i.e. rela-

\textsuperscript{11} The correlations between the three Factorial grades and the length in words were as follows: Communication and Fluency: 0.47; Grammar and Vocabulary: 0.20; Pronunciation: 0.21.
tively few words in a test, is related to low Global grades (for further discussion of this observed tendency, see Chapter 14).

(b) Length of student production in minutes and Global grades assigned. The length of the student production was not only measured in students’ words but also in minutes. The shortest student’s production was about 8 minutes long, the longest slightly more than 20 minutes. Figure 8.8. illustrates the relationship between the Global grades assigned and the length of the student production in minutes. The results show no very clear relationships but the same tendency as the one described above can be observed; student production assigned a Low grade, and in particular the grade 2, tends to be short, i.e. between 8 and 10 minutes long. The correlation was 0.13 between the Global grades and the length in minutes.\textsuperscript{12} This correlation and Figure 8.8. show that there is no clear relationship between the length of the interviews in time (as opposed to in words, see above) and the grades awarded.

(c) The students’ first, second and third tests and the Global grades assigned. Finally, the relationship between the grades assigned and the order of the performance has been investigated. Figure 8.9. displays the distribution of the grades in the students’ first, second and third interviews of the testing session. Slightly more students were assigned low grades, in particular grade 2, in their first production than in the second or third. The opposite tendency can be observed as regards the high grade 4; fewer students were assigned this high grade in their first interview than in their second or third test.

\textsuperscript{12} The correlations between the three Factorial grades and the length in minutes were as follows: Communication and Fluency: 0.13; Grammar and Vocabulary: 0.06; Pronunciation: 0.11.
third. The order of the interviews does not seem to be of importance for the students who were assigned the top grade 5. The findings above are further discussed in Chapter 14.

The Global grades assigned were also compared with the order in which each interviewers assessed each test to see whether the order of the tests could have contributed to low or high grades (cf. the organisation of the tests in Appendix 3, and the effect of sequencing in 2.3. Reliability). The results of the investigation showed no clear tendencies. Both high and low Global grades were given by the individual interviewers at the beginning and end of the testing day. Nor did the Global grades assigned by one of the three interviewer categories show any clear tendency.

8.8. Grades in Relation to Background Variables

This section focuses on the Global grades assigned to student production in relation to a number of background variables. The Factorial grades are included in the discussion only where it seems relevant. First the distribution of the Global grades are studied relative to the students’ sex, age, educational profile (the course programme studied in the Swedish gymnasieskola) and their school grades in English from the previous autumn term. Then the relationship between the grades assigned and the results obtained in the National Test of English one year earlier are dealt with. Finally, students’ experience of English (time spent in English-speaking countries, time spent in other countries) in relation to the grades assigned in the tests is discussed.

Table 8.10. below illustrates the distribution of the Global grades for the boys and the girls.

By and large, the grades of girls’ and boys’ performance had a similar distribution of the three grades 4, 3 and 2. The top Global grade 5 was only assigned to girls’ production.
A great majority of the students were 19 years old when they took the test, five of them were 20 or 21 years old. The Global grades should be compared with the students’ age, although it might seem unlikely that one or two years of difference in age would play a significant role here. The results of this comparison are presented in Table 8.11.

Table 8.11. Global grades by age

<table>
<thead>
<tr>
<th>Grade Age</th>
<th>A/5</th>
<th>B/4</th>
<th>C/3</th>
<th>D/2</th>
<th>E/1</th>
<th>Total tests</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>4</td>
<td>28</td>
<td>30</td>
<td>10</td>
<td>0</td>
<td>72</td>
<td>24</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>31</td>
<td>39</td>
<td>12</td>
<td>1</td>
<td>87</td>
<td>29</td>
</tr>
</tbody>
</table>

A great majority of the students were 19 years old when they took the test, five of them were 20 or 21 years old. The Global grades should be compared with the students’ age, although it might seem unlikely that one or two years of difference in age would play a significant role here. The results of this comparison are presented in Table 8.11.

Table 8.11. shows that the age difference is not an important factor for the high Global grades assigned. A majority of the students who were 20 and 21 years old at the time of the testing were assigned average grades (9 out of 15). Of the interviews with the 19-year-olds, 42% were assigned the average grade 3, compared with 60% of the interviews with the 20–21-year-olds.

As pointed out above, students from three course programmes of the Swedish gymnasium were represented in the material of the Gothenburg project: four from “Natural Sciences”, twelve from “Social Sciences” and thirteen from “Economics” (cf. 7.2. Students). The Global grades assigned to the students are examined in relation to these three course programmes in Table 8.12.

Table 8.12. shows a slight tendency that the students who were taking “Natural Sciences” were assigned higher grades than other students. (The average grades are as follows: Natural sciences: 3.9; Social sciences: 3.4; Economics: 3.0.) This is not surprising. When the project was carried out in 1993, the course programme “Natural Sciences” was known to attract students with the highest grades from compulsory school.
At the time of the testing, a norm-referenced 1–5-point grading system was still in use, 5 being the top grade. No students included in the Gothenburg study had the lowest grade 1 in English. The students were self-selected which may account for the lack of truly poor students (cf. 7.2. Students). In Table 8.13, the Global grades are compared with the students’ grades in English at school. The correlation between the Global grades and the school grades is 0.47.

The results above give rise to the following three observations. Firstly, in general more students were given higher than lower grades in the oral tests compared with their school grades in English. Secondly, the Global grades assigned to the students with the two top school grades 4 and 5 do not show any major differences in distribution. They spread in a similar way across the three Global grades A/5, B/4 and C/3. (Students with the school grade 5: 2 (15%), 7 (60%), and 3 (25%) tests; students with the school grade 4: 2 (8%), 13 (55%) and 8 (33%) tests). Thirdly, the Global grades assigned to students who have the low grade 2 in English at school do not differ, proportionally speaking, from those assigned to students who had the average grade 3. (Students with the school grade 2: 2 (22%), 5 (56%), 1 (11%) and 1 test (11%); students with the school grade 3: 9 (21%), 23 (55%) and 10 tests (24%).) This indicates that some low-achievers had their strength in oral proficiency. According to this tendency observed, it may be assumed

<table>
<thead>
<tr>
<th>Grades</th>
<th>Course programme</th>
<th>A/5</th>
<th>B/4</th>
<th>C/3</th>
<th>D/2</th>
<th>E/1</th>
<th>Total tests</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural sciences</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Social sciences</td>
<td>2</td>
<td>13</td>
<td>18</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>Economics</td>
<td>0</td>
<td>11</td>
<td>18</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>39</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>31</td>
<td>23</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>87</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 8.12. Global grades by course programme

<table>
<thead>
<tr>
<th>Global grades</th>
<th>School grades</th>
<th>A/5</th>
<th>B/4</th>
<th>C/3</th>
<th>D/2</th>
<th>E/1</th>
<th>Total tests</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>13</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>9</td>
<td>23</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>31</td>
<td>39</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>87</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 8.13. Global grades versus school grades in English

At the time of the testing, a norm-referenced 1–5-point grading system was still in use, 5 being the top grade. No students included in the Gothenburg study had the lowest grade 1 in English. The students were self-selected which may account for the lack of truly poor students (cf. 7.2. Students). In Table 8.13, the Global grades are compared with the students’ grades in English at school. The correlation between the Global grades and the school grades is 0.47.

The results above give rise to the following three observations. Firstly, in general more students were given higher than lower grades in the oral tests compared with their school grades in English. Secondly, the Global grades assigned to the students with the two top school grades 4 and 5 do not show any major differences in distribution. They spread in a similar way across the three Global grades A/5, B/4 and C/3. (Students with the school grade 5: 2 (15%), 7 (60%), and 3 (25%) tests; students with the school grade 4: 2 (8%), 13 (55%) and 8 (33%) tests). Thirdly, the Global grades assigned to students who have the low grade 2 in English at school do not differ, proportionally speaking, from those assigned to students who had the average grade 3. (Students with the school grade 2: 2 (22%), 5 (56%), 1 (11%) and 1 test (11%); students with the school grade 3: 9 (21%), 23 (55%) and 10 tests (24%).) This indicates that some low-achievers had their strength in oral proficiency. According to this tendency observed, it may be assumed
that particularly the Factorial grades 2c (Pronunciation) assigned to these students should show the same tendency, and possibly more strongly (as opposed to Factor 2a, Grammar and Vocabulary, which is probably tested in many other contexts at school). Figure 8.10.\textsuperscript{13} tests this hypothesis.

Figure 8.10. confirms that students with the low grade 2 in English at school were not awarded consistently lower factorial grades for pronunciation than were students with the average school grade 3. The correlation between the Pronunciation grades and the School grades was 0.40.

Another tendency in Figure 8.10. is worthy of our attention. The results show a slight difference in the proportions of the grades assigned to the students who have the high grade 4 at school. In Table 8.13, 13 out of the 24 oral tests (i.e. 54\%) were assigned the Global grade 4, whereas in Figure 8.10., 8 of the 24 oral tests (i.e. 33\%) were assigned the Factorial grade 4 for Pronunciation. These students’ grades for Pronunciation were thus somewhat lower than their Global grades. No other such differences in proportions can be observed.

It is also of interest to study the relationship between the Global grades and the results the students had obtained in the National Tests of English. These results are in the form of the grades 1 to 5. The students had taken this test the year before the material of the Gothenburg project was collected, i.e. in the spring term of 1992.\textsuperscript{14}

As a whole there are certain similarities between the students’ Global grades of the oral tests and the National Test Results, which are based on overall proficiency, except that there was, at the time, no oral test included. The correlation coefficient between the Global grades and the Grades in the National Test is 0.40. It is interesting to notice that students with poor results in the National Test, i.e. the grades 1 or 2, may still have high or

\textsuperscript{13} Two tests were not given grades for Pronunciation. The students in these tests had grade 3 in English at school.

\textsuperscript{14} This National Test was for the three- and four-year study programmes (cf. the Introduction).
average Global grades assigned to their oral tests in the material of the Gothenburg project.

As was mentioned above, the students were asked about their experience of stays in English-speaking countries. The purpose was to see how many of the students had such experience and whether this background variable seemed to be of importance for the grades assigned. The answers in the questionnaire showed that one student (L) had spent 16 weeks in English-speaking countries. This was considerably more than was the case with the other students. The grades of student L’s three oral tests were overall 4 or 3 except for the grade for Grammar and Vocabulary in one assessment that was the low grade 2. This student was thus not among those who were assigned the top grades. The same tendency could be observed when the grades (the Global grades and the grades for Communication and Fluency) of twelve students who had experience from English-speaking countries15 were compared with those of seven students who had no such experiences. The mean Global grade was 3.15 for the former and 3.24 for the latter group of students (the mean grades for Communication and Fluency were for the former 3.51 and for the latter group 3.50). To sum up, experience from stays in English-speaking countries does not seem to have had any effects on the grades assigned.

### 8.9. Students’ Opinions about the Tests

The students were asked to answer 17 questions about the test in a questionnaire after the testing sessions (for the questionnaire, see Appendix 7). In what follows, the students’ answers to three of the questions are discussed:

<table>
<thead>
<tr>
<th>Global grades</th>
<th>Grades in the National Test</th>
<th>A/5</th>
<th>B/4</th>
<th>C/3</th>
<th>D/2</th>
<th>E/1</th>
<th>Total tests</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>18</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>14</td>
<td>18</td>
<td>4</td>
<td>0</td>
<td>36</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>30</td>
<td>34</td>
<td>10</td>
<td>0</td>
<td>78</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

The results from the National Test were not available for three students. One of these three students was assigned an E as the Global grade for his performance.
– whether the students thought they were given the opportunity to show their competence,
– what they thought about the tasks, and
– what they thought about the interviewers’ methods.

Twenty-six of the 29 students answered that they thought that they had been given the opportunity to show their competence in English in the oral proficiency interviews.

In the answers to the question regarding the tasks, it was evident that the students appreciated the discussion in task 3 (“to argue a case”) most. Task 3 was considered to be quite authentic-like and stimulating, although some of the students thought it was difficult to discuss some issues without any time for preparation. Another reason for their positive attitude to the last task was that they appreciated that they were given the opportunity to express their own opinions.

As for the interviewers’ methods in the interviews, the students tend to prefer the conversations with the Native Speakers; 23 students answered that the Native Speakers were easy to talk to, whereas only 13 to 16 students, i.e. half of them, thought so about the other two interviewer categories.

8.10. Summary

In this chapter the grades assigned to the student production were presented. The Global grade and the three Factorial grades were compared and the three interviewer categories and the grades assigned to the individual students’ three tests were taken into account. The relationships between the grades assigned and a number of background variables were discussed.

The results of the investigation of the Global grades indicated that more students were rated on the upper than on the lower half of the rating scale. The rating criteria of these grades indicate that the interviewers regarded the students’ output as “very good production” or “acceptable language in spite of errors”. The Factorial grades suggest that the students were considered as being slightly better at Communication and Fluency than at Grammar and Vocabulary. The grades for Pronunciation tended to cluster in the middle of the scale, particularly when assessed by the Native Speakers.

Generally, the School Teachers gave lower grades, both Global and Factorial than did the other two interviewer categories. The Native Speakers gave the highest grades for Communication and Fluency. Differences could be observed in the grades assigned by individual interviewers within an interviewer category.

A majority of the students were assigned the same grades in two or all
three of their tests, but some students were given two, or even three, different Global or Factorial grades.

As for background variables, it was shown that age, sex, and time spent in English-speaking countries did not seem to influence the grades assigned. The Global grades correlated slightly with the students’ grades obtained in English at school and with their results from the Swedish National Test of English, although students with low grades there were sometimes assigned high grades in the interviews, in particular for Pronunciation.

The results are further discussed and interpreted in Chapter 14.
In this chapter the ten linguistic features (cf. 6.2. Linguistic Features) will be discussed with respect to the criteria for classification and the principles adopted in the presentation of the results in Chapters 10 to 13. The ten linguistic features selected are as follows:

1. the third person singular marker,
2. the progressive,
3. abstract nouns,
4. modifiers of adjectives,
5. lexical errors and compensatory strategies,
6. well,
7. you know,
8. intonation,
9. stress, and
10. speech rate.

Illustrative examples will be given from the student performance for each linguistic feature.¹ (For the reasons and principles in the selection of linguistic features, see 1.3. Linguistic Analyses of Students’ Proficiency, and 6.2. Linguistic Features.)

9.1. Verbal Group

There are good reasons for discussing some common problems in the students’ use of verbal constructions (cf. 1.3. Linguistic Analyses of Student’s Proficiency). In general, the knowledge of how to make proper use of verbs has been claimed to be important from a communicative point of view in a study on reactions to non-native English (Hultfors 1986: 188). Further-

¹The student’s letter and the test number are added to each example, for instance U6 = Student U in test number 6 (cf. 7.6.).
more, teachers are fully aware of the importance of a student’s mastery of
the verbal group in communication, both in writing and in speaking, and
they also recognise that students at different levels can have great difficul-
ties with the use of verbs.

In the material of the Gothenburg project, the samples of student produc-
tion show great differences not only in the frequencies of the instances but
also in the types of some verbal expressions used. This is why the focus in
the present analysis is on both quantitative and relevant qualitative features
of use in the investigations of the students’ uses of the third person singular
marker and the progressive. Their distribution is compared with the three
test rating categories High, Average and Low in order to ascertain the im-
portance of these features for the impression of a successful oral produc-
tion.

9.1.1. The Third Person Singular Marker

Occurrences of the third person marker are discussed in Chapter 10 in rela-
tion to the test rating categories High, Average and Low (cf. 6.2. Classify-
ing Student Performance into Test Rating Categories), together with the
distribution of the instances across the interviewer categories (School
Teachers, University Teachers and Native Speakers) and the three tasks (to
entertain, to inform, to argue a case). Two background variables are also
taken into account (sex and grades in English at school).

In the classification, non-standard English uses include instances of the
third person singular marker not recognised in prescriptive terms, as op-
posed to standard English instances which occur in contexts where they are
expected to occur according to descriptions in school grammars (for a dis-
cussion of these terms, see 6.4. Defining Standard English and Non-stand-
ard English Instances).

The two categories of analysed instances are thus:

standard English:       he takes
non-standard English:
    type 1    *he take
    type 2    *they takes
    type 3    other non-standard English instances with the
                        third person singular marker

Examples (1), (2) and (3) below from the data are given to illustrate the
three subcategories of the non-standard English uses:

type 1: (1) and then he fall asleep (L87)
type 2: (2) his parents dies (G82)
type 3: (3) an animal who gots hurt (H83)
Co-ordinated structures such as *he plays and sings* occur in the material. Since instances such as *he plays and sing* emerged, every verb is counted as one instance. Self-corrected instances are not included in the discussion, nor are modals included as they lack number contrasts.

9.1.2. The Progressive

There is evidence from previous studies of students’ difficulties and the negative impression that may be caused by the inappropriate use of the progressive (cf. 1.3. Linguistic Analyses of Students’ Proficiency). The interviewers of the Gothenburg project also pointed out that the improper use of the progressive was a frequent feature in the student production. The following answer to a question in the interviewers’ questionnaire is given for illustration:

… och sedan så naturligtvis användningen utav den progressiva formen som mycket sällan användes på rätt sätt. (… and then of course the use of the progressive which was very seldom used in a correct way.) (University Teacher 1).

Besides the reasons discussed in 1.3, this answer suggests that the students’ use of the progressive in oral production is a relevant topic to investigate in the linguistic analysis. The three instances below illustrate the students’ use of the progressive in the material.

(4) “well it is really something if you’re coming as a tourist to Gothenburg (Z26)”

(5) “he makes a phone call and I think he’s calling the cops or the police and that’s how he gets out” (Q2)

(6) “or I think I would be banging on the door I think” (G82)

The variation in the use of the progressive and simple forms in present-day English may lead to difficulties when deciding whether students’ use of the progressive should be considered to be adequate or inadequate. Different criteria can be used in the assessment. In the present study, the instances were classified in standard English and non-standard English categories on the basis of the descriptions found in three Swedish school grammars (cf. 6.4. Defining Standard English and Non-standard English Instances). Only semantic and pragmatic contextual features were considered in the discussion and classification. Three categories of non-standard use of the pro-

---

2 167 instances of self-corrections of verbs, nouns, prepositions and pronouns emerged in the student production.

3 Instances of the non-standard English type (*the cans*) did not appear in the material.

4 The reason was that relatively few instances (15%) contained incorrect verb forms (errors of concord or tense).
gressive were identified in the analysis of the student production. According to the three school grammars consulted, the simple form is preferred in these contexts (Hargevik and Hargevik 1993: 179; Svartvik and Sager 1975: 141–143; Ljung and Ohlander 1992: 87–93). The three categories of non-standard use of the progressive are:

1. Habit (The student talks about what he/she regularly does.)
2. Objective circumstance (The student talks about facts or completed actions.)
3. Stative verbs (The student uses verbs which describe the state of things, for instance possess, like, contain, resemble, cost).

Four examples are given to illustrate the two categories of standard English and non-standard English instances and the various uses of the progressive in the student production. First two examples of standard English use and then two examples of non-standard English use are given.

**Standard English use:**
In (7), the student describes what she is doing while speaking in the interview.

(7) “now I’ve chosen to talk about my hobby my hobby is horses that’s why I’m looking out the window” (G68).

In (8), the student argues for lowering the age when one can start practising for getting a driving license.

(8) “if you get eh what do you say well you’ve been doing it for such a long time so it’s not a big deal any longer I think” (XX12).

**Non-standard English use:**
In (9), the student discusses whether the age for getting a driving license should be lowered or raised. In this discussion, he puts forward a fact from experiments. This instance is considered to be non-standard English according to the second criterion (cf. above, Objective circumstance).

(9) “it is showing in the experiments that in statistics that elderly drivers are better cause they are more experienced” (Y10) (instead for: it has been shown).

In (10), the student describes why one house is expensive to rent in Task 1 (cf. Appendix 4B). This instance is regarded to be non-standard English.

5 These three Swedish school grammars do not describe all functions of the progressive. For a discussion of the uses of the progressive, see for instance Rydén (1977).
according to the third criterion (cf. above, Stative verbs) because of the use of *cost* in the progressive.

(10) “yeah but I think it’s the area that’s costing you” (L42) (instead for: *the area that will cost you*).

In order to provide a picture of the contexts where the progressive is used in the student production, it is also of interest to investigate in what contexts the standard English instances appear. Task 1 was a narrative task accomplished with the help of a series of pictures which elicited instances of the progressive. The proportions of standard English instances that are clearly related to Task 1 in the students’ production are also presented in 10.2. The Progressive. The progressive in these instances is used to describe an action in progress in some pictures. These figures are compared with the instances of other uses of the progressive in the students’ production. The standard English instances were thus classified in two categories.

1. The progressive used to describe the action in progress of the pictures in Task 1.
2. The progressive used in other contexts.

Two examples are given to illustrate the two categories.

In (11), representative of the first category, the student describes the series of pictures in Task 1 (cf. Task 1 in Appendix 4C).

(11) “so then they are at the beach and *they’re drinking and eating and having fun,* so” (WW30)

In (12), the student discusses why it is better to start a family when one is old than when one is young.

(12) “you can’t have children when you are so young because you don’t have a job *you are* maybe *studying* and you can’t do both things have children and study I don’t think it’s too good” (D34)

According to differences in the complexity of the structures, the instances of the progressive were also described in terms of the progressive simple and the progressive complex. The progressive simple includes structures with *am, are, is, was, were* and a present participle, whereas the progressive complex includes structures with two or more verbs and a present participle. The progressive simple is taught at an elementary level in Swedish schools, and contains only one verb followed by the present participle. It could therefore be expected to be easier to learn. The progressive complex is introduced much later in Swedish schools and since it contains two or more verbal elements followed by a present participle, it is expected to be
more difficult to acquire for the students. Complex progressive forms in the
data include structures with can be, has been, have been, will be and would be
followed by the present participle.

In co-ordinated uses, such as in (13), every present participle was
counted as a separate instance. Instances of self-correction were excluded
from the discussion.

(13) “they're drinking and eating and having fun” (WW30)

Thus, in Chapter 10, the students’ use of the progressive is discussed and
examined against the test rating categories High, Average and Low. In line
with the presentation of the third person singular marker, the discussion of
the students’ use of the progressive comprises the distribution across the
three interviewer categories and the tasks of the tests. This is followed by a
discussion of the relationship to the same two background variables as with
the third person singular marker (the students’ sex and their school grades
in English).

9.2. Vocabulary

The students’ vocabulary is important for the impression of their oral profi-
ciency (cf. 1.3. Linguistic Analyses of Students’ Proficiency). This was also
discussed by the assessors in the Gothenburg project, particularly by the
Native Speakers. In the answers to the questions of the questionnaire, the
three Native Speakers gave the following answers to the question “What
aspects of language or linguistic variables were important in your assess-
ment?” (Question 5 in the questionnaire, cf. Appendix 6):

“Communication is most important. Willingness to communicate. Ability to deal
with vocabulary lacunae.” (Native Speaker 1)

“Accuracy + a wide range of vocabulary, stress and intonation, an ability to converse
with another person . . .” (Native Speaker 2)

“I think the vocabulary aspect is the one that most of all leads me to a high mark. So,
I think that first of all vocabulary and idiom would be one of my five aspects.”
(Native Speaker 3).

The students’ vocabulary was also discussed by two Native Speakers when
asked what could be particularly disturbing for the assessment. (Question 6
in the questionnaire, cf. Appendix 6):

“Grammatical inaccuracy, poor vocabulary. I found it difficult to assess pupils who
were short on ideas anyway.” (Native Speaker 2)

“It’s the reverse of what I’ve just been saying, the lack of range of vocabulary, the
misuse of certain words like mountain when cliff was the appropriate word. That
obviously leads to a lower mark.” (Native Speaker 3)
In the answers to questions 5 and 6 of the Questionnaire, two (of the three) School Teachers discussed vocabulary as a significant factor for the assessment. Two (of the three) University Teachers mentioned the erroneous use of words such as nature, instead of countryside, and sea instead of lake as being particularly disturbing. These comments made it evident that the students’ lexical errors were relevant to include in the survey of the students’ vocabulary.

The students’ vocabulary is investigated regarding their use of abstract nouns, modifiers to adjective and compensatory strategies in search for a word or phrase. Furthermore, the occurrence of lexical errors concerning twelve frequent words across the three test rating categories is described.

9.2.1. Abstract Nouns

The content in the student production and their capacity for abstract thinking may be factors that contribute to the grades assigned to the student production. Students who discuss more complicated matters, who argue abstract issues and thus tend to use many abstract nouns, may be assessed more favourably than students who refrain from doing so. The relationship between grades and the frequency of the types and tokens of abstract nouns in the student production may therefore be worth investigating.

My personal experience from assessing Swedish students’ written and spoken production of English at school also seems to indicate that students who have reached an advanced level of language proficiency tend to use abstract nouns in their essays and discussions. The frequency of abstract nouns seems to be evidence of the students’ mastery of discussing controversial issues and their having attained language proficiency. Spontaneous discussions with language teachers have given further support to this idea.

For illustration three extracts with examples of abstract nouns collected from the data are provided below.

(14) “No it’s I don’t think it’s a difference at that age it’s everyone sort like says it’s okay and that there is no problem later on when boys tend to continue to do that and girls sort like have to they don’t get the same attention.” (M63)

(15) “Because eh when you are old you have experience you know how it’s to be to how children are maybe in the world you have friends who has children or ...” (A31)

(16) “Yeah but we have done this if we hadn’t have a eh so much pollution and so on so many people wouldn’t have cancer if we weren’t smoking and it’s our fault.” (S4)
Among the abstract nouns identified in the student performance, uncountable nouns emerged. These nouns are extensively discussed in school grammars, which distinguish between countable nouns, uncountable nouns and nouns with dual class membership (Hargevik and Hargevik 1995; Ljung and Ohlander 1995; Svartvik and Sager 1995). It may be expected that the use of uncountable nouns may cause problems for the student, in particular when there is a countable equivalent in the student’s first language Swedish.

The distribution of standard English and non-standard English instances of abstract uncountable nouns in the student production of the three test rating categories was consequently of relevance to include in the presentation.

The abstract nouns identified in the student production were thus categorised as countable or uncountable nouns. Nouns of dual class membership were included in the first category. In what follows, six nouns from the student production are given to illustrate the two categories:

Countable nouns / Nouns of dual class membership: problem, result, solution
Uncountable nouns: advice, knowledge, progress.

Abstract nouns are thus included in the study for two main reasons. Firstly, their presence in the student production may indicate mature thinking and so influence the Global grades assigned positively; secondly, abstract nouns that are uncountable may lead to the use of non-standard English forms and perhaps influence the global grades assigned.

The term “abstract nouns” includes nouns that refer to concepts that are “typically non-observable and non-measurable” (Quirk et al. 1985: 247). The nouns included in the present study are nouns used in contexts where they have abstract meanings according to this definition. It is of course impossible to give a representative list of the abstract nouns used in the students’ production, and so the list presented is to some extent subjective. It is worth pointing out that in the selection of the types of nouns included, the decisive factor is the context and that the students give the noun an abstract meaning. The interpretation of each single instance is thus crucial. Some types of nouns occur both with abstract and concrete meanings. The following example illustrates an abstract meaning of a noun, key, also used with concrete meanings:

(17) “The story ends pretty sad and that’s you can say yeah you can say it’s a key to his book 1984.” (T5)

When the abstract noun is repeated by the student, by way of correction or clarification, the two are counted as one instance.

The classification was checked in Collins Cobuild English dictionary (1995) and Oxford advanced learner’s dictionary (1995).
9.2.2. Modifiers of Adjectives

Modifiers of adjectives are words used in contexts such as in

(18) “It’s a very good book I think.” (T5)
(19) “Men are quite scared.” (W23)

They are considered to be characteristic of the spoken language or an informal style (Crystal and Davy 1969: 110–113; Leech and Svartvik 1975: 100–101). In the present investigation modifiers of adjectives, for instance quite and very, stood out as a category that seemed to be used differently by the students in their oral production. At a first survey these modifiers seemed to be very frequent in certain students’ performance. Some students’ tendency to use many instances of modifiers in their speech provided grounds to investigate whether modifiers were only used to intensify or downgrade adjectives or also as one of the linguistic items that students could use as devices that contributed to an impression of fluency and influence the Global grades assigned. Therefore it seemed reasonable to investigate whether the frequency of modifiers of adjectives was a factor that could be related to the Global grades assigned to their oral production. Furthermore, previous research had shown that adverbs as modifiers, as one item in one subcategory of the “reinforcers of register”, were important for the assessment of fluency (cf. 1.3. Linguistic Analyses of Students’ Proficiency). For these reasons, modifiers of adjectives were selected as one of the subcategories in the present investigation of the students’ use of vocabulary.

In the presentation of the results (Chapter 11), the occurrence of modifiers of adjectives in the students’ production is described, both in terms of overall frequencies and in terms of the types of modifiers used. The scope of the discussion is limited to modifiers of adjectives. Only adverbs as pre-modifiers of adjectives were thus selected. As for modifiers of adverbs, very few instances emerged in the material, and they were not included in the investigation. Both the modifiers which were used as intensifiers, e.g. very, and as downgraders, e.g. a little were taken into consideration. When the modifier is repeated, as in (20), the modifier is counted as if used once.

(20) “really really good” (G37)

Examples (21) – (23) are given to illustrate the students’ use of modifiers in the production. These examples are from production with high frequencies. Three illustrative extracts are given.

(21) “... when you’ve read all the book and you feel well it’s rather sad because you have to leave this book and that’s a bit sad, it’s very impressive ...” (Y10)
“I’m thinking I come to think about my grandmother she is very funny in a way because when grandfather lived she was very boring and a little bit angry and I thought she didn’t like me ...” (L42)

“It’s a eh every pist have a sort of colour if it’s how dangerous eh difficult it and the green it’s very difficu ... easy eh it’s eh for children and blue it’s quite easy and red is eh quite difficult and black is very difficult.” (W23)

9.2.3. Students’ Lexical Errors and Compensatory Strategies

In order to be successful in their performance students obviously need to have a certain range of English vocabulary. In this section the focus is on the occurrence of lexical errors and the strategies the students use when they cannot find an English word to express a specific meaning. The questions raised are whether occurrences of lexical errors and compensatory strategies tally with the Global grades assigned and whether there are different strategies that characterise student production related to the three test rating categories High, Average or Low. The students are not given much time to plan how to cope with a situation where they have trouble finding appropriate words. In this respect the students in the oral interviews of the Gothenburg project are close to authentic speech situations (cf. the discussion of the complexity of spoken production in 2.2. Oral Proficiency).

The investigation of the students’ lexical errors includes a description of the inappropriate use of certain words. The selection of the words was based on one or more of the following three criteria. Firstly, the students were corrected by the interviewers when they used these words in the interviews or the interviewers commented on the erroneous use of these words, either on the Oral Mark Sheet or in the answers to the questions of the questionnaire.7 Secondly, the words are described as “false friends” or occur in the descriptions of “important groups of synonyms” in Rende and Petti (1990). Thirdly, based on my personal experience of teaching English in Sweden, the words were recognised as common English words where a Swedish student often goes wrong, partly because of interference from Swedish. The words investigated were:

(1) the erroneous use of alley instead of avenue,
(2) mark instead of brand or make
(3) lonely or lonesome instead of alone,
(4) little or small instead of young
(5) mountain instead of cliff, hill or rock, and
(6) nature instead of countryside, country, landscape or scenery.

7 Cf. the extract from an oral proficiency interview which provides an illustrative example of an interviewer who corrected a student who used mountain instead of cliff (7.1. The Material of the Gothenburg Project).
Furthermore, inappropriate uses of one word in the following word pairs were identified:

(7) can or know,
(8) learn or teach,
(9) live or stay
(10) make or do
(11) read or study, and
(12) unlock or lock up.

The interviewers commented on the students’ uses of strategies in the answers to the questions in the questionnaires used. The question “What five aspects do you think are significant for the grade?” yielded comments on the students’ strategies and ability to cope with vocabulary gaps:

“I look favourably at the student employing strategies such as asking for help. I don’t know what that’s called but I mean this using circumlocution.” (Native Speaker 3 in an answer to the questionnaire).

Native Speaker 1 writes about the importance of the students’ ability to deal with vocabulary gaps and the question “What did you notice in particular that could lead to a low grade?” led to the following answer given by a School Teacher:

“Words pronounced so that they weren’t understood and literal translations from Swedish.” (School Teacher 3 in an answer to the questionnaire).

Literal translations could be categorised as a compensatory strategy. The comments above by the assessors and previous research on student production (cf. 1.3. Linguistic Analyses of Students’ Proficiency) made it seem relevant to include a survey of the students’ uses of compensatory strategies in the present study.

The present survey of the students’ uses of compensatory strategies is in line with the approach used in the Nijmegen project, a study of strategies employed by Dutch students of English (Poulisse 1989: 12). The Nijmegen project used an approach where the focus of interest was on some types of strategies, deliberately setting aside other types.8

In the present study, the strategies are classified on the basis of the survey of compensatory strategies in Chapter 1 (cf. 1.3. Linguistic Analyses of Students’ Proficiency) Twelve types of strategies in two categories were distinguished: Strategies involving Swedish and Strategies in English. The

8 Reduction and message adjustment strategies seemed problematic to deal with within the scope of the present study. Other methods than the ones used would have been necessary, such as listening to recorded tests together with the students and/or interviewers immediately after the completion of the tests.
boundary between these categories is fuzzy and the types classified should be seen as points along a continuum. The twelve compensatory strategies identified in the material were as follows:

**Strategies involving Swedish**
1. Request for help in Swedish
2. Swedish words used
3. Swedish words pronounced in English
4. Swedish words explained in English
5. Swedish words added
6. Word coinage based on Swedish

**Strategies in English**
7. Request for help in English
8. Waiting for assistance
9. Substitution by superordinate or approximate terms
10. Word coinage based on English
11. Giving two alternatives
12. Paraphrase (Comparison, Description and Reformulation)

In what follows these twelve compensatory strategies are defined and illustrated with examples from the material.

1. Request for help in Swedish
The students asked in Swedish for the translation of a word.

(24) “Lantställe, vad heter det?” (E80) (="Summer cottage, what is it called?)
(25) “He cleans and … vad heter ‘stryka’?” (G68) (="How do you say ‘to iron’?
(26) “I don’t think they are medvetna … hur säger man det? (G82) (="aware of …how do you say it?

2. Swedish words used
The students used Swedish words and pronounced them in Swedish.

(27) “you have to be uppmärksam." (E80) (="attentive)
(28) “leaving me in the dagis” (H38) (="day-care centre)
(29) “you always see a three on the betyg” (G82) (="school report)

3. Swedish words pronounced in English
The students pronounced Swedish words in an English way, as if these words existed in the English language.
4. Swedish words explained in English
The students first talked about the concept with the help of the Swedish word and then explained the concept in English or translated it into English. It seems that the students were planning out loud and searching for the appropriate English word. In some instances, the English word is pronounced with a rising intonation and could be interpreted as a question (cf. the discussion of intonation in 1.3. Linguistic Analysis of Students’ Proficiency). In the following illustrative example, the interviewer shows his approval of the student’s use of herring.

(33) Interviewer: “What do you eat?”
Student: “Sill. Herring.”
Interviewer: “Yes, that’s herring yeah.” (F81)

5. Swedish words added
After the students had said the appropriate English word, they added the Swedish word. The translation may have been added because they wanted to be as precise as possible.

(34) “it’s normal for a summer house maybe I don’t know I don’t know we call it torp” (F36) (= summer cottage)

(35) “it’s a small animal it’s a water animal iller” (B77) (=polecat)

6. Word coinage based on Swedish
The students made up English words, often by means of literally translating Swedish expressions and words.

(36) “it’s a happy field I don’t know what to say merry-go-rounds and” (ZZ29) (=funfair)

(37) “if there isn’t a back door maybe fire door or something you can he could find and get out that way” (O90) (=emergency exit)

7. Request for help in English
The students asked for help in English about the appropriate English word. In the following example, the student asked about the pronunciation of a word.

(38) “industr …. indu what do you say?” (ZZ47)

154
8. Waiting for assistance
The students made a pause and thought about how to go on. In this pause, the interviewers started to help the students by way of suggesting words. Evidently, this strategy is largely dependent on the interaction and the interviewer’s initiative (cf. the discussion concerning the interviewer’s method in 1.2. Oral Proficiency Testing).

(39) Student: “Swedish women have lower …yeah.”
   Interviewer: “Pay?”
   Student: “Yeah, pay.” (F75)

(39) Student: “She’s more …”
   Interviewer: “Stricter?”
   Student: “Yes, stricter.” (H38)

9. Substitution by superordinate or approximate terms
The students used general terms or concepts close in meaning to the one searched for. It is obvious that the students at first seemed to search for the appropriate word but then chose to use an approximate or superordinate word. Or-tags, such as or what do you call it and or what do you say, tended to emerge in these instances.

(41) “he takes him as his *pupil* or whatever you can say” (U6) (=apprentice)

(42) “we make a *ring*” (E80) (= a wreath)

(43) “we eat *raw fish*” (S4) (= pickled herring)

10. Word coinage based on English
The students invented new, and non-existent, words with the help of their knowledge of English.

(44) “everybody says that Swedes are so *equalized*” (F75) (= equal)

(45) “but I think that this *youthness* that we have now it’s important too to I mean to pay with our feelings to find ourselves of course” (Q59) (= youth)

11. Giving two alternatives
In the answer to one of the questions in the questionnaire, one of the interviewers discusses the significance of the students’ feeling confident in their choice of words and structures. Students that give two alternatives to concepts they wish to describe might signal uncertainty and lack of self-confidence (for discussions of the significance of self-confidence and anxiety, and the language learner, see Tarone and Yule 1989: 133). The strategy called “Giving two alternatives” is regarded as a signal of the students’
difficulty in defining a phenomenon precisely; the students added a second word to be sure that the message was understood.

(46) “he discovered he can’t come out the movie or the cinema” (E80)

(47) “they went to have a picnic by the sea or by the ocean” (O61)

(48) “very near to the shop or the supermarket or whatever you call it” (D34)

12. Paraphrase
The students paraphrased in order to get round the lack of vocabulary. Three ways of paraphrasing were identified: a) comparison, b) description and c) reformulation.

Comparison
The students compared characteristics similar to or different from the concept. In the first illustrative example, the student pointed at similarities, in the second at differences.

(49) “it’s like a tree, and you dance around it” (K86)(= Midsummer pole)

(50) “the toilet is not an ordinary one” (D34) (= outdoor lavatory)

Description
The students gave a careful physical description with regard to size, shape, colour, material, features, function or location.

(51) “I don’t know what you call it it’s a pole with flowers and leaves and it looks like a cross with two circles under each arm” (YY13)(= maypole)

(52) “I don’t know what they are called, cottages which are white and then they have tree in …” (WW46) (= Tudor style)

Reformulation
The students started off but then went on to a paraphrase and reformulated the utterance to avoid the lack of vocabulary.

(53) “where they have different … you know they are different people who show their art” (O61) (=exhibitions)

(54) “I tie a … I don’t know the word in English but I tie them together so I have a ring to put on the head” (K86) (= wreathe)
9.3. Discourse Markers

As described in Chapter 1, (cf. 1.3. Linguistic Analyses of Students’ Proficiency), discourse markers, such as well and you know, are used in a wide range of contexts and have several discoursal functions in communication. The students’ use of discourse markers, meta-talk and the ability to fill pauses in speech with words could be expected to contribute to the impression of their proficiency in spoken English and the assessment in oral tests. In the present study, the students’ use of two selected discourse markers is analysed.

When the transcription of the student production was checked it was evident that the two discourse markers well and you know occurred frequently in certain students’ production. Other discourse phenomena such as question-tags were very rare. This is one of the reasons why the focus is on the students’ use of well and you know in the present study.

Another reason for the interest in the two discourse markers was that a great variation in the frequency of well in the students’ production could be observed. A number of students produced 20 instances, whereas other students did not use it at all. For that reason, it seemed interesting to study the relationship between the occurrences of well and the Global grades assigned. On the basis of the theory of possible new meanings and interpretations of an item used in a wide range of contexts (cf. 1.3. Linguistic Analyses of Students’ Proficiency) and the observed high frequencies of well in certain students’ spoken English, it was relevant to take into account what seemed to be the discoursal functions of the students’ uses of well and compare them with the grades assigned. Furthermore, on the basis of the findings by Riggenbach (1991) (cf. 1.3. Linguistic Analyses of Students’ Proficiency), it might be expected that well was one of the lexical fillers that characterise students’ speech assigned a High Global grade.

Based on my experiences of language teaching in Swedish schools, it seems that the appropriate use of discourse markers, such as well and you know, is a feature of the English language that was hardly ever explicitly taught at the time of the collection of the material. The students were left to pick up these features without the aid of specific instruction or teaching. This is nothing new and was pointed out many years ago by Catford (1959: 149):

I know of no practical English grammar for foreign learners which describes the use of oh, ah or the introductory or resumptive well. And yet the occurrences of these words, like the occurrences of all linguistic items are not random, they are systematic and presumably describable.

9 English grammars for advanced students at university level contain descriptions of discourse markers, such as well and you know, see for instance Johansson and Lysvåg (1986: 229).
One of the reasons for the non-existence of the teaching of discourse markers in most classroom situations may be that it is impossible to describe, for instance, the uses of *well* in traditional grammatical terms. Inadequate uses of verbs can be corrected quite easily whereas the appropriate or inappropriate uses of *well* may be more difficult to pin down. Yet, as Svartvik states (1980: 169):

... an inappropriate use of particles like *well* may have more unfortunate consequences for communication success than elementary grammatical errors.

Discourse markers such as *well* are also very hard to translate idiomatically into another language (Svartvik 1980: 169). Even though Swedish is close to English, the exact translations of *well* in the instances attested in the material are not evident. Discourse markers are thus used for certain pragmatic or contextual reasons and possibly appropriate use of *well* could be an indication of the students’ ability to think in English or a sign that they have adopted an English style which may influence the assessment.

Discourse markers can be defined as sequentially dependent elements which bracket units of talk (Schiffrin 1987: 31). In the present study, the use of two discourse markers according to this definition is discussed: *well* and *you know*.

The purpose of the present study is to give qualitative descriptions besides the quantitative analysis. The qualitative analysis, with the discoursal functions attested in the student production, can provide evidence of instances of *well* and *you know* that serve the functions that are described in literature as used by native speakers. This finding would confirm the relevance of frequency counts of *well* and *you know* in a study of assessment of oral proficiency.

9.3.1. *Well*...

In this study, the focus is on a specific use of *well*, namely its use as a discourse marker as defined by Schiffrin (1987). Schiffrin suggests two specific conditions that allow expressions to be used as discourse markers. Discourse markers have to be syntactically detachable from the sentence, and they have to be commonly used in initial positions of the utterance. Both these criteria are applicable to the instances of *well* included in the investigation.

The first two instances from extract (55) illustrate a student’s use of *well* as an initiator and as a marker of reformulation. The extract is from a discussion of school uniforms in Task 3.

(55) Interviewer: Would it be better to have a uniform?
Student: *Well* I don’t think no not in Sweden I think it’s I shouldn’t work, *well* if you yeah maybe it should be better I think ... (E74)

Intonation makes the second instance of *well* a discourse marker and not an adverb.

158
The third illustrative instance of *well* is a student’s use of *well* in a shift from description to indirect speech. Example (56) is from a description of two boys on a beach. The student uses *well* to emphasise that now she turns to a description of what the two boys are thinking.

(56) “and they got to the beach and they sat down they ate had a lot of fun and suddenly they saw this toy airplane aeroplane on a shelf on the on a rock rock and they thought *well* why don’t we go and get it.” (M63)

When the students repeated *well* several times in the same context, the instances were counted as one instance, as in example (57):

(57) Interviewer: Which one did you choose?
    Student: *Well, well, well* I must think a little bit.

Five categories of discoursal functions of *well* were used to classify the instances. These functions were found in the material and have been described in previous studies. The five categories are

(1) *well* in answers to the interviewers’ questions and requests,
(2) *well* in reported speech,
(3) *well* in repetitions,
(4) *well* in contexts where students express difficulties with English, and
(5) *well* as a finishing signal.

The instances of *well* that could not be categorised in the discoursal functions 1–5 are included in a sixth category. In this last category, no evident explanation could be given to the students’ use of *well*. The reasons for the occurrence of *well* may be, for instance, hesitation, use as a pause-filling device or simply an effort to make the utterance “sound English”. The criteria for setting up discoursal functions are based on structural clues in each instance. The instances of the category of indeterminate uses do not show structural similarities that are possible to categorise and these instances are labelled

(6) *well* in indeterminate contexts.

The six categories are commented on and exemplified below.

(1) *Well* in answers to the interviewers’ questions and requests

Brown (1977: 121) classifies *well* as a common introductory filler when the speakers are asked a question and they are expected to produce an immediate reply. Ball (1986: 117) describes *well* as a universal introductory filler in answers to almost any question; this filler is followed by a pause. *Well*
indicates that the speaker is about to produce a reply. The following two examples illustrate the students’ uses of *well* in answers to the interviewers’ questions and requests:

(58) Interviewer: Would it bother you if your girlfriend was ...?  
Student: *Well* that would bother me I think because I think ... (Y25)

(59) Interviewer: Now that’s gonna cost: Would that be a benefit for the town for example?  
Student: *Well* it’s always good to have a big stadium. You can have rock concerts and so on there. (Z26)

(2) *Well* in reported speech

*Well* is used when the speakers introduce speech with verbs such as *say* and *think*. *Well* is described as a member of the group of words which occurs as the first item in direct quotations (Schourup 1985: 22). Svartvik (1980: 175) compares quotation marks in written English with *well* in spoken English. They both signal the beginning of direct speech. After the verb *think*, *well* indicates that the speaker turns to what somebody else thinks. This is often in the form of a question, as if the speaker were speaking to her/himself. The following two examples illustrate the second discoursal function of *well*:

(60) “and he said *well* I deserve to go to the cinema” (Z11)

(61) “and they thought *well* why don’t we go and get it” (M63)

(3) *Well* in repetitions

Instances of *well* in repetitions were identified at an early stage of the analysis. It is evident that in oral production, students may change their ways of expressing what they wish to say and choose alternative words or structures. In doing so they may repeat what they have previously said. The structural criteria of this function are the students’ uses of two similar words or expressions before and after *well*. In example (61), the student decided to reformulate her description of her father and repeated *he is*:

(62) “but not with my father *he is well* maybe when it’s raining and he is out running he have to take his medicine but …” (B77)

(4) *Well* in contexts where students express difficulties with English

The students also used *well* as a marker of hesitation and reformulation. One evident reason for their pauses in speech is that they did not know the appropriate English words. In the instances of this category, the students expressed their difficulties with or their doubt about the correctness of the words selected. They did so by adding phrases such as *I don’t know what*
it’s called or what do I say? before or after well. These phrases are the structural criterion applied to distinguish this fourth function. In the following example, the student showed his doubt about the appropriate use of the word strength.

(63) “to have the strength well what do you say” (ZZ47)

(5) Well as a finishing signal
In its conclusive function, well indicates that the speaker is about to finish. Well is one item along with others, such as OK, so and all right, signalling that the speaker is willing to bring the conversation to an end before the final “Goodbye” (Wardhaugh 1985: 158; Ball 1986: 118). In the Gothenburg material, well precedes concluding comments or final remarks about a topic in a part of the test or in the entire test. Well indicates that the students have completed a task or the whole test by the use of well. Example (64) is from the end of Task 1 in a test.

(64) “well he falls asleep in the cinema” (YY13)

In example (65) the student concluded his description of the houses in Task 1 with an utterance initiated by well:

(65) “it’s also a small small house where you can sleep so there’s enough room for four people cause it’s quite small the house and it cost about 1600 Swedish crowns well that’s what I can tell you” (YY48)

(6) Well in indeterminate contexts
Instances that cannot be classified into the five categories above seem to be used in a number of different contexts. Among these contexts, there were two contexts that could be identified. They are well as a hesitation marker or as a pause-filling device, and in reformulations or shifts of speaker orientation. There were three reasons for including the instances in the sixth category. Firstly, it was not possible to identify the definite reason for the students’ use of well. Secondly, the borders between the discoursal functions of the instances attested were not clear-cut. Thirdly, there were no structural clues that could be used in the classification of the instances.

Nevertheless, the two possible discoursal functions of well distinguished in certain instances in this category may be discussed. Firstly, well seems to function as a pause-filling device when the students pause to consider how to continue a sentence. Well is then a device serving to give speakers time to think (Schourup 1985: 79; Wardhaugh 1985: 142). According to Wardhaugh (1985: 149) pauses generally tend to get filled with items such as well by the speaker or by other participants in conversations. Aijmer and
Olsson (1990: 77) discuss the use of well as a word used when the speaker hesitates or searches for a word, or when the speaker needs a kind of mental pause signalling: “let me consider it for a while”. Example (66) illustrates a student’s use of well when he seems to hesitate or to fill a pause.

(66) “and well we dance around this tree with lots of flowers and things like that sing sing songs and” (O90)

Secondly, in some other instances well seems to be used as a means of signalling a shift in speaker orientation, for instance from a description to an evaluation. In this function well served as a means of cohesion. Schiffrin (1985: 659) calls this kind of instances “reflexive frame breaks”. Among the instances identified, shifts occurred, for example, when students shifted from describing to evaluating or from telling a story to adding a personal opinion. The evaluation or the addition of the personal opinion was then introduced by well. Example (67) illustrates this use. In this example the student shifts from the discussion of a book title to the observation of the occasion when she read the book.

(67) Interviewer: Strange title, isn’t it?
Student: Pretty but it makes it more interesting well it was a long time ago I read it but it’s about a murderer and (J85)

In example (68) the student shifts from his evaluation of a trip to Paris to his account of it:

(68) “it was a very interesting experience because I loved the city and well we lived in old Paris and at a hotel and everything was okay” (Y51)

9.3.2. You know ...

The second discourse marker investigated, which is you know, is of interest because of factors, such as its significance for the interaction, its characteristic of giving signals of a more informal level of conversation, the described plea for co-operation adhered to it, and as has already been mentioned, its frequency in spoken English (cf. 1.3. Linguistic Analyses of Students’ Proficiency). Furthermore, you know differs from well as regards its functions. You know is used in other contexts and with different meanings. You know was also included in the present study because of its frequent occurrence in the material (other types of discourse markers of interest, such as you see, appeared in remarkably fewer instances).

Only the instances of you know that can be described as discourse markers according to the criteria by Schiffrin (1987) presented above (cf. 9.3.1.) are included in this study. This means that only those instances of you know
which can be left out of the structure without changing the meaning of the utterance to any noticeable extent, are considered. The two examples (69) and (70) are given to illustrate the students’ use of you know:

(69) “so he was locked inside the cinema ... you know screaming for help” (B77)

(70) “well I drive a lot but you know I think you have to have had your driving license for ...” (E80)

In the oral tests, the interviewers could be expected to take some initiatives towards less formal conversation. For that reason, the interviewers’ uses of you know in the interviews that also contain instances of students’ use are taken into account. It is of interest to see whether the interviewers started using you know before the students did so.

Three additional aspects of you know are of interest to consider in the discussion on the basis of results from previous studies (cf. 1.3. Linguistic Analyses of Students’ Proficiency). Firstly, the finding that you know is used more by women than by men (Östman 1981: 72; Coates 1993: 116–118) suggests that it is relevant to include a discussion of the proportions of the instances in boys’ and girls’ production in the material. Secondly, the finding that you know is particularly used in narrative parts of conversations makes it interesting to investigate the distribution of the instances across the three tasks (Östman 1981:16). Thirdly, the fact that you know seems to occur to a great extent in medial position of the utterances (Erman 1987: 50) is worth taking into account in the investigation of the instances attested.

Three discoursal functions of you know were identified in the material. Instances that could not be classified in one of the categories described above were included among the indeterminate instances.

1. Expressing an opinion
You know is used together with statements of the students’ opinions and together with I think, I don’t think, I thought, I like and I feel like. It seems that the students use you know as a softening item by way of making the point-of-view stated in a more indirect and less controversial way. The students seem to have tried to reach mutual understanding as to their standpoint. In (71) the student argues for starting a family late in life:

(71) “because I think you must have a good job and you know you have to live a little by your own first” (L42)

2. Word searching
On some occasions the students search for words or use words which they do not know whether they are appropriate or not. In quite a few instances,
you know is combined with phrases such as I don’t know the word or what do you say. In these reformulations the students seem to use you know to provide the extra time they need or to hedge words that they suspect that they have used incorrectly. In this discoursal function, instances of you know are also combined with word-coinage or Swedish words (cf. 9.2.3. Students’ Lexical Errors and Compensatory Strategies). In (72) the student is unsure of the correct use of language course:

(72) “there is a I don’t know what the word is but you know this language course and I was ...” (XX49)

3. Repetition
In instances placed in this category the students seem to use you know to provide extra time in their planning of what they were going to say. There is no topic-shift but they go on discussing the same topic. The indication of this continuation is the repetition of at least two words after you know. This is the criterion of the instances assigned this discoursal function. In (73) the student explains why there is no predetermined age for starting a family and uses mature enough both before and after you know.

(73) “I think you never be mature enough if you think in that way you know when you are thirty you probably say oh no I’m not mature enough” (R58)

4. Indeterminate instances
This category includes instances of you know when it is not used for obvious reason. The reasons could be for instance, vocabulary gaps or need for clarifications of what has been said but evidently there could be other reasons that made the students include you know in their speech. In the examples (74) and (75), there may be several reasons for the students’ use of you know, for instance as an indication of switching from deference to camaraderie (cf. 1.3. Linguistic Analyses of Students’ Proficiency). In (74) the student discusses the fact that many girls who are interested in horses choose other hobbies when they are 17. In (75) the student describes the fact that he moved some years ago.

(74) “yes you know in the age of 17 mostly girls stop riding” (G68)
(75) “you know I moved here two years three years ago so ...” (O61)
9.4. Aspects of Pronunciation

9.4.1. Intonation

The students’ intonation is of interest in the investigation mainly for three reasons. Firstly, intonation is mentioned four times in the rating criteria for Pronunciation (Factorial grade 2c) (cf. the rating criteria in Appendix 8). Secondly, the interviewers gave comments on the students’ intonation on the Oral Mark Sheet (cf. Appendix 10) and they also commented on the students’ intonation or the grading of pronunciation in the answers to the questions of the questionnaire (cf. Appendix 6). The following examples are given to illustrate these comments on intonation and pronunciation:

“intonation somewhat monotonous” (comment on B71 by Native Speaker 1)

“pronunciation, I find it very difficult to put on an A B C D E scale and the criteria were very difficult to apply” (comment in answers to questionnaire by Native Speaker 3)

Thirdly, at an early stage of the investigation of the students’ intonation, it was evident that there were instances of unidiomatic intonation in the students’ performance. This unidiomatic intonation was identified as rising intonation in declarative sentences. Because of these reasons, it could be assumed that the students’ intonation could influence the grades assigned.

Intonation is dependent on many contextual factors (cf. 1.3. Linguistic Analyses of Students’ Proficiency). However, at least four functions expressed by intonation may be distinguished. The speaker can single out the word that carries the most important information. This can be called the information function of intonation. A second function is to indicate the communication function of the utterance, i.e. whether the speaker wants the utterance to be interpreted as for instance a statement, a question, or a request. Thirdly, the grammatical function may be distinguished: intonation distinguishes grammatical categories such as restrictive and non-restrictive relative clauses (Johansson and Rönnerdal 1993: 157). Finally, intonation can give indications of the speaker’s attitude, either towards the message or towards the interlocutor, and this is what can be named the attitudinal function.

Because of the complexity of intonation, illustrated above, it should be emphasised that it is not within the scope of the present study to provide a complete description of the students’ intonation in their production. Only results regarding one particular aspect of intonation in a specific context and observed in a limited number of interviews will be presented. The tendencies observed may nevertheless provide useful information about the diversity and the quality of the students’ intonation patterns in relation to the assessments.
As with the other linguistic features discussed above, the findings as regards the students’ intonation are compared with the normative descriptions found in school grammars and are discussed in terms of standard English and non-standard English uses. Thus no full-scale description of the students’ intonation is provided. Only one basic feature of English intonation has been studied: the nuclear falling intonation in declarative sentences.

The method used in the present study was purely auditory. No spectrograms or other technical devices were used. The researcher himself together with two phoneticians listened to and evaluated the students’ intonation.

9.4.2. Stress

Stress and rhythm, along with intonation discussed above, may play important roles in the assessment of Swedish students’ spoken English. Stress is discussed, for instance by Axelsson (1994), in a handbook of pronunciation for teachers of English in Sweden. The prominence of stress is described as “the feature that, above all, characterizes the rhythm of English and makes it different from Swedish” (Axelsson 1994: 17). In the present study, two limited investigations of the students’ use of stress are included: one concerns the students’ stress of lexical and function words in a selected sample, the other the stress pattern of the attested 69 instances of and so on in the material. The general principle of unstressed function words as opposed to stressed lexical words applies in both English and Swedish, but the contrast is stronger in English. In a study on a Finnish-English contrastive basis, it is argued that the most important role of stress is “to indicate the constituent structure of the sentence by structuring the flow of speech into specific dynamic patterns” (Lehtonen et al. 1977: 68). The contrast between the stress of lexical words that convey new information and unstressed functional words is one means for the speaker to indicate the structure of the sentence. A study of the students’ stress of the phrase and so on is useful because of the difference in stress in Swedish and English: the corresponding Swedish expression is stressed on the last word, whereas and so on is stressed on the middle word. This fact is pointed out by Johansson and Rönnerdal (1993: 135) who describe the use of and so on in a subsection among other examples where a Swedish speaker often goes wrong. The decision to do a special study of the students’ stress in the phrase and so on was based on the assumption that stress may be a factor influencing the assessment. Of course, the correct or incorrect use of stress in one particular instance may not reflect the general level of a student’s use of stress. Therefore this investigation was complemented with a study of the stress of lexical and function words in one-minute samples of nine students.
9.4.3. Speech Rate

In the rating criteria of the Gothenburg project, the terms fluent and fluency emerge twice. Factorial rating 2a has the heading “Communication and Fluency” and the highest grade of this Factorial rating contains the description “Comfortable and completely fluent” (cf. Appendix 8). The two concepts fluent and fluency can be associated with time and it is natural to speak of the temporal element of fluency (Saleva 1997: 54). Furthermore, speed appears in the criteria of the grade B of Factorial rating 2a: “Extended contributions at natural speed and tempo”. The students’ speech rate in relation to the Global grades assigned is therefore of interest to investigate.

The investigation of the students’ speech rate is based on the method and findings in the study by Lennon (1990) (cf. 1.3. Linguistic Analyses of Students’ Proficiency). One of the three factors that Lennon found to be a variable of fluency, namely “pruned words per minute” was chosen for a closer scrutiny. The present analysis covers the first minute in Task 1. There are three reasons for this decision. Firstly, Task 1 was a storytelling task completed with the help of a series of pictures similar to the one used in Lennon’s study (cf. 7.4. Tasks). Secondly, Task 1 lasted for at least one minute in all the oral tests, and the students were rarely interrupted by the interviewers during this first minute. Thirdly, the students had a series of pictures to follow, and for that reason it could be expected that the student production is more comparable than when they discussed different topics, as they did in Task 2 and Task 3. In the samples, the introductory phase, i.e. before the students began to tell the story, was excluded. The narration was often prompted by the interviewers with phrases such as “Ok, tell me the story” or “Please start”. Furthermore, it should be mentioned that there was an additional reason for the decision to focus on the first minute of Task 1. It may be argued that assessors decide about the grade at a very early stage of an oral proficiency interview. For that reason, it was relevant to investigate the students’ speech rate in the very early stages of their production.

After this introductory survey of the linguistic features included in the present study, the following chapters will present the results from the investigation of these features in the students’ production.

---

11 Fillmore (1979: 93) distinguishes four types of native speaker fluency. One of these four types includes a temporal element: “the ability to fill time with talk”.
10.1. The Third Person Singular Marker

10.1.1. Overall Distribution

The material yielded 799 instances of the third person singular marker when all the standard English and non-standard English uses were counted. The distribution of the instances in student production across the three test rating categories High, Average and Low, and the distribution of normalised figures are given in Table 10.1.

The results show, as expected, that student production assigned a High grade has a higher proportion of standard English instances (75% of the total) and that student production assigned a Low or Average grade contains a higher proportion of non-standard English instances (31% and 43%, respectively, of the total). The normalised figures per 1,000 words show that non-standard English instances occur more often in student production assigned a Low grade (4.7) than in production assigned a High grade (2.2) and that standard English instances are frequent in production assigned an Average grade (8.0). Furthermore, the normalised figures

Table 10.1. *Uses of the third person singular marker by test rating categories*

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard English</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total instances (proportion of all)</td>
<td>799</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>6.7</td>
<td>75%</td>
<td>8.0</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>6.3</td>
<td>57%</td>
<td>7.3</td>
<td>70%</td>
</tr>
<tr>
<td><strong>Non-standard English</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total instances (proportion of all)</td>
<td>72</td>
<td>100%</td>
<td>19</td>
<td>100%</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>2.2</td>
<td>25%</td>
<td>3.6</td>
<td>31%</td>
</tr>
<tr>
<td></td>
<td>4.7</td>
<td>43%</td>
<td>3.2</td>
<td>30%</td>
</tr>
<tr>
<td>Total instances of both categories</td>
<td>287</td>
<td>100%</td>
<td>416</td>
<td>100%</td>
</tr>
<tr>
<td>Number of oral tests</td>
<td>32</td>
<td>100%</td>
<td>37</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>100%</td>
<td>82</td>
<td>100%</td>
</tr>
</tbody>
</table>
show that the standard English instances occur as often in student production rated Low (6.3) as in student production rated High (6.7). Obviously many standard English instances do not change the overall impression of a performance that contains many non-standard English instances. Regardless of grade, student production contains both standard English and non-standard English, but a higher proportion of non-standard English instances emerged in student production assigned an Average and, in particular, a Low grade.

10.1.2. The Occurrence of the Third Person Singular Marker in Individual Students’ Production

In a single test, non-standard English instances of the third person singular marker may appear either together with standard English instances or together with non-standard English instances of other types. This fact can possibly explain the interviewers’ tolerance of non-standard English instances in the assessment. This could be the case particularly for the 72 non-standard English instances in the student production assigned a High grade (cf. Table 10.1). The proportions of standard English and non-standard English instances in individual students’ production are provided in Table 10.2. In this presentation, all production is categorised in three groups depending on the proportions of standard English and non-standard English instances found. The three categories distinguished are:

1. The majority of the instances are standard English.
2. There are an equal number of standard English and non-standard English instances.
3. The majority of the instances are non-standard English.

According to the results, the majority of the student production (54 out of the total 82 oral tests, 66%) contains more standard English than non-standard English instances. Relatively more oral tests of individual students’ production assigned a Low grade contain a majority of non-standard Eng-

| Table 10.2. Proportions of the third person singular marker in individual students’ production |
|-----------------------------------------------|----------------|----------------|----------------|----------------|
|                                               | High           | Average        | Low            | Total          |
| 1. Majority standard English                  | 22 (68%)       | 25 (67%)       | 7 (54%)        | 54 (66%)       |
| 2. An equal number                            | 5 (16%)        | 4 (11%)        | 2 (16%)        | 11 (13%)       |
| 3. Majority non-standard English              | 5 (16%)        | 8 (22%)        | 4 (30%)        | 17 (21%)       |
| Total oral tests                              | 32 (100%)      | 37 (100%)      | 13 (100%)      | 82 (100%)      |
lish instances (30%) than do oral tests of individual students’ production assigned a High grade (16%). As could be supposed, the figures give evidence that most individual students’ production contains a majority of standard English instances, and moreover, that a higher proportion of individual students’ production assigned an Average (22%) or Low (30%) grade than production assigned a High grade (16%) contains a majority of non-standard English instances.

10.1.3. Linguistic Factors

Types of Non-standard English Instances of the Third Person Singular Marker

The non-standard English instances of the third person singular marker are of three types (see 9.1.1. The Third Person Singular Marker). It is of interest to see how the instances of the three types are distributed across the three test rating categories High, Average and Low.

Type 1 (*he take) is generally the most common type of the three non-standard English types in the material. Moreover, Type 1 is by far the most common type for the student production assigned a Low grade (80% as opposed to 49% and 48%) whereas production of other test rating categories tends to have instances of Type 2 or 3, too. Type 2 is almost as frequent as Type 1 in student production assigned a High or Average grade.

The Type of the Subject

The type of subject used in the non-standard English instances of the third person singular marker might account for the distribution across the test rating categories. In Table 10.4., the 242 non-standard English instances are categorised in four categories depending on the type of the subject: pronoun, noun, co-ordinated use and others (the last category includes omissions of subjects and instances in relative clauses). The following examples of non-standard English Type 1 are given to illustrate the four categories:

<table>
<thead>
<tr>
<th>Table 10.3. Proportions of non-standard English types of the third person singular marker by test rating categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Average Low Total</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Type 1 (*he take)</td>
</tr>
<tr>
<td>Type 2 (*they takes)</td>
</tr>
<tr>
<td>Type 3 (other types)</td>
</tr>
<tr>
<td>Total instances</td>
</tr>
<tr>
<td>Total oral tests</td>
</tr>
</tbody>
</table>
The most common type of subject with the categories of the non-standard English instances is a pronoun. This is probably the case for standard English uses too. In Type 2, differences between types of subjects were less marked than in Type 1. The types of subject with non-standard English instances and the three test rating categories show no clear correlation. Thus no great differences are found in the proportions of the types of subject across the three test rating categories.

**Verbs Ending in Sibilants**

As is shown in Table 10.3., the omission of the third person singular marker is the most common non-standard English type. Next, the verbs used in the instances of omission of the third person singular marker are examined, to see whether the final phonemes of these verbs may, at least partly, account
for the occurrence of non-standard English instances. The verbs ending in voiced or voiceless sibilants require the ending [iz] and it is probable that students need to make an extra effort in adding this ending. The sibilants of these verbs might have a knock-out effect on the use of the third person singular marker by some students, who would adopt the principle of economical language, production with the least possible effort (cf. 1.3. Linguistic Analyses of Students’ Proficiency).

The material yielded 17 instances with 13 types of verbs ending in sibilants: catch, change, close, dance, finish, fix, manage, notice, pass, reach, realise, use and wish.

The figures are definitely very low but nevertheless interesting. The proportion of non-standard English instances of verbs ending in sibilants is higher (53%) than the proportion of non-standard English instances of all the verbs in the material (30%, cf. Table 10.1.). The standard English and non-standard English instances appear particularly in student production assigned an Average grade. No clear tendency can be observed across the three test rating categories.

10.1.4. Occurrences of the Third Person Singular Marker in Relation to Interviewer Categories, Tasks and Background Variables

In this subsection, the role possibly played by the background of the interviewers is first discussed. Then the distribution of the instances in relation to the tasks of the students’ production is elaborated on, and finally the focus is on two background variables: sex and the students’ school grades.

The proportions and normalised figures of the third person singular marker in comparison with test rating categories and by interviewer categories can be seen in Table 10.6.

High frequencies of the third person singular marker are found in student production assessed by School Teachers (160, 255 and 61). This finding is

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard English instances</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>8 (47%)</td>
</tr>
<tr>
<td>Non-standard English instances</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>9 (53%)</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>17 (100%)</td>
</tr>
</tbody>
</table>
further discussed below. Moreover, there is a higher proportion of standard English than non-standard English instances in all student production irrespective of interviewer category or test rating category. One interesting difference should be pointed out, however. For two interviewer categories, School Teachers and University Teachers, the normalised figures show that the frequency of non-standard English instances is higher in student production assigned an Average or Low grade (ST: 4.9 and 6.4; UT: 2.3 and 3.3) when compared with student production assigned a High grade (ST: 2.5; UT: 1.3). No similar tendency is found in student production with Native Speakers.

As mentioned above, there are more instances of the third person singular marker in student production with School Teachers than in production with the other two interviewer categories. The topics of the tasks may, at least partly, account for the differences in the recorded frequencies. The topic of Task 1 in the interviews with School Teachers was a storytelling task about a single person whereas the topics of Task 1 in the other interviews were about two houses and about two boys (cf. Appendices 4A, 4B and 4C). The distribution of the 799 instances in the three tasks of the student production can be seen in Table 10.7.

The third person singular marker was more common in student production assessed by School Teachers, and Task 1 in this production contains a large proportion of these instances. The topic, to tell the story about Mr Brown, seems to account for the differences in frequencies. The topic of Task 1 with School Teachers provided a context for the students to use the
The types of third person singular marker described above more often than the topics of Task 1 in the tests with the other two interviewer categories.

Two background variables are readily accessible, namely the students’ sex and their grades in English in school. The investigation showed that there is no relation between the students’ sex and the use of the third person singular marker: the distribution of standard English and non-standard English instances in the 12 boys’ production and in the 17 girls’ production shows no major differences.

The distribution of the standard English and non-standard English instances in comparison with the students’ grades in school is presented in Table 10.8. (grades 1 to 5, 5 being the top grade). The low number of students with the grade 2 should be taken into account when the results are interpreted.

Production by students who had high grades in English in school contains a high proportion of standard English instances. Not surprisingly, the higher the student’s grade is, the higher is the proportion of standard English instances in the production.

### Table 10.7. The third person singular marker versus the tasks of the tests

<table>
<thead>
<tr>
<th></th>
<th>ST</th>
<th>UT</th>
<th>NS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1 (to entertain)</td>
<td>333 (70%)</td>
<td>42 (42%)</td>
<td>93 (42%)</td>
<td>468 (59%)</td>
</tr>
<tr>
<td>Task 2 (to inform)</td>
<td>105</td>
<td>36</td>
<td>60</td>
<td>201</td>
</tr>
<tr>
<td>Task 3 (to argue a case)</td>
<td>38</td>
<td>23</td>
<td>69</td>
<td>130</td>
</tr>
<tr>
<td>Total instances</td>
<td>476 (100%)</td>
<td>101 (100%)</td>
<td>222 (100%)</td>
<td>799 (100%)</td>
</tr>
<tr>
<td>Number of oral tests</td>
<td>29</td>
<td>24</td>
<td>29</td>
<td>82</td>
</tr>
</tbody>
</table>

### Table 10.8. The third person singular marker in comparison with school grades in English

<table>
<thead>
<tr>
<th>Grade</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard English instances</td>
<td>83 (86%)</td>
<td>158 (74%)</td>
<td>265 (67%)</td>
<td>51 (54%)</td>
</tr>
<tr>
<td></td>
<td>Non-standard English instances</td>
<td>16 (14%)</td>
<td>55 (26%)</td>
<td>128 (33%)</td>
<td>43 (46%)</td>
</tr>
<tr>
<td>Total</td>
<td>99 (100%)</td>
<td>213 (100%)</td>
<td>393 (100%)</td>
<td>94 (100%)</td>
<td>799 (100%)</td>
</tr>
<tr>
<td>Total number of students</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>3</td>
<td>29</td>
</tr>
</tbody>
</table>
10.1.5. Summary

In this survey of the students’ use of the third person singular marker, the results indicated that student production assigned a High grade contained a higher proportion of standard English instances than student production assigned an Average and, in particular, a Low grade. Student production assigned a High grade also contained a number of non-standard English instances but a tendency was observed that these instances were often accompanied by more standard English instances in the same production. The most common non-standard type was Type 1 (*he take). Student production rated Low contained a much higher proportion of this non-standard English type than other student production did. No other remarkable differences were found.

When the students omitted the third person singular marker, the most common type of subject was a pronoun. Instances with a non-standard English addition of the third person singular marker did not show this tendency. No noteworthy differences were found in the types of subject used in the non-standard English instances of student production or in the distribution of the non-standard English instances of the sibilant-final verbs across the three test rating categories High, Average or Low.

The instances of the third person singular marker were more common in interviews completed together with School Teachers as interviewers. The topic of Task 1 seemed to account, at least partly, for this difference in frequency.

An interesting tendency could be observed in the proportions of standard English and non-standard English instances in student production assigned different grades by the three interviewer categories. In student production assessed by Native Speakers, there was no noteworthy proportional difference between the student production assigned High, Average and Low grades. In the student production assessed by School Teachers and University Teachers, there was a higher proportion of non-standard English instances in production assigned an Average or Low grade.

As to the background variables, the proportion of standard English and non-standard English instances tallied with the students’ grades in English in school.

10.2. The Progressive

10.2.1. Overall Distribution

The material yielded 228 instances of the progressive, and they are evenly spread across the students’ production of the three test rating categories.
The distribution of the 228 instances, the totals (n), means (m), standard deviations (s), correlation (r) and the normalised figures per 1,000 words are provided in Table 10.9. The results show that 106 instances of the progressive appear in the 32 oral tests assigned a High grade, 96 instances in the 37 oral tests assigned an Average grade and 26 instances in the 13 oral tests assigned a Low grade.

The progressive is slightly more frequent in student production assigned a High grade than in other student production. All the nine oral tests that contain more than five instances of the progressive were assigned a High or Average grade. Only one oral test rated High (out of the total of 32) does not contain any instances of the progressive. The normalised figures per 1,000 words indicate no great differences between student production rated High and production rated Low. The correlation is 0.20.

Table 10.9. includes all instances of the progressive in the material. It can be assumed that there are differences of type in the total frequencies presented above; these are described below.

Table 10.9. The progressive by test rating categories

<table>
<thead>
<tr>
<th>Instances of the progressive</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10 or more</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>n (instances)</td>
<td>106</td>
<td>96</td>
</tr>
<tr>
<td>m (instances)</td>
<td>3.3</td>
<td>2.6</td>
</tr>
<tr>
<td>s (instances)</td>
<td>2.4</td>
<td>3.2</td>
</tr>
<tr>
<td>normalised figures*</td>
<td>3.3</td>
<td>2.7</td>
</tr>
</tbody>
</table>

* For a description of the calculation of the normalised figures, see 6.4. Issues to Take into Account: The Amount of Student Speech, the Length of the Oral Tests and the Correlation Coefficients.
10.2.2. Linguistic Factors

The Progressive Simple and the Progressive Complex

The material (228 instances) yielded 199 instances of the progressive simple, and 29 instances of the progressive complex. Expectedly, the former is thus a much more frequent construction than the latter. The distribution over test rating categories is shown in Table 10.10.

The complex forms were rare in the material irrespective of test rating category. There is thus nothing remarkable in the distribution of the two forms in student production of the three test rating categories.

Table 10.10. The progressive simple and the progressive complex by test rating categories

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The progressive simple</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total instances</td>
<td>94</td>
<td>82</td>
<td>23</td>
<td>199</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>2.9</td>
<td>2.2</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>The progressive complex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total instances</td>
<td>12</td>
<td>14</td>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Total instances</td>
<td>106</td>
<td>96</td>
<td>26</td>
<td>228</td>
</tr>
<tr>
<td>Total oral tests</td>
<td>32</td>
<td>37</td>
<td>13</td>
<td>82</td>
</tr>
</tbody>
</table>

Standard English and Non-standard English Instances

The proportions of the standard English and non-standard English instances across the three test rating categories High, Average and Low are presented in Table 10.11.

No great differences are found in the way the standard English and the non-standard English instances are distributed, but the slight tendency for student production assigned a High grade to contain a higher proportion of standard English instances than does student production rated Average or Low can be distinguished.

In what follows, the types of standard English and non-standard English instances are discussed.

Types of Standard English and Non-standard English Instances

Task 1 was a narrative accomplished with the help of a series of pictures which could elicit the use of the progressive (cf. Table 10.16.) In what follows, the proportions of standard English instances that are clearly related to the topics of Task 1 in the student production are discussed.
progressive in these instances is used to describe an action in progress in some pictures. The distribution of these instances (describing an action in progress in the pictures provided in Task 1) is compared with the distribution of other uses of the progressive in the test, i.e. the progressive is used in other contexts than to describe the pictures of Task 1. The three test rating categories are taken into account. It is of interest to see whether a high proportion of the standard English instances of a test rating category appears in other contexts than action in progress.

The results show that in almost half of the standard English instances (41%) the students used the progressive to describe an action in progress found in the pictures of Task 1. The proportions are similar irrespective of test rating category.

The 22 non-standard English instances attested, classified here according to semantic criteria, i.e. uses in inappropriate contexts will now be scrutinized. The instances that are non-standard English owing to the semantic context were identified to be of three types (cf. 9.1.2. The Progressive).

Most of the non-standard English instances belong to the two categories labelled Habit and Objective circumstances. The raw figures are admittedly very low but nevertheless interesting. A slight tendency can be observed that student production assigned a High grade contains non-standard English instances which refer to objective circumstances and facts. Students in production assigned an Average grade, on the other hand, used the progressive in an inappropriate way when they discussed their habits and actions which were completed.

### Table 10.11. Standard English and non-standard instances of the progressive by test rating categories

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard English</td>
<td>99 (93%)</td>
<td>84 (87%)</td>
<td>23 (88%)</td>
<td>206 (90%)</td>
</tr>
<tr>
<td>Non-standard English</td>
<td>7 (7%)</td>
<td>12 (13%)</td>
<td>3 (12%)</td>
<td>22 (10%)</td>
</tr>
<tr>
<td>Total instances</td>
<td>106 (100%)</td>
<td>96 (100%)</td>
<td>26 (100%)</td>
<td>228 (100%)</td>
</tr>
</tbody>
</table>

### Table 10.12. Types of standard English instances of the progressive by test rating categories

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Action in progress</td>
<td>40 (40%)</td>
<td>35 (42%)</td>
<td>10 (43%)</td>
<td>85 (41%)</td>
</tr>
<tr>
<td>2. Other uses</td>
<td>59 (60%)</td>
<td>49 (58%)</td>
<td>13 (57%)</td>
<td>121 (59%)</td>
</tr>
<tr>
<td>Total instances</td>
<td>99 (100%)</td>
<td>84 (100%)</td>
<td>23 (100%)</td>
<td>206 (100%)</td>
</tr>
</tbody>
</table>
As seen in the description above, the students use the progressive in several contexts which can be described as non-standard use. They probably also use the simple form in instances where the progressive would have been the appropriate form to use. These non-standard English instances, however, are not included in the present survey.

### 10.2.3. The Occurrence of the Progressive in Individual Students’ Production

The proportions of standard English and non-standard English instances in individual students’ production could be one of the factors that influences the assessment. It is thus of interest to discuss the proportions of the two categories in individual students’ production across the three test rating categories High, Average and Low.

Most non-standard English instances (a total of 82%; 41% and 41%, see Table 10.14.) appear in student production that contains either only standard English instances or both standard English and non-standard English instances. The distribution across the test rating categories shows a slight tendency for instances in production rated Average to occur together with both standard English and other non-standard English instances. The fig-

| Table 10.13. Non-standard English instances of the progressive by test rating categories |
|---------------------------------|------|------|------|------|
|                                 | High | Average | Low | Total |
| 1. Habit                        | 2    | 9      | 1    | 12    | 55%  |
| 2. Objective circumstances      | 4    | 2      | 2    | 8     | 36%  |
| 3. Stative verbs                | 1    | 1      | 2    | 2     | 9%   |
| Total of non-standard English instances | 7    | 100%   | 12   | 100%  | 22   | 100% |

As seen in the description above, the students use the progressive in several contexts which can be described as non-standard use. They probably also use the simple form in instances where the progressive would have been the appropriate form to use. These non-standard English instances, however, are not included in the present survey.

| Table 10.14. Non-standard English instances of the progressive by test rating categories |
|---------------------------------|------|------|------|------|
|                                 | High | Average | Low | Total |
| Together with only standard English | 4    | 2      | 3   | 9     | 41%  |
| Together with other non-standard English | 2    | 2      | 2   | 9%    |
| Together with both categories    | 9    | 9      | 9   | 41%   |
| The only instance in the student’s production | 1    | 1      | 2   | 9%    |
| Total                            | 7    | 12     | 3   | 22    | 100% |
ures are very low throughout but they can possibly account for the tolerance of non-standard English instances in student production assigned a High grade in comparison with production assigned an Average grade.

### 10.2.4. Occurrences of the Progressive in Relation to Interviewer Categories, Tasks and Background Variables

In this subsection, the focus is on the occurrence of the standard English and non-standard English instances of the progressive in relation to interviewer categories, the three tasks and two background variables, namely sex and the students’ grades in English in school. The distribution of the instances in relation to interviewer categories shows whether student production assessed by one category of interviewers in a particular test rating category contains higher frequencies in general or a higher proportion of standard English or non-standard English instances than does the production assessed by other categories of interviewers in the same test rating category, or whether the instances are evenly distributed. The frequencies are listed for each test rating category (see Table 10.15.).

The normalised figures show the there are more standard English than non-standard English instances in all student production irrespective of interviewer category and test rating category. Furthermore, three tendencies can be noticed.

Firstly, when the normalised figures are compared within each interviewer category, student production assigned a High grade practically al-
ways contains more standard English instances than other student production rated Average or Low by this interviewer category (ST: 3.4, 2.2 and 3.3; UT: 2.1, 1.1 and 1.1; NS: 3.9, 3.0 and 3.2). Secondly, student production rated High by Native Speakers contains no non-standard English instances, as opposed to production rated High by School Teachers (0.2) or University Teachers (0.5). All the non-standard English instances in production rated by Native Speakers were thus found in student production assigned an Average and Low grade (0.7, 1.1). Thirdly, generally the students seem to use the progressive more often in their production assessed by Native Speakers than in other production. The explanation may lie in the topics of the tasks that were different (though the tasks in the tests were identical).

The topics of Task 1 (to entertain) in the student production rated by School Teachers and Native Speakers were storytelling tasks. The students were to tell a story with the help of a series of pictures. Task 1 in the tests assessed by University Teachers was a description with no course of events (cf. 7.4. Tasks and Appendices 4A, 4B and 4C).

Most of the instances of the progressive in the student production assessed by School Teachers and Native Speakers are found in Task 1 (47% and 38%; see Table 10.16.). The topics of Task 1 with these two interviewer categories provided a context for the use of the progressive. This fact can, at least partly, account for the higher frequencies of instances in production with two of the interviewer categories, namely those whose topics naturally elicited the progressive.

The relationship to two background variables was also investigated: the students’ sex and their grades in English in school. No strong tendencies were found in the distribution of the standard English and non-standard English instances and the students’ sex.

The distribution of the instances with respect to the students’ school grades in English is shown below. Table 10.17. provides the proportions of

<table>
<thead>
<tr>
<th>Task 1 (to entertain)</th>
<th>ST</th>
<th>UT</th>
<th>NS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 2 (to inform)</td>
<td>23</td>
<td>21</td>
<td>35</td>
<td>79</td>
</tr>
<tr>
<td>Task 3 (to argue a case)</td>
<td>20</td>
<td>8</td>
<td>31</td>
<td>59</td>
</tr>
<tr>
<td>Total</td>
<td>81</td>
<td>40</td>
<td>107</td>
<td>228</td>
</tr>
</tbody>
</table>

Table 10.17. The distribution of standard English and non-standard English instances in student production according to the students’ school grades in English (ST = School Teachers; UT = University Teachers; NS = Native Speakers).
the standard English and non-standard English instances in relation to the students’ grades in English in school (grades 1 to 5, 5 being the top grade).

The figures above do not show any clear tendencies except perhaps for a very slight propensity for students with low grades in English in school to have a lower proportion of standard English instances in their production.

10.2.5. Summary

In this section the occurrence of the progressive in the student production was described. The three test rating categories, the interviewer categories and the tasks of the tests were taken into account. The proportions of the progressive simple and progressive complex and the distribution of standard English and non-standard English instances in individual students’ production were presented. In addition, the distribution of the types of standard English and non-standard English instances was discussed. Furthermore, the relationship between the occurrence of the progressive and two background variables was investigated.

Instances of the progressive were found in student production of the three test rating categories but oral tests which contained no instances were more often rated Average or Low than High. A great proportion of the instances were of the type of the progressive simple. Instances of the progressive complex were rare.

Student production assigned a High grade thus contained slightly more instances of the progressive than other student production. The great majority of all instances attested were of standard English (90%). The non-standard English instances were found in production of all the three test rating categories. Most of them appeared together with standard English instances or with both standard English instances and other non-standard English instances.

Table 10.17. The progressive in comparison with school grades in English

<table>
<thead>
<tr>
<th>Grade</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard English instances</td>
<td>22</td>
<td>96%</td>
<td>73</td>
<td>88%</td>
<td>84</td>
</tr>
<tr>
<td>Non-standard English instances</td>
<td>1</td>
<td>4%</td>
<td>10</td>
<td>12%</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>100%</td>
<td>83</td>
<td>100%</td>
<td>90</td>
</tr>
<tr>
<td>Total number of students</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>3</td>
<td>29</td>
</tr>
</tbody>
</table>

the standard English and non-standard English instances in relation to the students’ grades in English in school (grades 1 to 5, 5 being the top grade).

The figures above do not show any clear tendencies except perhaps for a very slight propensity for students with low grades in English in school to have a lower proportion of standard English instances in their production.

10.2.5. Summary

In this section the occurrence of the progressive in the student production was described. The three test rating categories, the interviewer categories and the tasks of the tests were taken into account. The proportions of the progressive simple and progressive complex and the distribution of standard English and non-standard English instances in individual students’ production were presented. In addition, the distribution of the types of standard English and non-standard English instances was discussed. Furthermore, the relationship between the occurrence of the progressive and two background variables was investigated.

Instances of the progressive were found in student production of the three test rating categories but oral tests which contained no instances were more often rated Average or Low than High. A great proportion of the instances were of the type of the progressive simple. Instances of the progressive complex were rare.

Student production assigned a High grade thus contained slightly more instances of the progressive than other student production. The great majority of all instances attested were of standard English (90%). The non-standard English instances were found in production of all the three test rating categories. Most of them appeared together with standard English instances or with both standard English instances and other non-standard English instances.
The grades assigned by the three interviewer categories did not account for any differences in the proportions of standard English and non-standard English instances. However, for the Native Speakers, student production rated High contained no non-standard instances, whilst production rated Average and Low by them did so. The progressive was also more frequent in student production assessed by Native Speakers, and by School Teachers. This difference in frequency may, at least partly, be explained by the topics in Task 1.

The students’ standard English and non-standard English use of the progressive did not tally with the students’ grades in English in school or with the students’ sex.
11.1. Abstract Nouns

11.1.1. Overall Distribution

The material yielded 579 tokens of abstract nouns. Of these, 278 instances occur in the 32 oral tests assigned a High grade, 271 in the 37 oral tests assigned an Average grade and 30 instances in the 13 oral tests assigned a Low grade. The overall distribution, normalised figures, means (m), standard deviations (s) and the correlation (r) are provided in Table 11.1.

The results show that abstract nouns occur in 75 out of the total of 82 oral tests. Student production with no instances was assigned different grades.

Table 11.1. Abstract nouns (tokens) by test rating categories

<table>
<thead>
<tr>
<th>Instances of abstract nouns</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1–3</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>4–6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>7–9</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>10–12</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>13–15</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>16–18</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>19–21</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>22–24</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>25–27</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>28 and more</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>n (instances)</td>
<td>278</td>
<td>271</td>
</tr>
<tr>
<td>m (instances)</td>
<td>8.7</td>
<td>7.3</td>
</tr>
<tr>
<td>s (instances)</td>
<td>9.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Normalised figures of instances</td>
<td>8.7</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Note: Global grades and occurrences of abstract nouns = 0.26*
Most student performance contains between one and six abstract nouns. All oral tests with more than six instances were rated High or Average. The normalised figures reveal that student production rated High or Average tends to contain more abstract nouns than student production rated Low and the correlation is 0.26.

The distribution of the types, as opposed to tokens, of abstract noun in the student production of the three test rating categories is presented below. The distribution, normalised figures, means, standard deviations and the correlation are given in Table 11.2. above.

Table 11.2. Abstract nouns (types) by test rating categories

<table>
<thead>
<tr>
<th>Instances of types of abstract nouns</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10 and more</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>n (instances)</td>
<td>173</td>
<td>189</td>
</tr>
<tr>
<td>m (types)</td>
<td>5.4</td>
<td>5.1</td>
</tr>
<tr>
<td>s (types)</td>
<td>4.8</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Most student performance contains between one and six abstract nouns. All oral tests with more than six instances were rated High or Average. The normalised figures reveal that student production rated High or Average tends to contain more abstract nouns than student production rated Low and the correlation is 0.26.

The distribution of the types, as opposed to tokens, of abstract noun in the student production of the three test rating categories is presented below. The distribution, normalised figures, means, standard deviations and the correlation are given in Table 11.2. above.

Table 11.2. shows that many oral tests (38 tests out of the total 82, 46%) contain between one and three types of abstract nouns. Fourteen (44%) oral tests assigned a High grade and 15 (40%) oral tests assigned an Average grade contain more than four types of abstract nouns. All the tests assigned a Low grade contain fewer than five types. The results obtained show a difference between student production assigned a High or Average grade on the one hand, and student production assigned a Low grade on the other hand. In the former test rating categories there are more types of abstract nouns than in the latter category. The results are in line with the tendencies observed in Table 11.1.
1.1.2. Linguistic Factors

Countable and Uncountable Abstract Nouns

In this section, the interest is on the quality of the abstract nouns in terms of countable (or count) and uncountable (or non-count) nouns. In the material, 119 types of abstract nouns appeared, and they were categorised as countable or uncountable nouns (cf. 9.2.1. Abstract Nouns). The proportions of the instances in the two categories in the production of the three test rating categories are presented in Table 11.3.

Table 11.3 shows that a great majority of the instances are countable nouns irrespective of test rating category.

Table 11.3. Countable and uncountable abstract nouns by test rating categories

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Countable nouns / Nouns of dual class membership</td>
<td>256</td>
<td>250</td>
<td>28</td>
<td>534</td>
</tr>
<tr>
<td>Countable nouns of dual class membership</td>
<td>92%</td>
<td>92%</td>
<td>93%</td>
<td>92%</td>
</tr>
<tr>
<td>Uncountable nouns</td>
<td>22</td>
<td>21</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>Uncountable nouns</td>
<td>8%</td>
<td>8%</td>
<td>7%</td>
<td>8%</td>
</tr>
<tr>
<td>Total</td>
<td>278</td>
<td>271</td>
<td>30</td>
<td>579</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

11.1.2. Standard English and Non-standard English Instances

Standard English and Non-standard English Instances

The students may have difficulties with the standard English use of the uncountable nouns (cf. 9.2.1. Abstract Nouns). For that reason, the distribution of the standard English and non-standard English instances of uncountable nouns is described.

The 45 instances of uncountable abstract nouns were classified as either standard English or non-standard English. The classification was based on the descriptions of common uncountable nouns in school grammars (cf. 6.4. Defining Standard English and Non-standard English Instances). The standard English uses from the student production are given in (76) – (77).

(76) “if you gain more confidence from being who you are ...” (YY28)
(77) “I don’t think I would have enough courage” (WW46)

Examples of non-standard English uses comprise instances when uncountable nouns are used as countable nouns, as in (78) – (79).

(78) “you can’t test a knowledge and you can’t describe a knowledge in marks because ...” (D79)
(79) “it has made a big progress” (M63)
These instances of non-standard English use above show the type found in the majority of the non-standard English instances. In these instances, students’ difficulties lie in their use of uncountable nouns as countable nouns.

Table 11.4. shows the proportions of standard English and non-standard English uses of uncountable nouns in student production of the three test rating categories. The raw figures are extraordinarily low (two instances) for student production rated Low. Since students assigned a Low grade generally used so few instances of abstract nouns, the interest is focused in particular on the instances in production rated High as opposed to those in production rated Average.

The majority of the instances were used according to the descriptions in school grammars. The proportion of standard English instances is higher in production rated Average than in production rated High.

**Types of Abstract Nouns**

As described above, 119 types of abstract noun appeared in the material. Different abstract nouns may characterise student production rated High, Average or Low. Certain types of abstract noun attested may be classified as belonging to two categories. These two categories are abstract nouns which were given in the title of the tasks, and abstract nouns with Swedish equivalents such as *problem* (the same word appears in Swedish). As English is close to Swedish, it may be expected that students who were assigned a Low grade were reluctant to leave Swedish and rather used English words which have similar equivalents in Swedish. Consequently, student production rated High would then contain more, as it were, less common abstract nouns and abstract nouns which are entirely different from Swedish. It may thus be expected that there are higher frequencies of these two categories of abstract nouns in student production rated Average or Low than in production rated High.

In Table 11.5. the distribution of the 119 types of abstract noun in student production of the three test rating categories is presented. The number of tokens for each type attested is also provided.

### Table 11.4. Standard English and non-standard English uses of uncountable abstract nouns by test rating categories

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard English uses</td>
<td>13 (59%)</td>
<td>16 (76%)</td>
<td>2</td>
<td>31 (69%)</td>
</tr>
<tr>
<td>Non-standard English uses</td>
<td>9 (41%)</td>
<td>5 (24%)</td>
<td>0</td>
<td>14 (31%)</td>
</tr>
<tr>
<td>Total</td>
<td>22 (100%)</td>
<td>21 (100%)</td>
<td>2</td>
<td>45 (100%)</td>
</tr>
</tbody>
</table>
Table 11.5. *The distribution of 119 types of abstract nouns by test rating categories*

<table>
<thead>
<tr>
<th>Abstract Noun</th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ability</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>advantage</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adventure</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>advice</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alternative</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>anger</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>appreciation</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>atmosphere</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>attention</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>attraction</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>background</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>beginning</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>calmness</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>capacity</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>career</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>cause</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>celebration</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>childhood</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>choice</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>circumstance</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>comfort</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>competition</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>confidence</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>connection</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>contact</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>control</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>courage</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>culture</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>decision</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>democracy</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>difference</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>difficulty</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>direction</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>disadvantage</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>disappointment</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disaster</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>education</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>effect</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>end</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>equality</td>
<td>5</td>
<td>9</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>excitement</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>experience</td>
<td>14</td>
<td>10</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>explanation</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fantasy</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fault</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>feeling</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>freedom</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>future</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>glamour</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>happening</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

188
<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>history</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>hobby</td>
<td>4</td>
<td>18</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>honour</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>hope</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>idea</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>identity</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>impression</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>influence</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>interest</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>key</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>knowledge</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>life</td>
<td>8</td>
<td>9</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>love</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>maturity</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>meaning</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>meeting</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>mind</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>mistake</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>moment</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>mythology</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>occasion</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>opinion</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>opportunity</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>opposite</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>option</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>part</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>patience</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>peacefulness</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>personality</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>pleasure</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>point</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>possibility</td>
<td>8</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>position</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>power</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>prejudice</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>pressure</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>problem</td>
<td>25</td>
<td>21</td>
<td>1</td>
<td>47</td>
</tr>
<tr>
<td>profit</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>progress</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>reality</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>reason</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>relation</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>relationship</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>resistance</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>respect</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>responsibility</td>
<td>11</td>
<td>6</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>result</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>revenge</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>right</td>
<td>4</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>risk</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>sake</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>school</td>
<td>5</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>similarity</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Table 11.5. reveals that the majority of the types (73 of the 119 types attested) occur once or twice in the material. The absolutely most frequent type is **way** (69 tokens) followed by **problem** (47 tokens). Other types with more than ten tokens are **celebration**, **difference**, **equality**, **experience**, **hobby**, **idea**, **life**, **responsibility**, **school**, **solution** and **time**. Most of the tokens of **difference**, **experience** and **responsibility** occur in student production rated High. **Ability**, **celebration**, **equality**, **hobby**, **life**, **meeting** and **progress** were given in the instructions of the tasks and this fact can explain their high frequencies. Thirty-one tokens of these seven types appear in student production rated High, 47 tokens in student production rated Average and 9 tokens in student production rated Low. The proportions of the tokens of these seven types given in the instructions compared with all instances (production rated High 11%; production rated Average 17%; production rated Low 30%) show that the high-achievers used types which were not given in the instructions to a greater extent than the other students did. This is also what could be expected. The proportion of the types of nouns with similar Swedish equivalents such as **problem** and **atmosphere** (Swedish: **problem** and **atmosfär**), however, is not remarkably higher in production rated Low (17%) or Average (34%) than in production rated High (32%).

Certain nouns tended to be used repeatedly by the students irrespective of test rating category, and these nouns seemed to be central in the descriptions and the discussions together with the students who participated in the interviews. Among these types, a group of five emerges. These types are **idea**, **problem**, **school**, **time** and **way**.

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>situation</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>solution</td>
<td>9</td>
<td>9</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>strength</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>subject</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>support</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>suggestion</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>symbol</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>system</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>thought</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>12</td>
<td>18</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>topic</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>tradition</td>
<td>2</td>
<td>4</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>trouble</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>truth</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>urge</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>way</td>
<td>38</td>
<td>29</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>278</td>
<td>271</td>
<td>30</td>
<td>579</td>
</tr>
</tbody>
</table>
11.1.3. Occurrences of Abstract Nouns in Relation to Interviewer Categories, Tasks and Background Variables

In this section occurrences of abstract nouns are described in relation to the interviewer categories across the three test rating categories, and regarding the distribution in the three tasks. Finally two background variables, namely the students’ sex and their grades in English in school, are taken into account.

Table 11.6. provides the distribution and the normalised figures of the types and tokens of abstract nouns in student production of the three interviewer categories across the three test rating categories High, Average and Low.

In student production assessed by Native Speakers (NS), there is a clear tendency that the higher the grade, the higher are the normalised figures for abstract nouns. This applies both to the types and the tokens. Student production assessed by the other two interviewer categories does not show this clear tendency when the distribution in production rated High and Average is compared. Student production rated Low, on the other hand, always tends to contain few instances irrespective of interviewer category.

The tests consisted of three tasks (cf. 7.4. Tasks and Appendices 4A, 4B and 4C). The topics of Task 3 were controversial issues to discuss, with three alternatives to choose from in each interview. The topics could be expected to elicit the use of some abstract nouns to a certain extent.

The distribution of the abstract nouns across the three tasks is of interest to investigate in order to see whether abstract nouns clustered in one of the three tasks and if differences could be observed when the topics of the three categories of interviewers are taken into account.

Table 11.7. shows that there is an identical tendency in production assessed by the members of the three interviewer categories: instances of

<table>
<thead>
<tr>
<th></th>
<th>ST</th>
<th>UT</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Av</td>
<td>Low</td>
</tr>
<tr>
<td>Total instances</td>
<td>105</td>
<td>116</td>
<td>15</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>9.9</td>
<td>9.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Total types</td>
<td>64</td>
<td>82</td>
<td>10</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>6.0</td>
<td>6.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Number of oral tests</td>
<td>10</td>
<td>13</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 11.6. Abstract nouns by interviewer categories
(ST = School Teachers; UT = University Teachers; NS = Native Speakers)
Abstract nouns are most frequent in Task 3. Obviously, Task 3, irrespective of interviewer category, provided contexts for the students to use abstract nouns.

It is also of interest to investigate the distribution of the instances in the tasks in relation to the test rating categories. Since frequent uses of abstract nouns tend to characterise student production rated High (cf. Table 11.1.), the results would be consistent if a majority of the abstract nouns used by students with a High grade emerged in Task 3. The proportions are presented in Table 11.8.

Table 11.8. Abstract nouns and the tasks of the tests by test rating categories

<table>
<thead>
<tr>
<th>Task 1 (“to entertain”)</th>
<th>Task 2 (“to inform”)</th>
<th>Task 3 (“to argue a case”)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>UT</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>10%</td>
<td>20</td>
<td>13%</td>
</tr>
<tr>
<td>85</td>
<td>36%</td>
<td>51</td>
<td>32%</td>
</tr>
<tr>
<td>127</td>
<td>54%</td>
<td>87</td>
<td>55%</td>
</tr>
<tr>
<td>236</td>
<td>100%</td>
<td>158</td>
<td>100%</td>
</tr>
<tr>
<td>Task 1 (“to entertain”) High</td>
<td>Average</td>
<td>Low</td>
<td>Total</td>
</tr>
<tr>
<td>32</td>
<td>12%</td>
<td>27</td>
<td>10%</td>
</tr>
<tr>
<td>70</td>
<td>25%</td>
<td>114</td>
<td>42%</td>
</tr>
<tr>
<td>176</td>
<td>63%</td>
<td>130</td>
<td>48%</td>
</tr>
<tr>
<td>278</td>
<td>100%</td>
<td>271</td>
<td>100%</td>
</tr>
<tr>
<td>Number of oral tests</td>
<td>32</td>
<td>37</td>
<td>13</td>
</tr>
</tbody>
</table>

abstract nouns are most frequent in Task 3. Obviously, Task 3, irrespective of interviewer category, provided contexts for the students to use abstract nouns.

It is also of interest to investigate the distribution of the instances in the tasks in relation to the test rating categories. Since frequent uses of abstract nouns tend to characterise student production rated High (cf. Table 11.1.), the results would be consistent if a majority of the abstract nouns used by students with a High grade emerged in Task 3. The proportions are presented in Table 11.8.

Table 11.8. gives support to the hypothesis above: in student production rated High, abstract nouns tend to cluster in Task 3 (63%), whereas student production rated Average or Low shows more similar percentages of abstract nouns in Tasks 2 and 3 (42% and 48% for the Average group, and 43% and 50% for the Low group).

In what follows, occurrences of abstract nouns in relation to the students’ sex and their grades in English in school are discussed. The distribution of the abstract nouns in relation to the students’ grades in English is presented in Table 11.9. (1 to 5, 5 being the top grade).

It is evident in Table 11.9, that the students with the top grade 5 in English in school used both more instances and more types of abstract nouns than the other students did. The high figures for students with the grade 2 in
English in school may, at least partly, be explained by the remarkably high figures of one of the students (Q). This student had noticeable lacunae in other respects of his total performance, for instance grammatical accuracy (cf. 14.4. The Individual Student’s Grades).

The results for the second student parameter, the students’ sex, are given in Table 11.10. which shows that boys’ production contains more instances of abstract nouns than girls’. The normalised figures of the types show the same strong tendency.

### Table 11.9. Abstract nouns in by school grades in English

<table>
<thead>
<tr>
<th>Grade</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total instances (tokens)</td>
<td>175</td>
<td>118</td>
<td>193</td>
<td>93</td>
<td>579</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>15.6</td>
<td>5.5</td>
<td>5.6</td>
<td>10.2</td>
<td>7.5</td>
</tr>
<tr>
<td>Total instances (types)</td>
<td>101</td>
<td>83</td>
<td>140</td>
<td>61</td>
<td>385</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>9.0</td>
<td>3.9</td>
<td>4.1</td>
<td>6.7</td>
<td>5.0</td>
</tr>
<tr>
<td>Total number of students</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>3</td>
<td>29</td>
</tr>
</tbody>
</table>

### Table 11.10. Abstract nouns by the students’ sex

<table>
<thead>
<tr>
<th></th>
<th>Boys’ production</th>
<th>Girls’ production</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total instances (tokens)</td>
<td>345</td>
<td>234</td>
<td>579</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>9.9</td>
<td>5.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Total instances (types)</td>
<td>221</td>
<td>164</td>
<td>385</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>6.3</td>
<td>3.9</td>
<td>5.0</td>
</tr>
<tr>
<td>Total number of students</td>
<td>12</td>
<td>17</td>
<td>29</td>
</tr>
</tbody>
</table>

In this section the distribution of the 579 abstract nouns was discussed. The results were compared with the three test rating categories. The interviewer categories, the tasks in the tests and two student background variables were taken into account. Three linguistic factors, namely the occurrence of countable and uncountable nouns, the correctness in the students’ use of uncountable nouns and the distribution of the 119 types of abstract nouns attested across the test rating categories, were described.

Student production assigned a High or Average grade contained more abstract nouns, both as to types and tokens, than did student production rated Low. In Native Speakers’ assessment, student production rated High
contained more instances and types of abstract nouns than student production rated Average did.

Abstract nouns were more frequent in Task 3 than in the other tasks, in particular in student production assigned a High grade.

The majority of the types of abstract noun attested were represented by one or two tokens in the material. Student production rated High contained a higher proportion of types not given in the instructions of the tests. A great majority of the abstract nouns found in the student production belonged to the category of countable nouns.

The analysis of the relation to the two background variables showed a strong tendency for students with the top grade in English in school to use many instances and types of abstract nouns in their production. Boys’ production contained more abstract nouns than girls’.

11.2. Modifiers of Adjectives

11.2.1. Overall Distribution

The material yielded 728 instances of modifiers of adjectives. Table 11.11. shows the distribution, normalised figures, means (m), standard deviations (s) and the correlation (r).

There emerged 315 modifiers in the 32 oral tests rated High, 339 in the 37 oral tests rated Average and 73 in the 13 oral tests rated Low. Most oral

<table>
<thead>
<tr>
<th>Instances of modifiers of adjectives</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0–3</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>4–6</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>7–9</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>10–12</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>13–15</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>16–18</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>19–21</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>22–24</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>25–27</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>n (instances)</td>
<td>315</td>
<td>339</td>
</tr>
<tr>
<td>m (instances)</td>
<td>9.8</td>
<td>9.2</td>
</tr>
<tr>
<td>s (instances)</td>
<td>5.1</td>
<td>6.5</td>
</tr>
<tr>
<td>Normalised figures of instances</td>
<td>9.8</td>
<td>9.4</td>
</tr>
</tbody>
</table>

194
tests contain up to twelve modifiers (63 out of the 82 tests, 77%). The frequency of instances in the individual student’s production ranges from none to 26 modifiers. The normalised figures in Table 11.10. show that frequencies do not vary greatly across the three test rating categories. There is only a weak tendency of more instances in student production rated High or Average compared with production rated Low. The correlation is 0.23.

Not only frequencies but also the types of modifiers used should be considered. Twelve types of modifiers of adjectives were found. In the individual student’s production, up to six types were used. Table 11.12. gives the frequencies of the types of modifier in the students’ production of the three test rating categories. Furthermore, the normalised figures, means, standard deviations and the correlation are provided.

In a great majority of the production (71 oral tests, 87%), the students used one to four types of modifiers. Student production which contains five types or more tends to have been rated High. It should be noticed that the normalised figures show that student production rated Low contains more types of modifiers of adjectives than other student production in the material. This unexpected tendency means that different types of modifiers of adjectives were more frequent in student production rated Low than in other student production.

Table 11.12. Modifiers of adjectives (types) by test rating categories

<table>
<thead>
<tr>
<th>Instances of types of modifiers of adjectives</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>n (types)</td>
<td>106</td>
<td>103</td>
</tr>
<tr>
<td>m (types)</td>
<td>3.3</td>
<td>2.8</td>
</tr>
<tr>
<td>s (types)</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>Normalised figures of types</td>
<td>3.3</td>
<td>2.9</td>
</tr>
</tbody>
</table>
11.2.2. The Types of Modifier

In this subsection the focus is on the distribution of the twelve types of modifier attested in the material across the three test rating categories High, Average and Low. The twelve types of modifier are presented in order of descending frequency in Table 11.13.1 Examples are given to illustrate the students’ use of them.

Not surprisingly, the list of types of modifiers in Table 11.13. reveals that very is by far the most frequent modifier of adjectives (382 instances). Quite is much more frequent than the other ten types (114 instances). These ten other types (others than very or quite) are all represented by 44 or fewer instances each. The distribution of the twelve types across the three test rating categories is provided in Table 11.14. Percentages are given showing the relative frequency of each modifier within each test rating category.

As can be seen from Table 11.14., very is by far the most common modifier of adjectives regardless of test rating category. Two other tendencies are of interest. Firstly, the higher the grade, the lower is the proportion of very, and the higher is the proportion of alternative modifiers such as really or pretty. Students’ in production rated High thus used alternative types to very. There is a higher proportion of the types a bit, a little bit, pretty, really and rather in production assigned a High grade than in production rated

---

1. Quite a bit and a little bit were grouped together with a bit as a bit is the head of the expression.
Average or Low. Secondly, there is the opposite tendency for the type so. There is a higher proportion of so in student production assigned an Average or Low grade than in production rated High.

11.2.3. Occurrences of Modifiers of Adjectives in Relation to Interviewer Categories and Background Variables

The members of the different interviewer categories may be more or less influenced by the varying frequency of modifiers of adjectives in their assessment. It is therefore of relevance to present the distribution of the modifiers in the test rating categories across interviewer categories.

Table 11.15. shows that modifiers of adjectives tend to be more frequent in student production assessed by University Teachers. This may probably be explained by the topics of Task 1. The topic of Task 1 in the interviews assessed by University Teachers was a description and a comparison of two houses (cf. 7.4. Tasks). It is likely that this topic provided a context for a frequent use of modifiers of adjectives. Furthermore, the results show very disparate tendencies in the distribution in the test rating categories across the three interviewer categories but one evident tendency should be commented on. Native Speakers seem to be influenced by the variety of modifiers of adjectives used, to judge from the normalised figures of student production rated High (ST: 2.8; UT: 3.7; NS: 7.4). The other normalised fig-

<table>
<thead>
<tr>
<th>Modifier</th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>very</td>
<td>150</td>
<td>47%</td>
<td>186</td>
<td>55%</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>63%</td>
<td>6</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>16%</td>
<td>4</td>
<td>10%</td>
</tr>
<tr>
<td>little bit</td>
<td>18</td>
<td>15%</td>
<td>11</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>11%</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>quite a bit</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>so</td>
<td>8</td>
<td>2%</td>
<td>27</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>8%</td>
<td>41</td>
<td>6%</td>
</tr>
<tr>
<td>pretty</td>
<td>25</td>
<td>8%</td>
<td>14</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7%</td>
<td>44</td>
<td>6%</td>
</tr>
<tr>
<td>really</td>
<td>18</td>
<td>6%</td>
<td>13</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0%</td>
<td>31</td>
<td>4%</td>
</tr>
<tr>
<td>rather</td>
<td>13</td>
<td>4%</td>
<td>9</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1%</td>
<td>23</td>
<td>3%</td>
</tr>
<tr>
<td>totally</td>
<td>6</td>
<td>2%</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1%</td>
<td>8</td>
<td>1%</td>
</tr>
<tr>
<td>a little</td>
<td>4</td>
<td>1%</td>
<td>3</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0%</td>
<td>7</td>
<td>1%</td>
</tr>
<tr>
<td>fairly</td>
<td>1</td>
<td>1%</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1%</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Total</td>
<td>315</td>
<td>100%</td>
<td>339</td>
<td>100%</td>
</tr>
</tbody>
</table>

* Percentages are approximate due to low raw figures.
ures do not tally with the Global grades assigned to the student production by the three interviewer categories.

In what follows, the relation to two background variables is presented: the students’ grades in English in school and the students’ sex. The distribution of the modifiers of adjectives and the normalised figures are presented in relation to the students’ grades in English in school in Table 11.16.

The results show no convincing relationship between the frequency of modifiers of adjectives and the students’ grades in English in school. Neither tokens nor types of modifiers of adjectives show any noteworthy tendencies.

In what follows the relationship between the students’ sex and the distribution of the modifiers of adjectives is described. The proportions and normalised figures are given in Table 11.17.

The results show a slight tendency for girls’ production to contain more types and more tokens of modifiers of adjectives than do boys’ production.
11.2.4. Summary

In this section the occurrence of modifiers of adjectives in the student production was presented. The three test rating categories and the three interviewer categories were taken into account. The distribution of the twelve types of modifiers of adjectives was presented in the three test rating categories. Finally, two background variables were taken into account.

The results indicated a slight tendency for student production assigned a High or Average grade to contain more instances of modifiers than did production assigned a Low grade. The results were not decisive, however. Furthermore, the results showed that students in production rated High by Native Speakers used more types, i.e. showed variety in their use of modifiers.

The presentation of the distribution of the twelve types of modifiers showed that very stood out as the most frequent modifier. The modifiers pretty, rather and really tended to be more frequent in student production assigned a High grade than in other production. There was the opposite tendency for so.

Frequent uses of modifiers of adjectives did not tally with the students’ grades in English in school. There was a weak tendency of there being more modifiers of adjectives in girls’ production than in boys’.

### Table 11.17. Modifiers of adjectives by the students’ sex

<table>
<thead>
<tr>
<th></th>
<th>Boys’ production</th>
<th>Girls’ production</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total instances (tokens)</td>
<td>287</td>
<td>440</td>
<td>727</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>8.2</td>
<td>10.5</td>
<td>9.5</td>
</tr>
<tr>
<td>Total instances (types)</td>
<td>99</td>
<td>143</td>
<td>242</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>2.8</td>
<td>3.4</td>
<td>3.2</td>
</tr>
<tr>
<td>Total number of students</td>
<td>12</td>
<td>17</td>
<td>29</td>
</tr>
</tbody>
</table>

11.3. Students’ Lexical Errors and Compensatory Strategies

11.3.1. Occurrences of Lexical Errors

The material yielded 106 instances of lexical errors of the twelve types selected for the investigation (cf. 9.2.3. Students’ Lexical Errors and Compensatory Strategies) in 48 of the 82 oral tests of the material.\(^2\) These 48 instances were selected according to the criteria defined in 9.2.3.
tests which contain at least one lexical error spread across the three test rating categories High, Average and Low. The distribution, normalised figures, means (m), standard deviations (s) and correlation (r) are presented in Table 11.18.

Most student production contains no errors or no more than one error (68 of the 82 oral tests, 83%). Sixteen (50%) of the oral tests rated High do not contain any instances of the lexical errors investigated. Fourteen tests (38%) rated Average and four tests (30%) rated Low also fall into this category. The normalised figures show that student production rated Low tends to contain more lexical errors (1.9) than production rated Average (1.3) or High (1.3).

In what follows, the distribution of the number of types of lexical error in the three test rating categories is presented. The distribution, normalised figures, means (m), standard deviations (s) and correlation (r) are given in Table 11.19.

The distribution and the normalised figures of the number of types of lexical errors in each test rating category tally with the results of frequencies described in Table 11.18. The normalised figures show that student production rated Low generally contains more types of lexical errors than other production (1.4 compared with 0.7 and 1.0). There is a difference between Tables 11.19. and 11.18. when student performances rated High and Average are compared. In Table 11.18. production rated High and Average yield the same results whereas in 11.19. student production rated Average contains slightly more types than production rated High.

### Table 11.18. Lexical errors (tokens) by test rating categories

<table>
<thead>
<tr>
<th>Instances of lexical errors</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>16 (50%)</td>
<td>14 (38%)</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32 (100%)</td>
<td>37 (100%)</td>
</tr>
<tr>
<td>n (instances)</td>
<td>42</td>
<td>47</td>
</tr>
<tr>
<td>m (instances)</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>s (instances)</td>
<td>1.9</td>
<td>1.3</td>
</tr>
<tr>
<td>Normalised figures of instances</td>
<td>1.3</td>
<td>1.3</td>
</tr>
</tbody>
</table>

In what follows, the distribution of the number of types of lexical error in the three test rating categories is presented. The distribution, normalised figures, means (m), standard deviations (s) and correlation (r) are given in Table 11.19.
11.3.2. Students’ Compensatory Strategies

There were 224 instances of compensatory strategies found in the material. These instances are indications of students’ lexical lacunae or their obvious uncertainty about the appropriateness of words used. The students used Strategies involving Swedish, for instance request for help in Swedish, and Strategies in English, for instance paraphrase. 

Table 11.19. Lexical errors (types) by test rating categories

<table>
<thead>
<tr>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32</td>
</tr>
<tr>
<td>m (instances)</td>
<td>0.7</td>
</tr>
<tr>
<td>s (instances)</td>
<td>9</td>
</tr>
<tr>
<td>Normalised figures of instances</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Table 11.20. Compensatory strategies by test rating categories

<table>
<thead>
<tr>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32</td>
</tr>
<tr>
<td>m (instances)</td>
<td>2.2</td>
</tr>
<tr>
<td>s (instances)</td>
<td>1.8</td>
</tr>
<tr>
<td>Normalised figures of instances</td>
<td>2.2</td>
</tr>
</tbody>
</table>

1 For a discussion and description of the term “compensatory strategy”, see 1.3. and for the classification, see 9.2.3.
satory strategies across test rating categories, normalised figures, means (m), standard deviations (s) and correlation (r) are presented in Table 11.20.

The normalised figures show that student production rated Low contains a considerably greater number of instances where the students resorted to compensatory strategies than in other student production (5.9 compared with 2.2 and 2.9). There is only a weak tendency for more instances to occur in student production rated Average than in student production rated High. A single instance, or none, of a compensatory strategy in a test is more frequent in student production rated High (16 tests, 50%), than in production rated Average (12 tests, 32%), or Low (1 test, 8%). The correlation is –0.35.

In what follows the distribution of the twelve types of strategy selected (cf. 9.2.3. Students’ Lexical Errors and Compensatory Strategies) is described across the three test rating categories.

The distribution of the total number of instances of the two categories (Strategies involving Swedish and Strategies in English) shows that students generally use more Strategies in English (71%) than Strategies involving Swedish (29%). This tendency is strongest however for students rated High (80% / 20%).

### Table 11.21. Twelve types of compensatory strategy by test rating categories

<table>
<thead>
<tr>
<th>Strategies involving Swedish</th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Request for help in Swedish</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>2. Swedish words used</td>
<td>7</td>
<td>14</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>3. Swedish words pronounced in English</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>4. Swedish words explained in English</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5. Swedish words added</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>6. Word coinage based on Swedish</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total Strategies involving Swedish</strong></td>
<td>14</td>
<td>34%</td>
<td>17</td>
<td>33%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategies in English</th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Request for help in English</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>8. Waiting for assistance</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>9. Substitution by superordinate or approximate terms</td>
<td>9</td>
<td>15</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>10. Word coinage based on English</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>11. Giving two alternatives</td>
<td>23</td>
<td>21</td>
<td>7</td>
<td>51</td>
</tr>
<tr>
<td>12. Paraphrase</td>
<td>13</td>
<td>11</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td><strong>Total Strategies in English</strong></td>
<td>55</td>
<td>80%</td>
<td>69</td>
<td>67%</td>
</tr>
</tbody>
</table>

| **Total Strategies used** | 69 100% | 103 100% | 52 100% | 224 100% |
| **Total number of oral tests** | 32   | 27       | 13 | 82    |
The most common type of strategy used is “Giving two alternatives” (51 instances). There is a higher proportion of these instances in student production rated High (23 instances, 33%) than elsewhere (Average 21 instances, 20%; Low, 7 instances, 13%).

The interaction between the interviewers and students during the tests and the interviewers’ background may influence the students’ choice of type of strategy. It may be expected, for instance, that student production assessed by Swedish interviewers contains a higher proportion of Strategies involving Swedish than production assessed by Native Speakers. When the students realised that the interviewers were Swedish they may have turned to the Swedish language rather than trying to cope in English when faced with vocabulary difficulties or when they found it difficult to express what they wanted to say. For that reason, a survey of the distribution of the 224 instances of compensatory strategies across the three interviewer categories is provided in Table 11.22.

<table>
<thead>
<tr>
<th>Strategies</th>
<th>ST</th>
<th>UT</th>
<th>NS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategies involving Swedish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Request for help in Swedish</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2. Swedish words used</td>
<td>17</td>
<td>4</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>3. Swedish words pronounced in English</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>4. Swedish words explained in English</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5. Swedish words added</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>6. Word coinage based on Swedish</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Total Strategies involving Swedish</td>
<td>37</td>
<td>37%</td>
<td>13</td>
<td>22%</td>
</tr>
<tr>
<td>Strategies in English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Request for help in English</td>
<td>14</td>
<td>0</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>8. Waiting for assistance</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>9. Substitution by superordinate or approximate terms</td>
<td>18</td>
<td>9</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>10. Word coinage based on English</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>11. Giving two alternatives</td>
<td>17</td>
<td>15</td>
<td>19</td>
<td>51</td>
</tr>
<tr>
<td>12. Paraphrase</td>
<td>8</td>
<td>17</td>
<td>8</td>
<td>33</td>
</tr>
<tr>
<td>Total Strategies in English</td>
<td>64</td>
<td>63%</td>
<td>46</td>
<td>78%</td>
</tr>
<tr>
<td>Total Strategies used</td>
<td>101</td>
<td>100%</td>
<td>59</td>
<td>100%</td>
</tr>
<tr>
<td>Total number of oral tests</td>
<td>29</td>
<td>24</td>
<td>29</td>
<td>82</td>
</tr>
</tbody>
</table>

The most common type of strategy used is “Giving two alternatives” (51 instances). There is a higher proportion of these instances in student production rated High (23 instances, 33%) than elsewhere (Average 21 instances, 20%; Low, 7 instances, 13%).

The interaction between the interviewers and students during the tests and the interviewers’ background may influence the students’ choice of type of strategy. It may be expected, for instance, that student production assessed by Swedish interviewers contains a higher proportion of Strategies involving Swedish than production assessed by Native Speakers. When the students realised that the interviewers were Swedish they may have turned to the Swedish language rather than trying to cope in English when faced with vocabulary difficulties or when they found it difficult to express what they wanted to say. For that reason, a survey of the distribution of the 224 instances of compensatory strategies across the three interviewer categories is provided in Table 11.22.
The results show that students were more inclined to use compensatory strategies, and in particular Strategies involving Swedish, in student production assessed by School Teachers (ST: 101 instances; UT: 59 instances; NS: 64 instances). Another tendency should also be commented on: a higher proportion of strategies of request for help, in Swedish or in English, appeared in student production with School Teachers than with other interviewers. The tests with School Teachers contain 20 instances (20%), the ones with University Teachers 1 instance (2%) and the ones with Native Speakers 7 instances (11%).

11.3.3. Summary

In this section the discussion included a description of occurrences of lexical errors related to a selected list of words and the compensatory strategies the students used to solve vocabulary difficulties. The compensatory strategies could also signal their uncertainty about the appropriateness of an English word or phrase. The three test rating categories High, Average and Low as well as the three interviewer categories were taken into account.

According to the results, most lexical errors emerged in student production rated Low. Student production rated High contained fewer instances, both in terms of types and tokens, than other student production. A slight tendency of more types of lexical errors in student production rated Average, compared with production rated High, could be observed.

The survey indicated that student production rated Low contained considerably more instances of compensatory strategies than other student production. As regards the types of compensatory strategies used by the students, the results showed that the students in production assigned a High grade used a higher proportion of Strategies in English than Strategies involving Swedish compared to students in the Average or Low category where Swedish was resorted to more often.

When the interviewer categories were compared, it was found that student production assessed by School Teachers contained a higher proportion of Strategies involving Swedish than production assessed by the members of the other two interviewer categories. It also contained more requests for help in English and Swedish than production assessed by University Teachers or Native Speakers.
CHAPTER 12
Discourse Markers

12.1. *Well*...

The present description of *well* in the student production is structured as follows. First the occurrence of *well* in the three test rating categories High, Average and Low will be surveyed. Then the discoursal factors will be described. Finally, the relationships to the three interviewer categories, the tasks of the tests and two background variables will be dealt with.

12.1.1. Overall Distribution

The data yielded 407 instances of *well* as a discourse marker. The distribution, normalised figures, means (m), standard deviations (s) and the correlation (r) are provided in Table 12.1. The proportions (in percentages) within the test rating categories that contain no instances are also given.

The results show that the 407 instances were spread across the three test rating categories and that the frequencies in the individual students’ production range from none up to more than 15 instances. It may be observed that 24 oral tests (29%) contain no instances and that higher proportions of oral tests rated Average (35%) or Low (38%) contain no instances compared with student production rated High (19%). Six oral tests contain relatively high frequencies with more than 15 instances (16, 16, 19, 19, 26 and 27 instances). These oral tests were rated High or Average. The majority of the oral tests that contain 10 or more instances were rated High (12 out of the 18 oral tests). The normalised figures give support to the observation that student production rated High contains more instances of *well* (7.0) than student production rated Average (4.5) or Low (2.6). The correlation is 0.31. In what follows, the contexts of *well* attested in the students’ production are described.
12.1.2. Discoursal Factors

Mere frequency does not give a complete picture of the students’ uses of well. For that reason, there is an account of the discoursal factors identified.¹

The Discoursal Functions of well

Well can be used in a number of discoursal functions. In what follows, the list of the six categories of discoursal functions described in Chapter 9 is provided (cf. 9.3.1. Well…). The proportions of the discoursal functions are compared with respect to the three test rating categories High, Average and Low.

The material yielded instances of five types of discoursal functions of well:

1. well in answers to the interviewers’ questions and requests,
2. well in reported speech,

¹ An investigation of the collocations of well was carried out. Well did not appear in recurrent combinations with other words, however; 44 of the 407 instances (11%) of well occurred together with I don’t know, I mean, I think or then.

Table 12.1. Well by test rating categories

<table>
<thead>
<tr>
<th>Instances of well</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>6 (19%)</td>
<td>13 (35%)</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>15 and more</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32 (100%)</td>
<td>37 (100%)</td>
</tr>
<tr>
<td>n (instances)</td>
<td>223</td>
<td>161</td>
</tr>
<tr>
<td>m (instances)</td>
<td>7.0</td>
<td>4.4</td>
</tr>
<tr>
<td>s (instances)</td>
<td>6.5</td>
<td>6.3</td>
</tr>
<tr>
<td>Normalised figure</td>
<td>7.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>
(3) well in repetitions,

(4) well in contexts where students express difficulties with English, and

(5) well as a finishing signal.

The remaining instances were classified as

(6) well in indeterminate contexts.

The distribution of the six types across the three test rating categories High, Average and Low is provided in Table 12.2.

The results show that nearly half of the instances occur in indeterminate contexts (48%). By far the most frequent discoursal function identified is well used in answers to the interviewers’ questions and requests (44%). This is the only discoursal function identified in student production rated Low. In student production rated High, well appeared in connection with reported speech and in conclusions. Well was used in repetitions and in contexts of linguistic difficulties mainly by students assigned an Average grade.

12.1.3. Occurrences of well in Relation to Interviewer Categories, Tasks and Background Variables

In this section, the distribution of the instances in the student production with respect to the interviewer categories, the tasks of the tests and two background variables are discussed.

The purpose of investigating the instances in relation to interviewer categories is to find out whether student production which contains many instances of well was assigned a High grade by all three categories of interviewers or not. The total and normalised figures in student production rated High, Average and Low by the School Teachers (ST), the University Teachers (UT) and the Native Speakers (NS) are provided in Table 12.3.
The normalised figures show a clear difference in the distribution of the instances of well in student production across the three interviewer categories. Student production rated High by School Teachers (ST) in particular, and to some extent by University Teachers (UT), contains more instances (9.9 and 5.8) than student production rated Average (5.0 and 2.1) or Low (1.2 and 3.0) by them. The instances in production rated by Native Speakers (NS) do not show this tendency (5.2, 5.2 and 5.4). This finding is discussed in 14.3. Interviewer Categories, and in 15.3. Discourse Markers.

A very high rate of well in one student’s production could explain the observed differences between grades assigned in relation to the use of well. Moderate uses of well may result in positive assessments, whereas student production with high frequencies could result in another grade. For that reason, it was of interest to investigate the grades of the tests that contain a remarkably high frequency. Oral tests which contain more than ten instances of well were investigated in order to see whether this hypothesis could account for the different tendencies observed in Table 12.1. The results of the investigation showed that the distribution of the oral tests with ten or more instances of well confirmed the tendencies observed in Table 12.4. (High frequencies of well were rated differently by the interviewer categories.). Eight of the ten oral tests which contain more than ten instances and rated by School Teachers or University Teachers were assigned a High grade whereas four of the eight oral tests with more than ten instances were rated High by the Native Speakers. The other four oral tests were rated Average.

A description of the distribution of the instances of well in the three tasks of the tests may provide useful information. It could be expected that the great majority of the instances emerged in the unprepared discussion of Task 3 (cf. 7.6. Tasks, and furthermore, the discussion of well as an evincive in 1.3. Linguistic Analyses of Students’ Proficiency). A high proportion of well in Task 3 could be evidence of the students’ need to use well in their

<table>
<thead>
<tr>
<th></th>
<th>ST</th>
<th>UT</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Av</td>
<td>Low</td>
</tr>
<tr>
<td>Total instances</td>
<td>105</td>
<td>61</td>
<td>5</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>9.9</td>
<td>5.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Total oral tests</td>
<td>10</td>
<td>13</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 12.3. Well by interviewer categories
(ST = School Teachers; UT = University Teachers; NS = Native Speakers)

The normalised figures show a clear difference in the distribution of the instances of well in student production across the three interviewer categories. Student production rated High by School Teachers (ST) in particular, and to some extent by University Teachers (UT), contains more instances (9.9 and 5.8) than student production rated Average (5.0 and 2.1) or Low (1.2 and 3.0) by them. The instances in production rated by Native Speakers (NS) do not show this tendency (5.2, 5.2 and 5.4). This finding is discussed in 14.3. Interviewer Categories, and in 15.3. Discourse Markers.

A very high rate of well in one student’s production could explain the observed differences between grades assigned in relation to the use of well. Moderate uses of well may result in positive assessments, whereas student production with high frequencies could result in another grade. For that reason, it was of interest to investigate the grades of the tests that contain a remarkably high frequency. Oral tests which contain more than ten instances of well were investigated in order to see whether this hypothesis could account for the different tendencies observed in Table 12.1. The results of the investigation showed that the distribution of the oral tests with ten or more instances of well confirmed the tendencies observed in Table 12.4. (High frequencies of well were rated differently by the interviewer categories.). Eight of the ten oral tests which contain more than ten instances and rated by School Teachers or University Teachers were assigned a High grade whereas four of the eight oral tests with more than ten instances were rated High by the Native Speakers. The other four oral tests were rated Average.

A description of the distribution of the instances of well in the three tasks of the tests may provide useful information. It could be expected that the great majority of the instances emerged in the unprepared discussion of Task 3 (cf. 7.6. Tasks, and furthermore, the discussion of well as an evincive in 1.3. Linguistic Analyses of Students’ Proficiency). A high proportion of well in Task 3 could be evidence of the students’ need to use well in their
planning of speech in an unprepared task. Furthermore, it could indicate that the discussion in Task 3 resulted in many answers introduced by well.

Table 12.4. Well and the tasks of the tests by test rating categories

<table>
<thead>
<tr>
<th>Task 1 (&quot;to entertain&quot;)</th>
<th>High (35%)</th>
<th>Average (20%)</th>
<th>Low (35%)</th>
<th>Total (29%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 2 (&quot;to inform&quot;)</td>
<td>67 (30%)</td>
<td>65 (41%)</td>
<td>9 (39%)</td>
<td>141 (35%)</td>
</tr>
<tr>
<td>Task 3 (&quot;to argue a case&quot;)</td>
<td>79 (35%)</td>
<td>63 (39%)</td>
<td>6 (26%)</td>
<td>148 (36%)</td>
</tr>
<tr>
<td>Total</td>
<td>223 (100%)</td>
<td>161 (100%)</td>
<td>30 (100%)</td>
<td>407 (100%)</td>
</tr>
</tbody>
</table>

Number of oral tests 82

Table 12.5. Well and the tasks of the tests by interviewer categories

<table>
<thead>
<tr>
<th>Task 1 (&quot;to entertain&quot;)</th>
<th>ST (36%)</th>
<th>UT (33%)</th>
<th>NS (18%)</th>
<th>Total (29%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 2 (&quot;to inform&quot;)</td>
<td>58 (34%)</td>
<td>25 (28%)</td>
<td>58 (39%)</td>
<td>141 (35%)</td>
</tr>
<tr>
<td>Task 3 (&quot;to argue a case&quot;)</td>
<td>51 (30%)</td>
<td>34 (39%)</td>
<td>63 (42%)</td>
<td>148 (36%)</td>
</tr>
<tr>
<td>Total</td>
<td>171 (100%)</td>
<td>88 (100%)</td>
<td>148 (100%)</td>
<td>407 (100%)</td>
</tr>
</tbody>
</table>

Number of oral tests 82

planning of speech in an unprepared task. Furthermore, it could indicate that the discussion in Task 3 resulted in many answers introduced by well.

Table 12.4. shows that the instances were quite evenly spread among the three tasks of the tests and that the students of the three test rating categories do not show different tendencies in their uses of well regarding the distribution across the tasks of the tests. Below the distribution in the three tasks in relation to the three interviewer categories is elaborated on.

The figures in Table 12.5. show no great differences in distribution except for one tendency that should be commented on. Task 1 in student production assessed by the Native Speakers (NS) contains a lower proportion of instances (18%) than student production with the other interviewers (36%, 33%). This finding is discussed in 15.3. Discourse Markers.

In what follows, the focus is on the relationship between the occurrence of well and two background variables: the students’ sex and the students’ grades in English in school. First the distribution of the instances in girls’ and boys’ production is discussed.

The results in Table 12.6. show that the boys were inclined to use well for various purposes, such as to hesitate and fill pauses in speech, more often than the girls did.

As regards the second student parameter, namely the students’ grades in English in school, the distribution and normalised figures of well in com-
The normalised figures in Table 12.7. show the clear tendency that students who had the top grades 4 and 5 in English in school more often used well than did the other students. The normalised figure for the students with the low grade 2 is remarkably high, however. This figure can be explained by the fact that the performance by two of the three students who had the low grade 2, hardly contains any instances at all, whereas one student (Q) used well up to 19 times in one interview. The normalised figure for students who had the low grade 2 is consequently highly influenced by the instances in the performance by one individual student.

### 12.1.4. Summary

In this section the students’ use of well was described. The test rating categories, the three interviewer categories, the tasks of the tests and the relation to two background variables were taken into account. Furthermore, a survey of the discoursal functions was provided.

Generally, most instances of well emerged in student production rated High. Individual students’ production which contained high frequencies was assigned a High or Average grade. These were the tendencies in production rated by School Teachers or University Teachers. The distribution of the instances in student production assessed by Native Speakers did not show this pattern. Their grading did not seem to have been influenced by the occurrence of well.
The majority of the instances of *well* were in indeterminate contexts, and probably used as an item used in for instance hesitations or as a gap-filling device. Another identified function was in answers to the interviewers’ questions and requests. Student production rated High contained a larger proportion of instances in other discoursal functions, such as in conclusions, than in answers to questions and requests.

The boys used *well* slightly more often than the girls did and the frequency of *well* correlated with the students’ grades in English in school. In spite of some intra-group variation, it was clear that students with high grades in English in school tended to use *well* more often.

### 12.2. *You know...*

The structure of the present section is as follows. First the occurrence of *you know* is presented in the student production across the three test rating categories High, Average, and Low. Then collocational factors along with the position of *you know* in the utterance are described. After these presentations, the discoursal functions of *you know* are taken into account. Furthermore, it is of interest to check the occurrence of *you know* across the student production rated by the three interviewer categories together with a presentation of the interviewers’ uses of *you know* in the interviews. For that reason, a subsection is included where the Global grades assigned to the students’ production are compared with the uses of *you know* by the members of the three interviewer categories. Finally, surveys of the distribution across the tasks of the tests and the relationship to two background variables are provided.

#### 12.2.1. Overall Distribution

There were 94 instances of *you know* in 34 oral tests. Student production which contains instances spread across the three test rating categories High, Average, and Low. Table 12.8. shows the distribution, normalised figures, means (m), standard deviations (s) and the correlation (r). Furthermore, the proportions (in percentages) of production within the test rating categories that contain no instances are given.

The majority of the oral tests (59%) contain no instances of *you know*. No clear tendencies can be observed when the proportions of oral tests with no instances in the three test rating categories are compared. The normalised figures show some interesting tendencies, however. *You know* is slightly more frequent in student production assigned an Average or Low
grade (1.3 and 1.9) than in production rated High (0.9). Discoursal functions and syntactic contexts may account for these differences in the students’ uses.

12.2.2. Collocational and Discoursal Factors

In this subsection, the distribution of the linguistic items that occur together with *you know* is presented. Then, a survey of the position of *you know* in the utterances in terms of turn-initial, medial, and turn-final position is provided. Finally, the distribution of the instances in four discoursal functions is described. In each survey, the proportions in the three test rating categories High, Average, and Low are compared.

**Collocational Factors**

The occurrence of recurrent linguistic items together with *you know* may provide useful information. These different linguistic items could possibly provide evidence of stylistic features of the student performance in one of the three test rating categories High, Average, and Low.

Only one type of collocation occurred more than twice in the material: *and you know*/*you know and*. Twelve instances were found of this collocation and the great majority of them were in student production rated High: 9 out of the 12 instances (75%). When the proportions of the instances of

<table>
<thead>
<tr>
<th>Instances of <em>you know</em></th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>0</td>
<td>18 (56%)</td>
<td>23 (62%)</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

| n (instances) | 30 | 47 | 17 | 94 |
| s (instances) | 1.6 | 2.1 | 2.1 | 1.9 |

Normalised figures 0.9 1.3 1.9 1.2
and you know you know and were compared with the total instances in one test rating category, the highest proportion was found in student production rated High (9 out of the total 30 instances, 30%).

The Position in the Utterance

The 94 instances were categorised as markers in turn-initial, medial, or turn-final position. The proportions of the instances in the three positions across the three test rating categories High, Average and Low are provided in Table 12.9.

The results show that a great majority of the instances are in medial position irrespective of Global grades assigned. The results are in line with previous research and native speakers’ use (Erman 1987: 50) described in Chapter 1 (cf. 1.3. Linguistic Analyses of Students’ Proficiency).

Discoursal Functions

Three discoursal functions of you know were attested: expressing an opinion, word searching, and you know in contexts with repetition. Remaining instances of you know were classified as indeterminate (cf. 9.3.2. You know…). In what follows, a survey of the instances in the three discoursal functions and the indeterminate instances across the three test rating categories High, Average and Low are provided.

The figures in Table 12.10. show two tendencies worth commenting on. Firstly, the majority of the instances (66%) are in the category “Indeterminate instances”, and the tendency is evident irrespective of test rating category. Secondly, the remaining instances which were categorised show differences in distribution mainly across the two test rating categories High and Average. You know emerged in contexts of students’ vocabulary gaps and in repetitions in student production rated Average whereas student production rated High contains more instances in contexts when the students expressed their own opinion.

Table 12.9 You know and the position in the utterance by test rating categories

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn-initial position</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Medial position</td>
<td>22 (73%)</td>
<td>31 (66%)</td>
<td>13 (76%)</td>
<td>66 (70%)</td>
</tr>
<tr>
<td>Turn-final position</td>
<td>6</td>
<td>13</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>Total instances</td>
<td>30 (100%)</td>
<td>47 (100%)</td>
<td>17 (100%)</td>
<td>94 (100%)</td>
</tr>
<tr>
<td>Number of oral tests</td>
<td>32</td>
<td>37</td>
<td>13</td>
<td>82</td>
</tr>
</tbody>
</table>
12.2.3. Occurrences of you know in Relation to Interviewer Categories, Tasks and Background Variables

There are at least two reasons why it is highly interesting to take into account the three interviewer categories in the case of you know. Firstly, the interviewers’ background can account for the different grades assigned to the student production that contains many instances. You know is described as stigmatised in some contexts (Schiffrin 1987) and the interviewers may show their approval or disapproval of the students’ idiomatic or unidiomatic use of such an item. It is therefore of interest to see whether members of one of the three interviewer categories reacted in a negative way towards the students’ use of you know.

Secondly, as described in Chapter 1 (cf. 1.3. Linguistic Analyses of Students’ Proficiency), you know is an item which may be used as a signal towards a more personal and relaxed conversation. It may be expected that the first step in this direction is taken by the interviewers in oral proficiency interviews. Evidently, there may be other items that work in the same direction as you know, but nevertheless it is of great interest to investigate the students’ use (cf. Table 12.12) and the interviewers’ use (cf. Table 12.13) of you know in the interviews.

The occurrence of the 94 instances of you know in student production assessed by the members of the three interviewer categories (School Teachers, University Teachers and Native Speakers) is provided in Table 12.11.

Table 12.11. shows remarkable differences in the distribution of you know in the student production across the three interviewer categories. Student production rated Low by Native Speakers (NS) contains 13 instances of you know and has a high normalised figure (7.0). Very few instances emerged in student production rated Low by the other two interviewer categories (UT: 2, ST: 2). Student production rated by University Teachers shows the opposite tendency: student production rated High contains many instances (17 instances, normalised figure: 1.5).
The focus of interest is now on the relationship of the students' use of *you know* to the interviewers' use of *you know* in the interviews. The results so far have indicated that the students' use of *you know* rather characterises student production rated Average or Low than High, at least when assessed by Native Speakers. The use of *you know* can be interpreted as a means of creating a more personal and intimate conversation. Since the interviewers were expected to conduct the conversation in the oral test, they might have regarded it as inappropriate if the students used linguistic items that did not fit into the interaction of the oral test. In the approach that follows, it is assumed that the students' uses of *you know* were considered to be more appropriate if they were accompanied by instances produced by the interviewers in the same interview. A higher proportion of oral tests that only contains the students' use of *you know*, and thus no interviewers' use, would be expected to be rated Average or Low rather than High.

The distribution of the tests that contain students' uses of *you know* compared with the tests where the interviewers used or did not use *you know* is presented in Table 12.12. Only the 34 oral tests that contain students' uses of *you know* are included in the presentation (cf. Table 12.8.). The three test rating categories High, Average and Low are taken into account.

A majority of the instances of *you know* emerged in student production when the interviewers did not use *you know* (65%). The results show that a higher proportion of the instances emerge in production rated High (50%) than Average (21%) when the interviewers also used *you know*.
The raw figures of the student production assigned a Low grade are low, but they still show the same tendency as the figures in production rated Average.

The distribution of the instances in the three tasks may provide information regarding whether certain tasks elicited the students’ use of *you know*. It may be expected that a high proportion of the instances emerged in Task 2, which was a narrative task (cf. the discussion of the occurrence of *you know* in narrative parts of conversation in 1.3. Linguistic Analyses of Students’ Proficiency). The distribution of the 94 instances of *you know* in the three tasks is given in Table 12.13. The three test rating categories High, Average and Low are taken into account.

A high proportion of the instances cluster in Task 2 (50%), which had narrative topics in which the students informed the interviewers about one of three issues (cf. 7.4. Tasks).

The relationship between two background variables and the occurrence of *you know* is included in the present investigation. The background variables are the students’ sex and their grades in English in school. The distribution of the instances of *you know* in boys’ and girls’ production is given in Table 12.14.

*You know* is more frequent in girls than in boys’ production. The results are in line with Östman’s findings that *you know* is used more frequently by women than by men (Östman 1981: 72).

### Table 12.13. *You know* and the tasks of the tests by test rating categories

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1 (“to entertain”)</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Task 2 (“to inform”)</td>
<td>17 (57%)</td>
<td>21 (45%)</td>
<td>9 (53%)</td>
<td>47 (50%)</td>
</tr>
<tr>
<td>Task 3 (“to argue a case”)</td>
<td>9</td>
<td>19</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>Total instances</td>
<td>30 (100%)</td>
<td>47 (100%)</td>
<td>17 (100%)</td>
<td>94 (100%)</td>
</tr>
<tr>
<td>Number of oral tests</td>
<td>32</td>
<td>37</td>
<td>13</td>
<td>82</td>
</tr>
</tbody>
</table>

### Table 12.14. *You know* by the students’ sex

<table>
<thead>
<tr>
<th></th>
<th>Boys’ production</th>
<th>Girls’ production</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total instances (tokens)</td>
<td>28</td>
<td>66</td>
<td>94</td>
</tr>
<tr>
<td>Normalised figures</td>
<td>0.8</td>
<td>1.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Total number of students</td>
<td>12</td>
<td>17</td>
<td>29</td>
</tr>
</tbody>
</table>
The figures in Table 12.15. give grounds for the following observation: students with the top grade 5 in English in school tended not to use *you know* in the oral proficiency interviews.²

### 12.2.4. Summary

This section discussed the occurrence of *you know* in the student production. Surveys were provided regarding the distribution across the three test rating categories, occurrences in student production rated by the three interviewer categories, and in the three tasks of the tests. Two discoursal factors were taken into account, namely the discoursal function and the position of *you know* in the utterance. Whether *you know* was used together with certain other words was also investigated. Furthermore, whether the interviewers used *you know* in the tests when the students did so was discussed. Finally, the relationship to two background variables was presented.

The results showed that *you know* characterised student production rated Average and in particular, student production rated Low. The grades assigned by the three interviewer categories showed different tendencies, however. High frequencies emerged in student production rated Low by the Native Speakers. The grades by the University Teachers showed the opposite tendency: high frequencies appeared in student production rated High.

It was not possible to classify a majority of the instances into obvious discoursal functions. Some tendencies could be observed in the distribution of the three discoursal functions which were possible to identify, however. Slightly more instances in student production rated Average or Low were found in contexts of word searching, whereas the instances in student production rated High tended to cluster in contexts of expressing their own opinion. What is more, a higher proportion of the instances in production by students assigned a High grade was in tests in which the interviewers also used *you know*.

² The results cannot be explained by the fact that three of the four students with the top grade 5 in English at school were boys. The three instances attested spread in different tests.
A majority of the instances appeared in the narrative part of the tests and as for the relation to background variables, two tendencies could be observed: the girls were more inclined to use you know than the boys were, and students with the top grade in English in school tended to refrain from using you know in their production.
13.1. Intonation

As discussed in Chapter 9 (cf. 9.4.1. Intonation), one particular aspect of intonation will be dealt with in the present study, namely, the students’ rising intonation in declarative sentences.

Two issues should be mentioned in connection with the students’ intonation. Firstly, instances of rising intonation in declarative sentences occurred when the student wanted to have a confirmation by the interviewer of the correctness of a structure or a word (cf. 1.3. Linguistic Analyses of Students’ Proficiency). The following extract from test (H38) illustrates this phenomenon. Student H had a rising intonation on fence.

(112) Student: ...you see you don’t have the eh fence
   Interviewer: Fence around
   Student: around so
   Interviewer: and no trees

Secondly, the students came from the west coast of Sweden and had a Swedish accent characteristic of that area. A special feature of that accent is rising intonation in declarative sentences.\(^1\) The possible influence of the Swedish accent on the students’ intonation in English was not discussed by the interviewers, however.

13.1.1. Results from the Observations of Students’ Intonation

The production by nine students was selected for an investigation of their intonation in declarative sentences. The sample contains three oral tests assigned a High grade, three assigned an Average and three a Low grade.

\(^1\) Rising intonation in declarative sentences may occur in English in first-language countries. The number of possibly correct intonation patterns in English is quite large (Minugh 1991: 66).
Both boys’ and girls’ production was represented in this sample. Extracts of one minute were selected from Task 1 when the students were interviewed and assessed by Native Speakers. The same task with the same category of interviewers was chosen since different tasks or topics could have influenced the results. The material used in this investigation of the students’ intonation thus consists of nine minutes of the students’ spoken production. Apart from myself, the intonation in these extracts was analysed by two persons: Dr. Pia Norell, senior lecturer at the English Department at Uppsala University, and Dr. Francis Nolan, senior lecturer at the Department of Linguistics at Cambridge University. The two researchers consulted were instructed to spot instances of “unidiomatic use” of intonation in declarative sentences in the one-minute-long extracts collected, and furthermore, to comment generally on the students’ intonation. No information was given about the Global or Factorial grades assigned to the student performance.

The results of the investigation of the students’ intonation were as follows:

1. More instances of “unidiomatic use” of intonation, i.e. divergence from the falling intonation in declarative sentences, were spotted in student production assigned Average or Low grades than High grades. The former contains more than twice as many instances as the latter (Student production graded High 4 instances; student production graded Average 10 instances; student production graded Low 12 instances).

2. Production by five of the six students assigned Average or Low grades was given the following comments on unidiomatic rises in utterance-final words: “end rises” or “many final rises”. The intonation in student production assigned a High grade was described as varied and full of self-confidence whereas the intonation of student production rated Low was described as “monotonous, flat or unidiomatic”.

3. The observations by one of the two experts (Dr. Nolan) were as follows (my paraphrase):

The students’ intonation cannot be described in terms of correct and incorrect uses. Their rising intonation in a number of declarative statements is not in itself disturbing or divergent from native speakers’ intonation (particularly young speakers). Other features in the students’ pronunciation are of significance to the assessment. These features may be rhythm, stress or even the length of some types of phonemes before some consonants, such as the length of the first phoneme pronounced in easy.

13.1.2. Summary

Unidiomatic intonation in declarative sentences was identified in certain students’ performance. The frequency of the instances of unidiomatic into-
nation tended to tally with the Global grades assigned, i.e. unidiomatic intonation emerged in performance assigned Low or Average grades. Consequently, the students’ idiomatic or unidiomatic intonation may, at least partly, have been of importance for the assessment of their pronunciation and the Global grades assigned.

Nevertheless, it was evident that other factors than rising intonation in declarative sentences were important for the overall impression of the students’ pronunciation. These factors could have been variation in rhythm and stress, or the pronunciation of individual words.

By way of conclusion, further research is needed to analyse the assessment of advanced Swedish students’ intonation to see whether other factors are at work or whether intonation in itself is decisive for the grades assigned in assessments of oral proficiency.

13.2. Stress: Some Relevant Issues

13.2.1. Overall Distribution of Non-standard English Stress: Sentences and Words

In this section, the students’ stress on function words as opposed to lexical words will first be discussed. Then a closer investigation of their stress of and so on will be presented.

In order to investigate the students’ stress in the oral proficiency interviews, their tendency to stress function words in one of the tasks was investigated. In this investigation, the lexical words included nouns, lexical verbs, adjectives, and adverbs. The function words were articles, auxiliary verbs, conjunctions, prepositions and pronouns.

The sample included nine extracts of one minute from nine oral tests, all from the same task. There were three extracts from student production rated High, three from student production rated Average and three from student production rated Low. The extracts used were the same as the ones used in the investigation of the students’ intonation (cf. 13.1.1. Results from the Observations of Students’ Intonation).

The method used was purely auditory. The nine extracts were listened to and observations were made of the occurrence of “unidiomatic” stress of function words by the researcher himself. The results showed that instances of stressed function words were only traced in student production assigned Average or Low grades. There were more unidiomatic instances in student production rated Low than Average (seven instances in Low, four instances in Average). Pronouns and prepositions occurred in a great majority of these identified instances of unidiomatic stress.
In the second investigation of the students’ stress, the interest was on the students’ stress of *and so on*. First the overall frequencies are provided across the test rating categories, and then follows the presentation of the students’ stress of the instances attested.

Table 13.1. provides a survey of the distribution of the 69 instances of *and so on* which emerged in the material. The fact that the instances of *and so on* spread across the three test rating categories is useful in the investigation of stress; differences in quality may be expected. The figures for student production rated Low are highly influenced by the occurrence in one student’s test, however (ten instances in one student’s test). When these ten instances are left out, the normalised figure for student production rated Low is 0.9.

The 69 instances were categorised as “idiomatic stress”, “unidiomatic stress” or “indeterminate stress”. Instances of “idiomatic stress” are those where the students’ stress was on *so in and so on*. When the students tended to stress *on in and so on*, and thereby seemed to use the stress of the corresponding Swedish expression (Swedish = *och så vidare*), the instances were categorised as “unidiomatic stress”. Instances that could not be classified into these two categories were placed in “indeterminate stress”. The proportions of the three categories “idiomatic”, “unidiomatic” and “indeterminate” stress across the three test rating categories High, Average and Low are provided in Table 13.2.

The results show that a great majority of the instances (55 of the total of 69, 84%) were pronounced in an “unidiomatic” way. It is evident that most

<table>
<thead>
<tr>
<th>Number of <em>and so on</em></th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Average</td>
<td>Low</td>
</tr>
<tr>
<td>0</td>
<td>18 (56%)</td>
<td>27 (73%)</td>
<td>7 (54%)</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32 (100%)</td>
<td>37 (100%)</td>
<td>13 (100%)</td>
</tr>
<tr>
<td>n (instances)</td>
<td>32</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>m (instances)</td>
<td>1</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>s (instances)</td>
<td>1.4</td>
<td>1.1</td>
<td>7.9</td>
</tr>
<tr>
<td>Normalised figure</td>
<td>1.0</td>
<td>0.5</td>
<td>2.0</td>
</tr>
</tbody>
</table>
of the students had not adopted an English stress pattern of *and so on*. The instances of “idiomatic stress” cluster in student production assigned a High grade. In student production assigned a Low grade, and to a large extent in production assigned an Average grade, the students pronounced *and so on* in an unidiomatic way.

### 13.2.2. The Occurrence of Non-standard English Stress in Relation to Background Variables

The eleven instances of “idiomatic stress” of *and so on* occurred in production by four students. Three of the four students were girls. All four students had the top grades four or five in English in school. Although the raw figures of “idiomatic stress” of *and so on* are admittedly very low, it may nevertheless be observed that “idiomatic stress” was found in production by students with a high grade in English at school. Furthermore, most of these students were girls.

### 13.2.3. Summary

In this part the students’ stress, firstly in function words and secondly, with focus on the students’ pronunciation of *and so on*, was discussed. The three test rating categories High, Average and Low and the students’ sex and grades in English at school were taken into account. The observations of the students’ stress of lexical and function words revealed that instances of “unidiomatic” stress of function words were found in student production rated Low and Average. No instances were traced in student production rated High.

The investigation of the students’ pronunciation of *and so on* showed that a great majority of the instances were pronounced in an “unidiomatic” way and that almost all the “idiomatic” instances were found in student production rated High.

---

Table 13.2. *The stress of* and so on *by test rating categories*

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Average</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Idiomatic stress”</td>
<td>10 (31%)</td>
<td>1 (5%)</td>
<td>0</td>
<td>11 (16%)</td>
</tr>
<tr>
<td>“Unidiomatic stress”</td>
<td>22</td>
<td>15</td>
<td>18</td>
<td>55 (84%)</td>
</tr>
<tr>
<td>“Indeterminate stress”</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>32 (100%)</td>
<td>19 (100%)</td>
<td>18 (100%)</td>
<td>69 (100%)</td>
</tr>
</tbody>
</table>

Number of oral tests: 32, 37, 13, 82
13.3. Speech Rate

13.3.1. Overall Distribution

A survey of the students’ speech rate in the student production will now follow. This survey is provided by means of investigating samples of student performance and the numbers of “pruned words” occurring during the first minute of Task 1. “Pruned words” comprise all words pronounced by the students, except for self-corrections, repetitions, comments on the narrating itself and questions addressed to the interviewers (cf. 9.4.3. Speech rate). The results in Table 13.3. provide the distribution across test rating categories, means (m), standard deviations (s) and correlation (r).

The students’ speech rate varied from 42 to 155 words during the first minute of Task 1. The correlation between the students’ speech rates and the Global grades assigned is 0.29.

The results show two tendencies. Firstly, the great majority of the students used 71 to 120 words in the first minute of Task 1 (in 67 of the 82 oral tests, 82%). These tests are evenly spread across the three test rating categories. Secondly, high speech rate is related to High grades and low speech rate is related to Low grades. The students who spoke fast, i.e. used more

<table>
<thead>
<tr>
<th>“Pruned words”</th>
<th>Number of tests per test rating category</th>
<th>Grade means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Average Low n</td>
<td>m s</td>
</tr>
<tr>
<td>42–40</td>
<td>0 1 0 1</td>
<td>3.0 0</td>
</tr>
<tr>
<td>51–60</td>
<td>1 0 0 1</td>
<td>5.0 0</td>
</tr>
<tr>
<td>61–70</td>
<td>0 0 3 3</td>
<td>1.7 0.6</td>
</tr>
<tr>
<td>71–80</td>
<td>5 5 3 13</td>
<td>3.2 0.8</td>
</tr>
<tr>
<td>81–90</td>
<td>2 6 2 10</td>
<td>3.0 0.7</td>
</tr>
<tr>
<td>91–100</td>
<td>6 8 1 15</td>
<td>3.4 0.7</td>
</tr>
<tr>
<td>101–110</td>
<td>4 9 1 14</td>
<td>3.3 0.6</td>
</tr>
<tr>
<td>111–120</td>
<td>5 7 3 15</td>
<td>3.1 0.7</td>
</tr>
<tr>
<td>121–130</td>
<td>3 1 0 4</td>
<td>4.3 0.8</td>
</tr>
<tr>
<td>131–140</td>
<td>2 0 0 2</td>
<td>4.0 0</td>
</tr>
<tr>
<td>141–150</td>
<td>2 0 0 2</td>
<td>4.0 0</td>
</tr>
<tr>
<td>151–155</td>
<td>1 0 0 1</td>
<td>4.0 0</td>
</tr>
<tr>
<td>n (oral tests)</td>
<td>32 37 13 82</td>
<td>r</td>
</tr>
<tr>
<td>m</td>
<td>107 96 88 99</td>
<td>(correlation between Global grades and speech rate) = 0.29**</td>
</tr>
<tr>
<td>s</td>
<td>23.4 16.9 17.7 20.7</td>
<td></td>
</tr>
</tbody>
</table>

Table 13.3. “Pruned words” per minute by test rating categories
than 120 words, were assigned a High grade (one student was assigned an Average grade). Three of the five students whose speech rate can be labelled as slow, i.e. between 42 and 70 words, were assigned Low grades.

### 13.3.2. Speech Rate in Relation to Interviewer Categories and Background Variables

The background and experiences of the interviewers could account for differences in the Global grades assigned to student production that contains samples of high (i.e. students who used more than 120 words the first minute) or low speech rate (i.e. students who used fewer than 70 words in the first minute). For that reason, it is of interest to survey the figures of the students’ speech rate in relation to the Global grades assigned by the members of the three interviewer categories.

Table 13.4 shows high average scores of student production rated High irrespective of interviewer category. The highest average scores and the highest individual scores attested are in student production rated High. The students who spoke fast in Task 1 tended to be assigned a High grade irrespective of interviewer category.

The presentation of the relationship between speech rate and the two background variables may provide useful information. The background variables are the students’ grades in English in school and the students’ sex. Table 13.5 shows the scores of pruned words per minute in comparison with the students’ grades in English (grades 1 to 5, 5 being the top grade).

The results show the tendency that students who had high grades in English in school spoke faster in the samples than those who had average or low grades. The investigation of the second background variable showed that there was no relationship between the students’ sex and their speech rate: the average score in the samples of the boys’ production was 101 words, and in the girls’ production the average score was 98 words.

### Table 13.4. Average scores of pruned words per minute by interviewer categories

(ST = School Teachers; UT = University Teachers; NS = Native Speakers)

<table>
<thead>
<tr>
<th></th>
<th>ST</th>
<th></th>
<th></th>
<th>UT</th>
<th></th>
<th></th>
<th>NS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Av</td>
<td>Low</td>
<td></td>
<td>High Av</td>
<td>Low</td>
<td></td>
<td>High Av</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Average score</td>
<td>112 97 92</td>
<td>94 91 89</td>
<td>116 99 78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest score</td>
<td>155 117 116</td>
<td>139 127 111</td>
<td>141 118 86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest score</td>
<td>74 76 66</td>
<td>57 42 70</td>
<td>96 71 70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of oral tests</td>
<td>10 13 6</td>
<td>12 8 4</td>
<td>10 16 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
13.3.3. Summary

In this section a survey of the students’ speech rate was provided. The method used was adopted from a study with German students of English (cf. 1.3. Linguistic Analyses of Students’ Proficiency, and Lennon 1990).

The results showed that students who spoke with a high speech rate (more than 120 words) during the first minute in Task 1 tended to be assigned High grades. Another tendency could be observed: students who had low speech rate in their production were assigned Low grades. This relationship between speech rate and the Global grades assigned was evident in the assessments by all interviewers, irrespective of interviewer category.

There was a relationship between speech rate and the students’ grades in English in school: students with high grades in English at school spoke faster than the other students did.

Table 13.5. “Pruned words” per minute by school grades in English

<table>
<thead>
<tr>
<th>Grades</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average scores</td>
<td>125</td>
<td>100</td>
<td>94</td>
<td>89</td>
<td>99</td>
</tr>
<tr>
<td>Highest scores</td>
<td>155</td>
<td>127</td>
<td>139</td>
<td>120</td>
<td>–</td>
</tr>
<tr>
<td>Lowest scores</td>
<td>91</td>
<td>57</td>
<td>70</td>
<td>42</td>
<td>–</td>
</tr>
<tr>
<td>Total number of students</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>3</td>
<td>29</td>
</tr>
</tbody>
</table>
Part 6:
Discussion

The purpose of Part 6 is to interpret the results of the project presented in Part 5. In Chapter 14, the findings and observations regarding the grades assigned will be discussed. In Chapter 15, the students’ linguistic competence as it was demonstrated in the oral tests will be considered.
In this chapter the results presented in Chapter 8 (Grades Awarded to the Students’ Oral Proficiency) are interpreted. Firstly, the students’ Global grades are discussed. Secondly, the distribution of the three grades for Communication and Fluency, Vocabulary and Grammar, and Pronunciation is analysed. Thirdly, the distribution of the grades assigned by the three interviewer categories is considered. Then, the assessment of the students whose production was assigned different grades by different assessors is discussed, and finally, the relationship between the background variables and the Global grades is expounded.

14.1. Global Performance Grades

Most of the students were assigned the Global grades B or C (cf. Table 8.1.). The interviewers’ overall impression was thus that the students were good or very good at speaking English. To a certain extent, the grades confirm the spontaneous comments the interviewers gave when they filled in the questionnaires immediately after the testing session: the student performance was considered good (cf. Chapter 8). More oral tests were graded according to the rating criteria of the upper than of the lower half of the scale. Using the rating criteria, the distribution of the students’ Global grades can be described in the following way: the student performance was acceptable or very good with minor inaccuracies, and a few students even showed fluent and overall correct use of English. Some students’ performance was often inaccurate or inappropriate but hardly any could be described as poor in many respects.

A great majority of the students were assigned different Global grades for their production in the three oral tests assessed by the three assessors (cf. Table 8.8.). Since the assessments of their performance were carried out by very qualified interviewers (cf. 7.3. Interviewers), this finding is notable.
and of great interest. The fact that four students were assigned three different Global grades on the five-step scale for their production and that 17 of the 29 students were assigned two different Global grades indicates how difficult it is to assess students’ oral performance in a reliable way when the assessors use a carefully worked out set of rating criteria (cf. Table 8.9.). The grades assigned by the three interviewer categories show no differences in this respect.

Several factors may explain these findings, such as the interviewers’ different interpretation of the rating criteria or the influence of certain linguistic factors on the assessment of the student production in the three tests. The results also show that the length of the performance and the order in which the oral proficiency interviews were carried out, at least for some of the students, may explain the grades assigned. However, it is obvious that the interviewers of the Gothenburg project assessed the students’ oral proficiency very differently and therefore gave the students remarkably different grades for their performance on the same day. The differences in the grades awarded will be discussed at some length as regards the interviewer categories (14.3. Interviewer Categories) and the grades assigned to the individual student’s production (14.4. Individual Student’s Grades). But even at this stage, it is evident that measures need to be taken, for instance in the form of teacher in-service training, to achieve acceptable reliability in grading students’ oral performance (see Reflections).

Eight of the 29 students were assigned the same Global grade by their assessors for their production in their three tests. These students’ performance was in the middle or near the upper half of the scale. This could be interpreted in two ways. The first interpretation is that the students whose proficiency in English could be characterised as being of average or medium standard or just above are more easily graded in a reliable way than other students. Their performance contains characteristics which the interviewers could match with the rating criteria “Very good production with minor inaccuracies” or “Acceptable language in spite of errors”. Another interpretation is that the students were assigned the same grades because the interviewers had a tendency to go for the grades of the middle of the grading scale (cf. 5.3. Sources of Variation). The interviewers tended to use the grades which were in the centre of the scale even though the student performance might have been closer to the rating criteria of other grades. It is not unreasonable to believe that both these explanations were of significance when all three interviewers gave the same grade.
14.2. Factorial Grades

The results of the three Factorial grades showed the same tendency as did the Global grades: many students were assigned different grades in the three assessments of the three tests. This would imply that even when the interviewers used the rating criteria of one particular aspect of the student performance, such as Grammar and Vocabulary, it could nevertheless be difficult for them to grade student production in a reliable way.

When the students were assigned the same grades for Communication and Fluency, or Grammar and Vocabulary by the three assessors, they were graded in the middle or just above the middle of the rating scale. (Eight students were assigned the same grade for Communication and Fluency, seven students were assigned the same grade for Grammar and Vocabulary, and eight students were assigned the same grade for Pronunciation, cf. Table 8.8.). This tendency is in line with the results obtained for the Global grades. When identical grades were given for Pronunciation, the student production was graded in the middle of the scale (grade C/3).

When the Factorial grades are studied, it can be noticed that the students’ strength lies in Communication and Fluency, and their weakness in Grammar and Vocabulary. Generally, the students were thus good at communicating what they wanted to say, but lacked vocabulary and made grammatical mistakes; some of the students even made basic errors, as the linguistic analyses of the material show. The grades assigned for Pronunciation show that the interviewers tended to find it difficult to make clear distinctions between different students’ pronunciation and thus went for the grades in the middle of the scale. One interpretation is that the students’ pronunciation was of a similar kind and that the Swedish school-leavers who participated in the Gothenburg project had all acquired good pronunciation but still with certain typically Swedish features which made the interviewers assign grades of the middle of the scale. Another interpretation could be that they came from the same region in Sweden (the Gothenburg area) and therefore were influenced by the same kind of interference from their Swedish accent. On the whole, it seems that the students were regarded to have acceptable intonation and rhythm, although their pronunciation showed evident traces of their Swedish accent.

As stated above, most students tended to be given different grades for their production in the three tests. The differences in the students’ Factorial grades were as follows. For Communication and Fluency, 5 students were assigned three different grades, and 16 students were assigned two different grades. For the second Factorial grade, Grammar and Vocabulary, 3 students were assigned three different grades, and 19 students were assigned two different grades. Finally, for Pronunciation, 6 students were assigned
three different grades, and 15 students were assigned two different grades (cf. Table 8.9.) Some tendencies in these differences in grades assigned could be identified. The degree of divergence attested in the individual student’s three tests suggests that the interviewers differed to a greater extent in their opinion of the students’ skills in Communication and Fluency, as well as in Pronunciation than of the resources the students had in Grammar and Vocabulary. The results give grounds for two interpretations. Firstly, it may be easier to assess students’ command of grammar and vocabulary than their pronunciation or their skill to communicate with fluency according to the descriptions in the rating criteria. This would imply that pronunciation, and communication and fluency are more difficult to assess reliably with a set of rating criteria than the students’ mastery of vocabulary and grammar. It may be easy to identify non-standard English verb forms or inappropriate uses of vocabulary but more demanding to decide on a grade for unidiomatic intonation or the lack of fluency in speech, in particular when the interviewers are requested to do so while the conversation is going on or shortly after it. Secondly, the individual student’s grammatical accuracy and use of vocabulary may be of a similar kind in his/her three tests, and this skill is thus assigned similar grades by the three interviewers, but the student’s ways of communicating and showing fluency may depend, for instance, on the interlocutor or on the task. This would explain why the grades assigned to Communication and Fluency are different in the student’s three tests. The differences in the grades assigned to Pronunciation might also be due to real differences in the performance in the individual student’s three tests with the three different interviewers.

By and large, on the basis of the results of the Gothenburg project, the first interpretation, i.e. that vocabulary and grammar are more easily assessed than communication and fluency, and pronunciation, seems the most likely.

14.3. Interviewer Categories

When the grades assigned by the members of the three interviewer categories are compared, three tendencies may be identified: relatively higher grades were given by the Native Speakers, relatively lower grades were given by the School Teachers, and regarding the grades for Pronunciation, the Native Speakers tended to assign the grades in the middle of the scale to the student production.

The grades awarded show that the Native Speakers generally had a somewhat more positive view of the students’ performance. The reasons for this may be that the Native Speakers interpreted the rating criteria so that higher
grades more often seemed to match the students’ performance. Higher
grades can be observed, in particular, for Communication and Fluency.
When all three interviewers gave different grades to a student, the Native
Speaker tended to be the one who assigned the highest grade. Another inter-
pretation is that the Native Speakers focused on “good” aspects in the stu-
dents’ performance such as the ability “to communicate and be fluent” and
not on their lack of vocabulary or the occurrence of grammatical errors. In
this way, the student production was given higher grades by the Native
Speakers than by the other interviewers.

The School Teachers gave lower grades than did the other two inter-
viewer categories. The School Teachers had the lowest means in their Glo-
bal grades, in their grades for Communication and Fluency, and in their
grades for Grammar and Vocabulary. When there was one deviant grade
given to a student’s production, the tendency was that the School Teacher
had assigned the low grade. This might be due to the possibility that the
School Teachers observed errors in the students’ uses of structures and vo-
cabulary and paid attention to (and let this influence the grades they gave)
grammatical errors since they recognised them from teaching situations at
school. This finding is in line with the discussion in Ellis (1994) (cf. 2.3.
Reliability). It implies that the School Teachers tended to focus more on
accuracy than on fluency, and thereby the students’ weaknesses influenced
their Global grades. Thus the School Teachers’ low grades for Grammar
and Vocabulary influenced their Global grades and their grades for Commu-
nication and Fluency (cf. the halo-effect in 5.3. Sources of Variation). The
School Teachers’ and the Native Speakers’ grades for Grammar and Vocab-
ulary tended to correlate more with the Global grades assigned than did the
other two Factorial grades. This tendency was not strong but is nevertheless
of interest (cf. the correlations in Tables 8.7.).

The Native Speakers gave the same grades to many students’ pronuncia-
tion. This could mean that the students showed less variation in their mas-
tery of English pronunciation than, for instance, in their grammatical accu-
ragacy. The three Native Speakers thus noticed little diversity in the Swedish
students’ pronunciation of English: the students’ pronunciation, intonation
and stress patterns showed interference from Swedish, and perhaps from
the Gothenburg area, in particular, which made the interviewers assign simi-
lar grades to their production. The fact that the grades for Pronunciation
tended to vary very little especially when assigned by the Native Speakers
gives support to this interpretation. This was the most evident tendency
among the interviewers to assign grades in the middle of the rating scale
without noticing differences among the students.

On the whole, regarding the interviewer categories, the results show that
the Native Speakers both were more appreciated as interviewers by the
students (cf. 8.9. Students’ Opinions about the Test), and to a certain extent, assessed the students differently than did the other interviewers in all respects except accuracy. The Native Speakers seemed to be more sensitive to pronunciation (and perhaps also better judges of that factor) and to the students’ vocabulary, and indifferent, or even negative, to the students’ use of the two discourse markers (well and you know). Furthermore, it may be assumed that speaking to a native speaker brings out the best in a non-native speaker because of the authentic feedback.

**Individual Interviewers**

School Teacher 2 was the interviewer who assigned the lowest grades. This may be because he, more than the other interviewers, focused on grammatical and lexical accuracy and inaccuracy and thereby assigned low grades to the student production.

Native Speaker 2 was the interviewer who assigned the highest number of identical grades to the student production when the grades are compared with the grades assigned by the other interviewers to the same student (cf. 8.5. Comparative Discussion of the Four Grades Given). This interviewer was a very experienced assessor of English, who had worked as an examiner for the Cambridge exams in Sweden for many years. It may be assumed that her background and skills contributed to grades that could be described as fair and reliable because of the fact that many of the students which she assessed were given the same grades for their production by the two other interviewers.

School Teacher 3 also gave the same Global grades and grades for Grammar and Vocabulary as did the other interviewers of the same students (cf. 8.5.). The fact that this interviewer knew seven of the ten students she interviewed from school may have contributed to grades that could be described as more reliable than most of the other ones; before the oral proficiency interviews started, this interviewer was aware of seven students’ real level of oral proficiency.\(^1\)

### 14.4. Individual Students’ Grades

As described in Chapter 8, twelve students were assigned different grades (Global and/or Factorial grades, or both) for their production in the three oral proficiency interviews. These individual students’ performance provides a valuable source for further analyses. In what follows, the grades and the performance of these students are discussed. The twelve students were

\(^1\) School Teacher 3 answered “yes” to the question “Do you know the candidate?” regarding seven of the ten students she interviewed.
assigned remarkably different grades by the three interviewers for their performance in the three tests (three different grades assigned, for instance 5 – 4 – 3, or two very divergent, 5 – 3 – 3, cf. Tables 8.1. to 8.4. in Chapter 8). The students of interest are students C, F, J, K, L, O, Q, R, S, T, XX and ZZ. Table 14.1. presents the grades of these twelve students which are discussed in this chapter. The grades to be discussed are in bold. (Table 14.1. shows the grades assigned to the student performance but not the order in which the interviews were carried out.)

The differences in the grades assigned to the students’ production in the three tests by the three interviewers might be explained by factual differences in the individual student’s three performances during the different interviews. This would imply that, for instance, a student was given the opportunity to show more of his/her competence in one of the tests than in the other two and was therefore given different grades. The interviewer’s methods or the order of the three interviews could, at least partly, explain some of the different grades attested. In the discussion below, the different grades are analysed with respect to the students’ linguistic and strategic competence, the interviewers’ methods, and the length as well as the order of the interviews.2 (The following abbreviations will be used in the discussion below: ST = School Teacher, UT = University Teacher, NS = Native Speaker. Example: UT 2 = University Teacher 2.)

2 For clarity, the grades in the discussion are described in the following way: A=5; B=4; C=3; D=2 and E=1.
**Student C**

Student C was assigned three different Global grades: 4 by ST 1, 2 by UT 3, and 3 by NS 2. The three grades for Pronunciation differ in the same way. Student C had his first interview with UT 3 and said in the interview that he was very nervous. His state of mind may have influenced his performance and resulted in his low grade. In the interview with NS 2, he spoke at length. This interview has the most student words in comparison with all other tests (2,010 words; the average: 948 words/interview). One interpretation that seems reasonable is that the communication worked well together with NS 2 and the student was given the opportunity to express himself freely. The student’s strength seems to lie in communication and fluency but not in accuracy. Two interviewers (UT 3 and NS 2) gave him the highest grades for Communication and Fluency in relation to the Global grades and the two other grades assigned by them. When the student was given the opportunity to speak at length, this may have influenced the grades. By way of conclusion, the order in which the interviews were made and the fact that one interviewer elicited student performance in ample quantities seem to account for the different grades assigned.

**Student F**

Student F was assigned three different grades for Communication and Fluency: 2 by ST 1, 3 by UT 3, and 4 by NS 3. One explanatory factor, as was the case with student C, could be the order in which the interviews were done. The lowest grade was assigned to the first interview, the highest to the last interview. Student F may have felt more self-confident in the last than in the first interview of the day. The length of the interviews and the number of words provide evidence of this hypothesis (560 student words in the first production, 1,060 and 1,010 student words in the second and third production; nine minutes in the first production, approximately 12 and 13 minutes in the second and third production).

**Student J**

Student J was assigned three different grades for Pronunciation: 5 by ST 1, 4 by UT 1, and 3 by NS 3. The other grades do not show similar differences. Again, the explanatory factor can be the order in which the interviews were made. The lowest grade was assigned by NS 3 in the student’s first production, and the highest grade by ST 1 in the student’s last production. Furthermore, student J spoke more in the last production compared with the other two (1,100 student words compared with 780 and 860 student words). Her speech rate is slower than that of many other students, and her performance contains few grammatical errors and many instances of the discourse markers *well*. The three different grades assigned could also be evidence of the
three interviewers’ different impression of the student’s pronunciation. Low speech rate together with other aspects of pronunciation, such as rhythm and stress (which the Native Speakers were more sensitive to in their assessment), could have been decisive factors for the different grades assigned. It may be concluded that the order of the interviews, the fact that the student was not very talkative and spoke slowly in her first production, and the interviewers’ different impression of the student’s pronunciation seem to explain the different grades awarded.

*Student K*

Student K was assigned three different grades for Grammar and Vocabulary: 2 by ST 2, 4 by UT 2 and 3 by NS 1. Again, the order in which the interviews were carried out shows the same pattern: the lowest grade was assigned to the student’s first production and the highest grade was assigned to the student’s last production. The production assessed by UT 2 cannot be taken into account since it was not recorded, so only the student’s tests graded 2 and 3 can be compared. The production graded 2 contains a significantly higher proportion of non-standard English instances of the third person singular marker. This may have influenced the School Teacher’s lower grade (15 non-standard English instances / 7 standard English instances compared with 7 / 6). The production graded 3 generally comprises more speech and contains more student words than the production graded 2 (1,165 student words and eleven minutes compared with 970 student words and approximately ten minutes). The production also contains few abstract nouns in comparison with other student production. This would imply that the student did not discuss abstract or controversial issues very much, in particular in task 3. The factors explaining the different grades seem to be the order of the interviews, a somewhat hesitant performance and the use of non-standard English verb forms in one of the tests.

*Student L*

Student L was assigned three different grades for the Global evaluation and for Pronunciation. The grades, both Global and for Pronunciation, were as follows: 2 by ST 2, 4 by UT 3 and 3 by NS 2. The lowest grade was assigned to the first production and the highest grade to the second production. The lengths of the interviews are from approximately ten to thirteen minutes. The shortest interview was assigned the lowest grade. The student’s performance can generally be described as hesitant, in particular when she is interviewed by UT 3 (although the highest grade was assigned in this interview). This student’s production contains few student words

---

1 The interview assessed by UT 2 was one of the five interviews which were not recorded due to technical problems (cf. 7.1.).
Fillers such as *I think so* and *or something* tend to occur in her performance as well as the discourse marker *you know*. The student’s use of fillers and *you know* may have had different effects on the three interviewers. When she was interviewed by NS 2, she tended to speak faster. The impression is that she had the courage to say more in the last production assessed by NS 2. It seems that the explanation of the different grades can be found, at least partly, in the order of the interviews. Furthermore, the student’s hesitant delivery and use of certain hesitation devices could have influenced the interviewers differently, both regarding the Global evaluation and the grades for Pronunciation.

**Student O**

Student O was assigned three different grades for Pronunciation: 2 by ST 2, 4 by UT 3, and 3 by NS 3. The three interviewers did not assign the same grade (the Global or the two remaining Factorial grades) to his production. Higher grades were given for Communication and Fluency than for Grammar and Vocabulary or Pronunciation by all three interviewers. The lowest grade for Pronunciation was given in the first interview (ST 2), the highest grade in the last interview (UT 3). This interview (assessed by UT 3) is also the longest (approximately 16 minutes) and contains the most student words (1,330). No great differences in linguistic features can be traced in his production in the three tests. Evidently the student had his strength in communication and fluency in the interviews. It seems that the length of the production and the order of the interviews explain the different grades.

**Student Q**

Student Q was assigned three different grades for Communication and Fluency, and two, but one grade apart, for Pronunciation. The grades for Communication and Fluency were as follows: 2 by ST 2, 3 by UT 2 and 4 by NS 2. The student’s grades for Pronunciation were the following: 5 by ST 2, and 3 by UT 2 and NS 2. Student Q was the only student who was given two different grades but two steps separated for Pronunciation (cf. Table 8.4.). It is worth observing the different grades assigned by School Teacher 1: 2 for Communication and Fluency, 5 for Pronunciation. Student Q is the only one out of the 29 participating students who had American pronunciation. His pronunciation is discussed in one of his oral tests:

“I asked you about this partly because when you came, when you started talking, I of course at once spotted your American accent because you have an American accent.” (University Teacher 2)

One possible explanation to this student’s different grades for Pronunciation could be the interviewers’ different impressions of the student’s Amer-
ican pronunciation. The rating criteria of the grades assigned indicate that ST 2 considered that the student had good intonation and rhythm (grade 5) whereas the other two interviewers thought there was an evident Swedish accent in his performance (grade 3). The student’s American pronunciation may have caused the different grades. Furthermore, Student Q frequently used various expressions to give him time to plan his speech, such as *I don’t really know* and *I mean*. One possible explanation is that his obvious American pronunciation together with his frequent use of certain expressions may have had different effects on the interviewers. His production may have been interpreted as being of poor quality and perhaps somewhat too colloquial, and that he had difficulties in expressing what he wanted to say, particularly when assessed by the School Teacher. The different impressions of his production led to the different grades both for Communication and Fluency.

**Student R**

Student R was assigned two different grades (one grade level apart) for the Global evaluation and for Communication and Fluency: 2 by ST 1, and 4 by UT 1 and NS 1. The high grades were given to the performance of the first and last oral tests. The tests are of similar lengths (between ten and eleven minutes long), but contain fewer student words than did other oral tests (790, 680 and 640 student words). She has a speech rate during the first minute of the tests which is close to the average. The differences in the grades assigned cannot be explained in terms of standard English or non-standard English instances. The performance could be described as careful and hesitant with few discourse markers and no meta-talk. No obvious explanation can be found of the School Teacher’s low grade. The School Teacher did not know her from school. In the comment on the ratings assigned added on the Oral Mark Sheet, the School Teacher discusses the fact that the student has very little to say about the topics of the tasks. One possible explanation to the different grades assigned to this student could be this student’s limited contributions in the interviews. Her production was very short and therefore the interviewers had different views about her level of proficiency.

**Student S**

Student S was assigned two different Global grades but two grade levels apart: 4 by UT 1, and 2 by ST 2 and NS 2. Three different grades were given for Communication and Fluency: 4, 3 and 2. The high Global grade was given in the first interview. The production contains few student words (620, 730 and 620 student words) and the tests which were assessed by ST 2 and UT 1 is shorter than the one assessed by NS 2 (the former approxi-
mately nine minutes long, the latter more than twelve minutes long). The student had a very good start with UT 1. She spoke fast during the first minute of that production. Her speech rate could have influenced the overall impression and the high grade assigned. Differences in grades could thus be explained by the successful performance in one of the tests and the fact that the grades were based on short student production.

**Student T**

Student T was assigned three different grades for Communication and Fluency: 5 by ST 2, 3 by UT 2 and 4 by NS 3. The highest grade was given in the last production. For Pronunciation he was also given three different grades: 2 by the ST 2, 4 by the UT 2 and 3 by the NS 3. ST 2 also gave a lower grade for Grammar and Vocabulary (grade 2) than did the other two interviewers (grade 4). We may observe the great difference in grades assigned by ST 2 (the grades 5 and 2). The other two interviewers gave this student the grades 4 or 3. The student’s performance is long and contains many student words (between 12 and 15 minutes long, and between 1,100 and 1,400 student words). Evidently the student had no difficulty communicating, but his performance was interpreted differently with regard to the rating criteria. The student’s performance contains many instances of metatalk. The following phrases from his performance are given for illustration:

- as I told you,
- as I said,
- how to say it,
- I don’t know the English word,
- I don’t know what it’s called in English,
- I’m going to start to describe …,
- I’m going to talk about…,
- so to say,
- that’s what you can say about …,
- the word is …,
- you can say

The student’s performance was commented on by the interviewers as follows:

“Eleven har talets gåva men språket och uttalet har många brister”. (The student has the gift of the gab but his language and his pronunciation have many deficiencies.) (School Teacher 2)

“Somewhat stiff” (University Teacher 2)

“…he used body language to suggest effectively ‘We shouldn’t just sit back with arms folded …’” (Native Speaker 3)

The student’s pronunciation of and so on is unidiomatic (four instances in the production assessed by the School Teacher, and two instances in the production assessed by the University Teacher). It seems that the order of the interviews and possible interference from Swedish (his production contains several examples of lexical and grammatical inaccuracies) led to the different grades assigned. The student was good at fluency, but his pronunciation and choice of vocabulary showed interference from Swedish. Evidence of this interference may be observed in high frequencies of the lexi-
cal errors investigated. A few non-standard English instances emerged together with many standard English ones of the third person singular marker and the progressive. It seems that ST 2 thought that these lexical errors together with certain grammatical errors were disturbing features in his production, whereas the two other interviewers did not think so. The interference from Swedish in his production may also have contributed to the interviewers’ different grades for Pronunciation.

**Student XX**

Student XX was assigned three different grades for the Global evaluation, for Grammar and Vocabulary, and for Pronunciation: 3 by ST 3, 4 by UT 3 and 2 by NS 3. Each interviewer thus assigned the same grade for the Global evaluation, for Grammar and Vocabulary, and for Pronunciation (cf. Table 14.1.). The lowest grades were given for the first production (NS 3) and the highest grades for the student’s last production (UT 3). The oral proficiency interviews are of the same length but the student spoke remarkably more in the interview assessed by School Teacher 3 (930 student words in comparison with 670 and 570 words). One explanation could be that ST 3 knew Student XX before they met in the interview; the student felt more self-confident together with an interviewer from her own school (cf. 14.3. Interviewer Categories). There are many instances of you know in her performance, which may have had different effects on the interviewers’ grades (cf. 12.2. You know…). The order of the interviews may also have influenced the student’s performance and thereby the different grades assigned.

**Student ZZ**

Student ZZ was assigned three different grades for the Global evaluation, Communication and Fluency, and Grammar and Vocabulary: 2 by ST 3, 3 by UT 2 and 1 by NS 3. Each interviewer consistently assigned the same grade for the Global evaluation, the two factorial ratings Communication and Fluency, and Grammar and Vocabulary. The lowest grade was given to the first production (NS 3) and the highest grade to the last production (UT 2). The production assessed by UT 2 is almost twice as long as the production of the other two tests (i.e. more than 20 minutes long). This interviewer seems to have liked the conversation (the interviewer laughs several times), and even though the student’s production is by far the longest one, there are not more than 1,060 student words. This is due to the fact that the student did not speak very much. In the production assessed by the Native Speaker, the student had certain difficulties expressing what he wanted to say. This interview is the only one in which the interviewer interrupts the test after Task 2, and throughout the interview the interviewer asks the student many questions, which the student answers very briefly. Student ZZ has overall
low figures for speech rate, grammatical correctness and range of vocabulary (for instance, abstract nouns). His production contains high frequencies of the lexical errors investigated and generally a higher proportion of non-standard English than standard English instances of the third person singular marker and the progressive. Student ZZ was generally assigned low grades for his production. No other production in the material was given grade 1 (assigned by the Native Speaker). His production contains many examples of incorrect and unidiomatic use of English. Obviously the Native Speaker, for some reasons, assessed these errors to be more serious for the overall impression of his performance than did the other two interviewers. Perhaps the order of the interviews, one interviewer’s technique, the student’s lack of initiative in one interview, and different views on the seriousness of lexical and grammatical errors may have contributed to the different grades.

14.5. Results in Relation to Background Variables

As already accounted for, the material of the Gothenburg project provided information regarding background variables, namely, the students’ sex, age, school grades in English and results from the National Test of English. The comparison between the grades assigned and these parameters showed that the top Global grade 5 were more often given to girls than to boys (cf. Table 8.10.). However, there is no similar tendency between the other Global grades assigned (4, 3, 2 and 1) and the students’ sex. This indicates that there were more very proficient girls than boys who participated in the Gothenburg project. Age did not seem to be of importance for a higher or lower Global grade (cf. Table 8.11.).

The relationship between the students’ school grades in English given at the end of the autumn term before the study was made and the grades assigned to their production was of interest in many respects. The Global grades assigned to the production of students who had the two top school grades (i.e. 4 or 5) did not show any major differences, proportionally speaking. Consequently, the performance of students who had the top grade 5 was not assigned remarkably higher grades than the performance of those who had grade 4 at school. This may be an indication that the teachers – of course – had taken other factors into consideration when they gave these students their school grades in English. These factors were, for instance, reading, writing, and the adequate use of grammar and vocabulary in writing. Nor did the grades of the production of students who had the low school grade 2 differ, proportionally speaking, from the grades assigned to the production of the students who had the average school grade 3. This
may perhaps indicate that low-achievers have their strength in oral proficiency but difficulties with other skills, such as writing or reading (cf. Table 8.13.).

The comparison between the grades assigned and the students’ results in the National Test of English given a little more than a year prior to the oral proficiency testing day supports the finding described above; students who had low or fairly low grades on the National Test tended to do so well in the oral tests of the Gothenburg project that they were given higher grades for these achievements (cf. Table 8.14.). This could be interpreted in the following way: low-achievers are fairly good at communication in oral interviews but often have difficulties with reading or writing. The students’ results in the National Tests indicated that these students were rather weak and had obvious gaps in other skills needed to master the English language, whereas in the oral test of the Gothenburg project they were given the opportunity to show oral skills, something which was not included in the National Test of 1992. These oral skills are aspects of interaction and communication which the students are not given the chance to show their command of in tests other than those similar to the ones used in the Gothenburg project.

14.6. Concluding Remarks

As a whole, the interviewers assigned generally positive (i.e. on the upper half of the rating scale), but quite often differing grades to the students who took part in the Gothenburg project. The students were often assigned different grades for the Global evaluation of their three oral tests. This tendency could also be observed in the three separate grades awarded to Communication and Fluency, Grammar and Vocabulary, and Pronunciation.

The interviewers seem to have appreciated the students’ ability to communicate and be fluent but, at least in the case of some of the students, also were influenced by their lack of vocabulary and the occurrence of grammatical errors in their performance. Many students were considered to have acceptable intonation and rhythm, but still an evident Swedish accent. All three interviewers gave the same grades particularly when students were awarded grades in the middle of the scale. The grades in the middle of the scale were also given to many students for their pronunciation, in particular when the performance was assessed by Native Speakers.

As regards the interviewer categories, the Native Speakers assigned higher grades to the student production than did the other interviewers, while on an average the School Teachers assigned the lowest grades.

Differences in grades assigned within the interviewer categories could be
observed. The individual interviewer’s grades of the individual student’s performance were of interest and the individual students’ production in the three oral tests which were assigned highly different grades was worthwhile analysing. These analyses form the basis of the following four suggested interpretations of the differences observed in the grades assigned.

– The order in which the interviews were done seems to have influenced the student performance and thus the grades assigned so that it, in most cases (but not all), became better and better.
– Student production which contained few words, due to factors such as students’ hesitant delivery and slow speech, seem to have been difficult for the interviewers to match with the rating criteria.
– The students’ uses of various fillers seem to have worked in different directions, i.e. influenced the assessment – and thus also the grades assigned – both some positively and some negatively.
– Whether the communication between the interviewer and the student worked well or not may have influenced the grades awarded. In other words, the method used by the interviewer to make the student feel confident might be of importance for the grading.

The comparison between the students’ school results (school grades and results in the National Test of English) and the grades awarded in the oral test showed that it was obvious that oral proficiency interviews give valuable additional information on the students’ communicative skills, skills which were not tested in the National Test of English in 1992. This was particularly noticeable for students with low grades in the National Test.
In this chapter, the results presented in Chapters 9 to 13 will be interpreted regarding the students’ use of certain linguistic features. The linguistic features were selected to cover four aspects of student production, namely morphology and syntax, vocabulary, discourse phenomena and aspects of pronunciation (cf. 6.2. Linguistic Features). The present chapter is structured with regard to these four aspects. Firstly, the students’ use of the verbal group will be discussed (the third person singular marker and the progressive). Secondly, certain aspects of the students’ vocabulary will be considered (abstract nouns, modifiers of adjectives, lexical errors and compensatory strategies). Thirdly, the occurrence of two discourse markers in the student production will be commented on (well and you know). Fourthly, the results from the investigations of the students’ pronunciation will be elaborated on (intonation, stress and speech rate). Finally, the interference from Swedish in student production will be discussed.

15.1. Verbal Group

In this section the students’ mastery of the verbal group in English will be discussed, first regarding their use of the third person singular marker and then considering their use of the progressive.

The Third Person Singular Marker

The third person singular marker occurred frequently in all student production. The classification into standard English and non-standard English instances in relation to the grades assigned showed that it was a matter of proportions: higher proportions of standard English instances in student production yielded a higher grade (cf. Table 10.1. and 10.2.). A high proportion of non-standard English instances led to low grades whereas single non-standard English instances were not of major significance for the
grades assigned. The most common type of the non-standard English instances was Type 1 (*he take) (cf. Table 10.3., and the discussion of the redundancy of the third person singular marker in 1.3. Linguistic Analyses of Students’ Proficiency). The type of subject did not seem to account for the distribution of the non-standard English instances in the production of the three test rating categories (cf. Table 10.4.). The analysis of verbs ending in sibilants showed that sibilant-final verbs were one of the explanations of the students’ non-standard English uses of the third person singular marker irrespective of test rating category (cf. Table 10.5.).

As for interviewer categories, assessors who were native speakers were more tolerant towards non-standard English instances than other assessors in a testing situation (cf. Table 10.6.). However, there were only thirteen oral tests assigned a Low grade altogether, and therefore only tendencies can be observed and reliable conclusions are difficult to draw. Nonetheless, this finding could be accounted for by the focus of the interviewers’ assessment on the student production: focus on message (content) rather than on form (language).

The students’ standard English and non-standard English uses of the third person singular marker tallied with their grades in English in school (cf. Table 10.8.). This could imply that the use of the third person singular marker is an indicator of general grammatical inaccuracy of importance for the students’ grades.

The Progressive

The mere number of instances of the progressive could not be expected to distinguish production of the three test rating categories High, Average and Low. The focus in the analysis was therefore on the types of instances observed. The high normalised figure of student production rated Low in comparison with student production assigned an Average grade might have been caused by the dominance of Task 1, relatively speaking, in the low-achievers’ production. Task 1 was expected to be the easiest task, and students with certain difficulties did not add very much to the action described with the help of some pictures. These pictures provided an evident context for using the progressive. The high normalised figure of the instances in production rated Low compared with production rated Average could thus be explained by the topic in Task 1 (cf. Table 10.9. and 10.16.). Mere frequency of the progressive was thus not enough to characterise student production rated High, Average or Low.

The proportions of the instances of the progressive simple and the progressive complex were not characteristic traits of any of the three test rating categories (cf. Table 10.10.). Not surprisingly, the difficulty in the use of the progressive tended to lie in the semantic context area rather than in the
application of standard English uses of verb forms. A great majority of the instances of the progressive were standard English instances, and no differences in distribution could be observed as regards standard English and non-standard English instances across the three test rating categories. Further conclusions are difficult to draw concerning the occurrence of the different contexts of the non-standard English instances and the grades assigned to the student production (cf. Table 10.13.).

In the individual student’s performance, there seemed to be a tolerance towards non-standard English instances when there were standard English instances in the same performance; there was a slight tendency that non-standard English instances in performance graded High more often occurred together with only standard English instances (cf. Table 10.14.). The results are in line with the tendencies found in the investigation of the third person singular marker (cf. above). For the production of students assigned an Average grade, the opposite tendency could be observed: a higher proportion of non-standard English instances occurred together with other non-standard English instances. The proportions of standard English and non-standard English instances in the individual student’s production seem to be a useful factor to include in the discussion of the relation between the students’ use of the progressive and the grades assigned. Furthermore, the results suggest that when there were high frequencies of the progressive, which could be explained by the topics of the task, it seemed that the interviewers noticed and were influenced by, the occurring non-standard English instances. As for interviewer categories, non-standard English instances were found in production rated High by School Teachers and University Teachers but not in production rated High by Native Speakers. When the distribution of the instances was compared with the students’ grades in school we could observe that students with high and average grades in English seemed to master the use of the progressive to the same extent.

Differences could be observed in the context of the instances of the progressive across the three test rating categories, for instance, a higher proportion of the instances in student production assigned a Low rating described the pictures in Task 1 than in other production. Students whose performance was rated High or Average used the progressive in other contexts.

We may conclude that the mere number of instances of the progressive could not distinguish production of the three test rating categories. No strong evidence could be found to support the idea that there are generally fewer non-standard English instances in production assigned a High grade compared with production assigned an Average or Low grade. In this context, the difficulty in determining whether instances were of standard English or non-standard English use should be mentioned. This fact may definitely have played a role in the assessment.
15.2. Vocabulary

In this section the students’ vocabulary will be discussed in two respects. Firstly the students’ use of two categories of vocabulary (abstract nouns and modifiers of adjectives) will be described. Secondly, the students’ lexical errors and their strategic competence to cope with gaps in vocabulary will be discussed.

Abstract Nouns
The results showed that a high frequency of abstract nouns (types and tokens) clearly distinguishes student production rated High or Average from student production rated Low (cf. Tables 11.1. and 11.2.). The occurrence of abstract nouns is thus a useful means of describing the difference between the language produced by high-achievers and low-achievers in oral proficiency tests of English. It seems that the ability to discuss or describe complicated matters with the help of items such as abstract nouns correlates with High grades in oral proficiency tests. The content and the topic in the production are probably of significance for the assessment. The investigation of types of abstract nouns across the test rating categories also revealed that students in production rated High more often used nouns which were not given in the instructions of the test. Furthermore, a majority of the instances appeared in Task 3 (cf. Table 11.6.).

The non-standard English use of uncountable nouns has low prominence in the material and does not seem to be a linguistic factor that influences the grades assigned to student production (cf. Table 11.4.). The students’ willingness to discuss and argue was probably of more significance than the proportion of standard English and non-standard English instances.

The results showed differences in boys’ and girls’ uses of abstract nouns. More types and tokens of abstract nouns were found in boys’ production (cf. Table 11.9.). This may indicate that the boys used a slightly different style than the girls in their performance. The frequency of abstract nouns could be one indication of this difference.

The grades from the three interviewer categories correlate with the frequencies of abstract nouns in student production assigned a Low grade: few instances correlate with Low grades. The frequencies of abstract nouns in production assigned Average and High grades tally with the grades set by Native Speakers. School Teachers and University Teachers do not seem to distinguish between these two categories of student production in this respect. One interpretation of this difference may be that Native Speakers tended to focus more on the message than the form in the students’ production. It is then assumed that production with many instances of abstract nouns was characteristic of students who were willing to develop their thoughts and attitudes. The idea that Native Speakers tended to focus more
on message than on form was pointed out by one of the interviewers in an answer to a question of the questionnaire after the completion of the tests. This interviewer’s comment gives support to the discussion above.

“I very rarely noticed anything going wrong with grammar. Probably there were lots of mistakes but they were simply ..., since I am listening as a native speaker, listening for the message, I didn’t notice them. I just once or twice noticed a don’t or a doesn’t which is trivial” (Native Speaker 3)

The results of the investigation of the distribution of abstract nouns across the three tasks suggest that students in production assigned a High grade were more inclined to discuss issues related to abstract matters in Task 3 than were the other students. Presumably, it was above all in Task 3 that high-achievers could show their knowledge of English abstract nouns. Furthermore, high frequencies of abstract nouns emerged in performance of students who had the top grade of 5 in English in school (cf. Table 11.8.).

We may conclude that many instances and types of abstract nouns characterised student production assigned a High or Average grade as opposed to a Low grade. Native Speakers tended to differ slightly from the other two interviewer categories in their grades in that student production with many instances of abstract nouns was rated High.

**Modifiers of Adjectives**

The results showed that the frequent use of modifiers of adjectives was slightly more common in production rated High than in production rated Average or Low (cf. Table 11.10.). Student production rated Low showed high figures regarding the types of modifiers used, which was quite unexpected (Table 11.11.). In general, however, the frequency of modifiers of adjectives showed a certain correlation with the Global grades assigned.

The presentation of the distribution of the types of modifiers used revealed that student production rated High contained modifiers such as pretty, really and rather as alternatives to the highly frequent very and also to the occurrence of so (cf. Table 11.12.) One possible explanation for the many instances of so in student production rated Average or Low may lie in the similarity to the Swedish word så: students who used so frequently were influenced by Swedish. When the three interviewer categories were taken into account, we could conclude that high frequencies of types of modifiers of adjectives appeared in production assigned a High grade by the Native Speakers. This tendency is in line with the results regarding abstract nouns (cf. above). It seems that the Global grades assigned by the Native Speakers, at least partly, were influenced by the students’ range of vocabulary.

The analysis of the relationship between the occurrence of modifiers of

---

1 Native Speaker 3 answered the questions by way of tape-recording.
adjectives and the students’ sex revealed a tendency for girls to be more inclined than boys to modify adjectives in their speech (cf. Table 11.15.).

We can conclude that mere frequency of modifiers of adjectives seemed to be insufficient to characterise student production rated High, Average and Low but the distribution of the types provided evidence not only of a relationship between a number of types of modifiers and the Global grades assigned, but also of differences across the interviewer categories.

**Students’ Lexical Errors and Compensatory Strategies**

As expected, more lexical errors of the types investigated appeared in student performance which were rated Low, whereas 50% of the production rated High did not contain any instances. The same tendency was observed in the frequencies of types of erroneous words used. Obviously the results are in line with the interviewers’ comments regarding the possibility that the students’ incorrect use of common words, such as *nature*, was a disturbing factor. It is thus likely that the inappropriate use of certain common words had a negative effect on the grades assigned.

Regarding the students’ use of compensatory strategies, the results showed that the performance assigned a Low grade contained considerably more compensatory strategies than other performance did. This is also what could be expected. Students who were assigned Low grades more often had to use compensatory strategies when they could not find an appropriate English word to express what they wanted to say.

The investigation of the distribution of the twelve types of compensatory strategies identified showed that, in general, the students used Strategies of English (for the categories of compensatory strategies, cf. 9.2.3. Students’ Lexical Errors and Compensatory Strategies): 71% of the instances of strategies were of this category. The tendency to use strategies based on the English language was stronger for students whose performance was assigned a High grade. This would imply that students who could efficiently compensate for vocabulary gaps, for instance by not switching to Swedish but instead giving two alternative words when they were uncertain of the appropriate word, were given better grades.

The interaction between the students and the interviewers may, at least partly, explain the students’ use of compensatory strategies. Two tendencies could be observed in the distribution of the strategies across the interviewer categories. Firstly, when the students were interviewed by the School Teachers, they were more inclined to use Swedish. Secondly, in these interviews the students more often asked for help (in Swedish or in English) from the interviewer to solve the difficulty. The fact that the students recognised the School Teachers from their own school probably explains these differences.
15.3. Discourse Markers

In the present section, the students’ use of two discourse markers will be discussed. The two discourse markers investigated are *well* and *you know*.

**Well**

The occurrence of *well* tends to characterise student production rated High rather than other student production. Students assigned a High grade thus used *well* more frequently than other students did and in this way they may have come closer to a native-like English style in their speech.

Obviously, it was not only a matter of frequency: *well* had important functions to fulfil in the students’ production and the instances of *well* indicate similarities between the students’ use and the descriptions of native usage of *well* presented above (cf. 9.3. Discourse Markers). Generally, it seems that the use of *well* which was not directly elicited by the interviewers characterises the language by the students assigned a High grade as opposed to *well* in answers, which was more frequent in performance rated Average or Low (cf. Table 12.3.). Students assigned a High Global grade used *well* in conclusions and in contexts with reported speech whereas the students in production assigned an Average grade used *well* in answers, with repetitions and when they had difficulties with the English language. The high-achievers’ use of *well* may be evidence of initiatives in their performance and their ability to structure speech in an appropriate way.

The survey of the distribution of the instances of *well* in relation to interviewer categories showed that the interviewers seemed to have different opinions about the positive effect of the students’ use of *well*. Two interpretations are possible. The first interpretation is that the non-native interviewers did not take into account the distinction between inappropriate and appropriate uses of *well*, whereas the native speakers did so. The second interpretation is that the students’ frequent use of *well* influenced the School and University Teachers to assign a positive grade whereas the students’ frequent use of *well* did not work in this direction with the Native Speakers. Since a frequent use of *well* characterises native spoken production, the first interpretation is more likely. The Native Speakers took into account features in the students’ use of *well* that may be difficult to observe for a non-native speaker. The students’ intonation and stress of *well* may be such a feature. Further research is needed to clarify the possible variation in the students’ use of *well*.

The investigation of student production that contained a remarkably high frequency of *well* showed the same tendency: the Native Speakers tended to be less positive in their grades of student production with many instances of *well* than were the other interviewers. Regarding the distribution across the
three tasks, the survey revealed that, as a whole, none of the three tasks made the students use *well* more frequently. The fact that there were few instances of *well* in Task 1 in production assessed by Native Speakers in comparison with the other two interviewer categories could be interpreted in the following way. The nature of Task 1 in the Native Speakers’ assessment gave fewer opportunities for hesitation or interaction with the interviewers. *Well* used in contexts of interaction with the interviewers was common since a high proportion of the instances of *well* emerged in answers to the interviewers’ questions and requests (cf. Table 12.3.).

The finding that *well* was more frequent in boys’ than in girls’ production suggests that there are differences between boys’ and girls’ performance (cf. Table 12.6.). Regarding the distribution in relation to the students’ grades in English in school, the results showed that although there may be great intra-group variation, students with high grades in English in school tended to use *well* more often than did the other students (cf. Table 12.7.).

We can conclude that *well* was more frequent in student production rated High than in production rated Average or Low and in particular in production rated by the non-native interviewers. Tendencies of qualitative differences in the uses of *well* in the three test rating categories could be distinguished. The idea that *well* is an important feature of fluency in spoken English is corroborated in the present investigation and of particular interest when focusing on differences in the assessments by native and non-native interviewers.

*You know*

Regarding the second discourse marker, the students’ uses of *you know* tended to be a characteristic trait of production rated Average or Low rather than High (cf. Table 12.8.). One interpretation of the results is that *you know* has meanings and functions in the students’ uses that made it inappropriate in their performance. *You know* used in certain contexts could be one of the features that contributed to a Low or Average grade.

The presentation of the collocations attested showed that the high-achievers, in particular, tended to link *you know* to the rest of their production with the help of *and*; *and* occurred together with *you know* in one third of the instances in production rated High.\(^2\) Concerning the position of *you know* in the instances attested, the great majority of all the instances were placed according to native usage irrespective of Global grade assigned (cf. Table 12.9). This could mean that the students tended to place *you know* idiomatically regardless of test rating category. As to the discoursal functions attested, the results revealed that in the majority of the instances, the

\(^2\)The raw figures were admittedly very low but nevertheless the tendency is of interest.
students used *you know* in a range of different contexts, and further study is necessary to pinpoint these discoursal functions (cf. Table 12.10.). The instances that were possible to categorise, however, showed differences in distribution: *you know* emerged in contexts of vocabulary gaps and in repetitions in student production assigned an Average or Low grade. Student production assigned a High grade contained more instances in contexts when the students expressed their own opinion. The results are not surprising. Students with production assigned an Average grade could be expected to have more difficulty in finding the appropriate words than students whose production was assigned a High grade. Instead of using *you know* in word searching, the high-achievers used it in contexts where they probably felt the need to soften their utterance.

The distribution of instances of *you know* across interviewer categories showed that the Native Speakers seemed to be more negative to the students’ uses of *you know* than the other interviewers and in particular, the University Teachers. It may be the case that *you know* was used in an unidiomatic way by the students, which only the Native Speakers took into account.

The results of the survey of the introduction of *you know* by students or by interviewers by test rating categories suggest that the students whose production was assigned a High grade were more aware of the situational implications of *you know* and tended to use it more often in appropriate situational contexts than were the other students (cf. Table 12.12.). The interviewers’ use of *you know* in interviews with students who were assigned a High grade could be an indication of the situational appropriateness of the students’ use. The fact that the students used *you know* when the interviewers also did so could be interpreted as evidence of the students’ ability to use *you know* in what could be referred to as “stylistically appropriate contexts”.

The results of the distribution of the instances of *you know* across the tasks are in line with the observations in Östman (1981: 16) described above (cf. 9.3. *You know …*). The instances clustered in Task 2 that contained a narrative topic. The presentation of the relationship to the students’ sex showed that *you know* was more frequent in production by girls than by boys (cf. Table 12.15.). The results may suggest that the Swedish students had adopted certain linguistic features that distinguish between male and female native speakers of English (cf. 9.3. *You know …*). Even if similar differences in language use could be found in Swedish and the students simply carry over their style from Swedish to English, it is nevertheless worthwhile highlighting that the English language produced by advanced Swedish students contained variation that is similar to that found in descriptions of native speakers’ production. One possible interpretation of the re-
results is that you know is a marker of an interactive and intimate style, which is more characteristic of girls’ than boys’ speech. As for the second student parameter, the students’ school grades, the results indicated that frequent use of you know tended to characterise the speech of students who did not have the top grade 5 in English in school (cf. Table 12.16.).

15.4. Aspects of Pronunciation

In this section the results from the investigations of the students’ pronunciation will be elaborated on with regard to their intonation, stress and speech rate.

Intonation

The findings summarised below are based on a small sample and with the help of three researchers’ impressionistic evaluation. They may still yield some useful information.

Intonation seems to vary in student production across the three test rating categories High, Average and Low. “Unidiomatic” intonation with rises in statement-final words and, in general, monotonous and flat intonation are more characteristic of student production rated Average or Low than High. It is worth pointing out however, that other features than intonation, such as stress and rhythm, seem to be more significant for the assessment of Swedish students’ spoken production.

Stress

The findings suggest that “unidiomatic” stress of function words is a prominent feature in the distinction between the high-achievers and the low-achievers. Students whose production was assigned a High grade had adopted an English-like stress pattern which they used in the extracts investigated.

The results of the investigation of the students’ use of and so on showed that the students rarely pronounced and so on in an “idiomatic” way, but the “idiomatic” instances that did occur were found in production rated High.

In general, the students’ stress patterns in their spoken English seem to play an important role in distinguishing student production rated High from student production rated Average or Low. High-achievers had adopted more “idiomatic” English stress patterns than other students.

Speech Rate

High speech rate characterises students who were assigned a High grade. No differences in distribution could be observed when the three interviewer categories were taken into account. Nor were clear tendencies found in the
distribution across girls’ and boys’ production. There was a relationship between speech rate and the students’ grades in English in school, however.

We can conclude that speech rate was probably one of the features in the student production that contributed to the High Global grades assigned.

15.5. Interference from Swedish

Interference from Swedish is of interest in the present study for at least four reasons. Firstly, references to the students’ mother tongue were made in the rating criteria, e.g. grade E of the Factorial rating 2a (Grammar and Vocabulary) had the wording: “Sometimes uses Swedish words”. The criteria of the same grade for Pronunciation, i.e. grade E of Factorial rating 2c, also referred to the student’s mother tongue, to “first language intonation”.

Secondly, the assessors made remarks on the interference from Swedish in their comments added to the grades. By way of examples, see the comments given below.

“Använder svenska ibland.” (= “Uses Swedish sometimes.” School Teacher 1, I 84)
“Intonation patterns not very English.” (Native Speaker 1, YY 28)

Thirdly, in their answers to the question “What is particularly disturbing in some students’ production?” (cf. question 5, the Interviewers’ Questionnaire in Appendix 6), the interviewers mentioned the interference from Swedish as one of these disturbing features. Two answers are cited for illustration:

“Swenglish, svenska som bara direktöversätts och låter mycket oengelskt.” (= “Swenglish, Swedish which is just directly translated and sounds very un-English.”) (School Teacher 3)
“Svensismer som it is för there is.” (= “Swedish ways of saying things like it is instead of there is.”) (University Teacher 2)

Fourthly, the investigation of the ten linguistic features revealed that parts of the students’ production showed evident interference from Swedish as regards grammar, vocabulary and/or pronunciation, e.g. the absence of the third person singular marker, the choice of abstract nouns which are similar in Swedish and English, and the stress pattern of and so on.

Overall, the students’ production is characterized by little interference from Swedish. However, if any interference is noticed, it pertains, in particular, to the students’ grammar, vocabulary and intonation and less to their pronunciation of individual sounds.

3 Cf. the Rating Criteria in Appendix 8.
very few oral tests were assigned the grade E for Grammar and Vocabulary, and/or for Pronunciation. Among the grades assigned, only one student’s production was given a grade according to the criterion “Sometimes uses Swedish words” (cf. above). In grammar, the students’ difficulties seem to lie in morphology and in the choice of words. In pronunciation, the assessors’ comments tend to describe interference in students’ intonation, whereas students’ pronunciation of individual sounds is regarded to be accurate. For illustration, Native Speaker 1 comments on four tests as follows:

“somewhat disturbing rhythm in speech but not obviously foreign” (Comment on test YY 28)
“intonation somewhat monotonous” (Comment on test B71)
“individual sounds OK” (Comment on test X24)
“intonation patterns not very English; individual sounds good” (Comment on test R18)

In this context, some remarks should be made concerning certain observations of word order in the students’ production. Errors in word order may be expected to occur together with errors of morphology or vocabulary, in particular in student production assigned Low grades. These errors could be anticipated to occur in structures with syntactic differences between Swedish and English, for instance, regarding the position of adverbials and the subject-verb order in certain contexts (Hargevik and Hargevik 1993: 3, 120; Ljung and Ohlander 1992: 291), but the students’ difficulties when they speak English in the oral interviews do not seem to concern word order. One evident explanation could be the fact that in many sentences, there are similarities between Swedish and English in this respect, and therefore students’ word order in English is the same as in Swedish. Another explanation could be that the students have mastered word order in English and thus have few errors even though there are differences between the two languages. There are structures in the material where errors could be anticipated because of differences in word order in Swedish and English. In these instances however, there is no interference from Swedish; the students’ use of English word order is correct. Six examples below illustrate this absence of interference in the students’ production. The contrastive structure identified comprises the subject-verb word order. Furthermore, illustrative examples are given of appropriate use of the preposition followed by the present participle.

(93) “… without clothes we almost look the same …” (YY 28)
(94) “… well usually they have a stronger message sometimes but it doesn’t have to be that either” (WW 30)
(95) “… well of course not but in many ways that isn’t the problem…” (Z26)

256
(96) “… well people could learn a lot from going back to school and do the things they are interested in …” (Y51)

(97) “… I think you got the sort of prejudices from looking at people, their clothing … there’s so many other things in life that you can spend your money on than buying expensive new clothing all the time…” (YY 28)

These examples also show the absence of interference in two other respects: the choice of preposition (“they are interested in”, cf. Swedish: intresserad av) and the use of there instead of it (“there’s so many things in life”, cf. Swedish: det finns).

In conclusion, the interference from Swedish in the students’ production was evident as regards for instance vocabulary and intonation but does not seem to be significant for the Global grades assigned. Furthermore, interference from Swedish was not emphasized by the interviewers. Generally, interference from Swedish was thus a less important factor for the assessments than one might have expected.

15.6. Concluding Remarks

The students who participated in the Gothenburg project represented a limited sample of the students from Swedish upper secondary school since they were all theoretically oriented in their studies in the gymnasieskola (cf. 7.2. Students). It could therefore, at least partly, be expected that these students would not show variation in their proficiency in English. The analysis of their performance showed evidence of interesting differences, however, regarding their use of the verbal group, vocabulary, compensatory strategies, the two discourse markers investigated as well as their pronunciation.

The student performance contained incorrect and correct instances of different kinds of the two features of the verbal group investigated. Examples of incorrect use were the absence of the third person singular marker after the pronouns he or she, and the use of the progressive with stative verbs such as cost. Correct use of the verbal group was observed, for instance, in the students’ use of the progressive to describe action in progress. Generally, a high proportion of grammatically correct instances appeared in a student’s performance which was rated High.

Students’ vocabulary was commented on as being significant for the overall impression of their production, in particular by the Native Speakers (cf. 9.2. Vocabulary). The results pointed at the significance of the frequency of abstract nouns, probably as an indication of abstract thinking, and the use of varied vocabulary, for instance by using different types of
modifiers throughout a performance. As expected, instances of lexical errors involving certain common words characterised student performance rated Low. Compensatory strategies were also more frequent in performance rated Low.

---

Table 15.1. **Rejected or confirmed sub-hypotheses as regards the ten linguistic features**

<table>
<thead>
<tr>
<th>Linguistic feature</th>
<th>Sub-hypothesis</th>
<th>Confirmation Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Group: The third person singular marker</td>
<td>There is a higher proportion of standard English instances in student production assigned a High grade than in other student production.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>Verbal Group: The progressive</td>
<td>There are fewer non-standard English instances in student production assigned a High grade compared with student production assigned an Average or Low grade.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Vocabulary: Abstract nouns</td>
<td>Abstract nouns characterise student production assigned a High grade. Frequent instances and types of abstract nouns, in particular nouns with no similar Swedish equivalent, are evidence of a more developed and advanced argumentation.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>Vocabulary: Modifiers of adjectives</td>
<td>More instances of modifiers of adjectives emerge in student production assigned a High grade than in other student production.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Vocabulary: Students’ lexical lacunae and communication strategies</td>
<td>Lexical errors are more frequent in student production assigned an Average grade, and, in particular, a Low grade. Compensatory strategies are more frequent in student production rated Low. Differences in occurrences of types of strategy can be observed across the test rating categories.</td>
<td>Partly confirmed</td>
</tr>
<tr>
<td>Discourse Marker: Well...</td>
<td><em>Well</em> is more frequent and is more idiomatically used in student production assigned a High grade than in other student production.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>Discourse Marker: You know</td>
<td><em>You know</em> is more frequent in student production assigned a High grade than in other student production. <em>You know</em> is used as a pause-filling device in the student production and as one of the means to establish interaction between the interviewers and the students.</td>
<td>Partly rejected</td>
</tr>
<tr>
<td>Aspects of Pronunciation: Intonation</td>
<td>“Unidiomatic use” of intonation in declarative sentences is more frequent in student production assigned an Average or Low grade than in student production assigned a High grade.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>Aspects of Pronunciation: Stress</td>
<td>Stress, both on functional as opposed to lexical words and on the phrase <em>and so on</em>, is more often found to be used according to English usage in student production assigned a High grade than in other student production.</td>
<td>Confirmed</td>
</tr>
<tr>
<td>Aspects of Pronunciation: Speech rate</td>
<td>Student production assigned a High grade contains samples with higher speech rate than the samples of the other student production.</td>
<td>Confirmed</td>
</tr>
</tbody>
</table>
The investigation of the students’ use of the two discourse markers provided useful information regarding both frequencies and discoursal features. High frequencies of *well* emerged in student performance rated High, particularly when assessed by non-native speakers. *You know* appeared in other student performance and often when the production was assigned a Low grade. Discoursal factors, such as if *you know* was used in contexts when the students were expressing their own opinions, seem to be related to the Global grades assigned.

The students’ pronunciation had an evident Swedish accent and showed differences in quality both regarding intonation and stress. The students’ speech rate during the first minute of their performance was another factor, besides stress and intonation, which seems to be related to the grades assigned.

The results of the investigation can be summarised with the help of rejecting or confirming the ten sub-hypotheses presented in Chapter 6 (cf. 6.2. Linguistic Features). In Table 15.1, the ten sub-hypotheses are confirmed or rejected.
Conclusion

In this final part of the study, the results will be summarised and conclusions will be drawn. Furthermore, some reflections will be provided on the testing of oral proficiency in English at schools.

1. Summary and Conclusions

In this section, conclusions will be drawn from the results presented in Chapters 8 to 13 and the interpretations given in Chapters 14 and 15. The aims of the study explained in Chapter 4 will provide the basis for the discussion. The conclusions should be considered against the aims and context of the Gothenburg project.

This project was carried out in 1993 when compulsory oral proficiency tests were about to be introduced as a part of the National tests in Sweden and while a norm-referenced grading system was still in use in the Swedish school system (cf. 1.1. Measuring Student Performance). The use and interpretation of rating criteria were not yet widely spread in assessments of language proficiency. The nine interviewers of the Gothenburg project did not study or discuss the rating criteria used during the sessions and given to them when they arrived before the assessments were carried out, something that has since of course become common practice together with the introduction of compulsory oral proficiency tests as part of the National tests in Sweden shortly after the Gothenburg study was carried out. Furthermore, the experimental situation of the oral tests and the limited number of participating students and interviewers should be kept in mind in the discussion of the observations that follow, as should the fact that the study was undertaken not to make an evaluation of Swedish school-leaving students’ general oral proficiency in English but to study interrater-reliability, that is, the degree to which different assessors tend to agree or disagree in their grading of the same students. The results and tendencies observed in the present study can nevertheless be of some interest.
As described in Chapter 4, the overall aim of the Gothenburg Project was
– to study to what extent different examiners tend to agree or disagree in
grading advanced Swedish students’ oral proficiency in English.

The material used in the present study was collected at the Foreign Lan-
guage Teaching and Testing Research Unit of the Department of Education
and Educational Research, Gothenburg University. The results have been
presented, discussed and interpreted in the present study. When the inter-
viewers assigned Global and Factorial grades to all 29 students’ perform-
ance and thus four different grades to each test, there was great variation in
the grades they awarded, both as far as grade levels in general were con-
cerned and in the grades they awarded the same students; in the case of 11
students the variation in opinion was remarkably great (e.g. three different
Global and Factorial grades were given by the three interviewers to the
same student). That there were great differences in the grades assigned by
such experienced and qualified interviewers as those who took part in the
Gothenburg project is an important finding of the present study. Among the
factors that may explain these differences, the following have emerged
from the analyses:
– the order of the interviews (lower grades were assigned to the students’
  first interviews),
– the length of the students’ production,
– the interviewers’ different interpretations of the rating criteria,
– their different views on the influence of Swedish on the students’ per-
  formance in English, and
– their slightly varying ability to make the students feel at ease and thus to
  produce language at their true level of proficiency.

Concerning the Factorial grades, the results show that the interviewers
tended to use a limited part of the rating scale for Pronunciation, in particu-
ar the Native Speakers.

Furthermore, by way of specification of the aim above, the first aim of
the analysis of the present study was two-fold:
– to study the relationships between the Global grades and the three Facto-
  rial grades given by each of the nine interviewers individually;
– to study the relationships between the Global grades and the three Facto-
  rial grades for each of the three interviewer categories.

The Global grades and the Factorial grades were closely related, that is, the
interviewers often assigned the same Global and Factorial grades to the
individual student’s production (cf. Table 8.5.). This tendency was particu-
larly strong in the grades assigned by the Native Speakers (cf. Table 8.7.).
As regards the relationships between the Global grades and the Factorial grades, the results indicated that the School Teachers’ and the Native Speakers’ grades for Grammar and Vocabulary showed the highest correlation with the Global grades assigned (ST: 0.77; NS: 0.89). As for the University Teachers’ grades, the Factorial grades for Communication and Fluency showed the highest correlation with the Global grades (0.90). However, on the whole and considering all the interviewers, no Factorial grade seems to have been more influential than the others when the interviewers assigned Global grades to the student production. Furthermore, the results showed that the Native Speakers generally gave slightly higher grades, whereas the School Teachers gave slightly lower grades than did the members of the two other interviewer categories. The Native Speakers and the School Teachers gave priority to different aspects in the student production. The Native Speakers seem to have paid more attention to the students’ communicative skills in their assessment whereas the School Teachers were more influenced by the students’ grammatical skills. In other words, there is a different emphasis on fluency versus accuracy in the assessments by the members of these two categories of interviewers.

The second aim formulated for the present study, was

– to investigate the students’ linguistic competence, and to some extent their strategic competence, in terms of their standard English or non-standard English use of certain “difficult features” in English, their vocabulary, their use of certain discourse markers, three aspects of their pronunciation and their ability to use efficient strategies.

The analysis of the students’ performance yielded useful information on their mastery of the English language with respect to their oral proficiency in an interview and as regards certain linguistic features investigated. The students had no difficulty using English in the different tasks of the Gothenburg project. Almost all students managed to speak about the three topics selected.¹ Their strength lay in their communicative skill; they were generally good at getting the message across. Regarding vocabulary, most students showed a varied and appropriate use of abstract nouns and modifiers of adjectives. Furthermore, they had adopted discoursal patterns from English usage in their use of the two discourse markers well and you know. These were employed in appropriate ways as hesitation devices or in contexts where the students expressed their own opinions. Their pronunciation could generally be considered to represent the middle of the rating scale: the students had good pronunciation of most of the individual sounds and acceptable intonation, but their foreign accent was evident. There was,

¹ One student (Student ZZ in the interview with the Native Speaker 3) did not complete Task 3 (argue a case).
however, variation in quality; there were also differences in speech rate and stress patterns.\(^2\) As regards syntax, it is worth special attention that no instances of non-standard English word order were observed. For the grammatical inaccuracies observed, interference from Swedish seemed to be an influential factor, for instance the distinction between *there* and *it* as the equivalents of Swedish *det*, and the inappropriate use of the progressive determined by semantic criteria, that is, the progressive was used in inappropriate contexts (cf. 10.2. The Progressive).

Finally, some interesting conclusions concerning two background variables could be drawn, namely, the relationship between the grades awarded in the interviews and the students’ sex and their school grades respectively.

**Sex.** The analyses of the oral proficiency interviews indicated that there were certain differences between boys and girls in their use of a number of linguistic features. As for vocabulary, abstract nouns tended to be used more by the boys, whereas modifiers of adjectives, for instance *very*, were more frequently used by the girls. Regarding the two discourse markers investigated, the results showed that *well* was more frequently used by the boys than by the girls. As to the use of *you know*, the girls used this discourse marker more frequently than the boys did, and this phenomenon is, interestingly enough, also found in native speakers’ speech (cf. 9.3.2. *You know…*).

**School Grades.** As for the relationship between the students’ school grades in English and their performance, two findings are of interest. Firstly, the students who had fairly low grades in English in school tended to be quite strong in oral proficiency. These students were generally not assigned remarkably low grades in the interviews. Secondly, most of the students who had top grades in English in school showed very good proficiency during the interviews, but some of these students’ performance did not differ remarkably from the other students’ performance. It seems reasonable to assume that there were “uneven profiles” among these students who perhaps had their strength in, for instance, reading and writing skills.

Finally, as a combination of the above first and second aims, the third aim of the present study was

- to investigate to what extent the students’ linguistic and strategic competence is graded differently by the three categories of interviewers, and especially to study the role played by the linguistic features in the Global grades assigned to the students by the three categories of interviewers.

The results of the linguistic analyses showed that five of the linguistic features seemed to have had different effects on the Global grades assigned by

\(^2\) This was observed, for instance, regarding their idiomatic or unidiomatic stress of *and so on*. 

263
the members of the three interviewer categories. The five features are the third person singular marker, abstract nouns, modifiers of adjectives, and the discourse markers *well* and *you know*. The students’ use of the progressive (although a slight tendency could be observed, cf. 15.1. The Progressive), their speech rate and the occurrence of lexical errors did not show this tendency. It was not within the scope of the investigations of the students’ intonation and stress to include the interviewer categories. In what follows the tendencies of differences across the interviewer categories are further discussed regarding communication, grammatical accuracy, number of certain categories of words, compensatory strategies, discourse phenomena and pronunciation.

First of all, regarding **communication**, it seems that the students who were assigned High grades tended to differ from the other students in their will to discuss abstract matters. The students who discussed abstract and more demanding topics, and who spoke at length and probably with many initial turns (i.e. they took the initiative), were assigned High grades. Whether abstract nouns, for instance, uncountable nouns, were used according to grammatical rules or not, did not seem to disturb the communication. This tendency was particularly obvious in the assessments by the Native Speakers.

Secondly, as for **grammatical accuracy**, the interviewers displayed a tolerance towards occasional non-standard English instances if the same linguistic features were also dealt with correctly in other parts of the same student’s performance. Non-standard English instances were more or less overlooked in student production assigned the highest grades as long as other instances of the same feature were grammatically correct. The following tendencies could be observed concerning the third person singular marker and the progressive: non-standard English instances emerged in the student performance of all three test rating categories (High, Average and Low). When the interviewers assessed the student performance, grammatical accuracy was obviously a matter of proportions rather than one of occasional instances in the individual student’s production. As expected, however, student production that was assigned High grades (i.e. the Global grades B/4 or A/5) generally contained a higher proportion of grammatically correct instances of the third person singular marker, for instance, than student production assigned Low grades (i.e. the Global grades E/1 or D/2). However, even though a student’s production was assigned the top Global grade, which had the rating criterion “A fluent and overall correct use of English”, there could be non-standard English instances. As for the interviewer categories, the results showed that non-standard English instances of the third person singular marker were more common in production assigned a Low grade than in production assigned a High grade by the
School Teachers and by the University Teachers. This tendency in distribution was not seen in production assessed by the Native Speakers.

The progressive was the second linguistic feature of the verbal group subjected to a detailed study. The results showed that a great majority of the instances were of standard English use and that there were no great differences in the distribution of standard English and non-standard English instances when student production was analysed in relation to the three test rating categories. Nor were there differences in the types of non-standard English instances attested. However, although few non-standard English were found, three tendencies could be observed. Firstly, the non-standard English instances in production rated High more often occurred together with only standard English instances than production rated Average or Low (cf. Table 10.15.). Secondly, these non-standard English instances, that is, instances in production rated High compared with instances in production rated Average or Low, were of a different type (not belonging to the category labelled Habit, but instead the category labelled Objective circumstance, cf. Table 10.14.). Thirdly, no non-standard English instance of the progressive was found in production rated High by the Native Speakers.

The Native Speakers seemed to be more tolerant of grammatical inaccuracy, as regards the third person singular marker, for instance, in their overall assessment of student performance than the School Teachers and the University Teachers. These results together with the fact that abstract nouns were more frequent in student production assigned higher grades by Native Speakers than by the other interviewers could be interpreted in the following way: the Native Speakers focused on the students’ communicative skills and the messages conveyed, while the School Teachers, in particular, tended to pay more attention to grammatical accuracy and the students’ grammatical repertoire.

Thirdly, mere numbers of certain categories of words in the student performance were not determining factors in the assignment of grades. Instead the different interviewers paid attention to types of words used and various discoursal factors. The types of modifiers of adjectives that the students employed (e.g. fairly instead of so), and in what discoursal contexts they used well, are examples of features that seem to be of significance for the assessment. The tasks and the topics of each task in the interviews seemed to be of significance for the distribution of words and discourse markers and thereby for the grades assigned.

Fourthly, when the students did not know the English word or expression they obviously wanted to use, they resorted to various compensatory strategies. Their use of strategies which were based on the English language, for example, paraphrase (as opposed to strategies involving Swedish, e.g. word coinage when the students translated Swedish expressions literally), was
considered to be a positive feature by the interviewers. In general, and not surprisingly, compensatory strategies were more frequent in performance assigned Low grades. The student performance which was assigned High grades showed a higher proportion of compensatory strategies involving English only. Efficient ways of solving gaps of vocabulary thus resulted in good grades.

Fifthly, discourse phenomena, such as the students’ use of hesitation devices, had different effects on the Global grades assigned. The discourse marker well was apparently associated with good fluency in spoken English when the performance was assessed by non-native speakers (high frequencies of well were observed mainly in performance assigned High grades by non-native speakers). The students’ use of you know, on the other hand, seemed to have had the opposite effect: frequent use occurred in production assigned a Low grade, particularly when the performance was assessed by Native Speakers, something which is worth noticing. Students’ use of discourse phenomena may thus influence assessors in different ways, and perhaps the Native Speakers could take into account unidiomatic use of discourse markers which were hard to identify for the non-native speakers.

Finally, as regards the students’ pronunciation, the results from the interviewers’ comments and their answers in the questionnaires did not provide evidence of the students’ Swedish accent being of significance for the assessment of their intonation in English (cf. 13.1. Intonation). It might be expected that the majority of the students spoke Swedish with an accent characteristic of the west coast of Sweden since they were from this region but this was not discussed in the comments to the assessments by the interviewers or by the two experts consulted on the students’ intonation (cf. 13.1.1. Results from the Observations of Students’ Intonation). The limited investigation of the students’ intonation in declarative sentences, their stress pattern and the survey of their speech rate the first minute of the oral test showed differences in quality across the three test rating categories. Furthermore, the interviewers’ comments together with the Factorial grades assigned for Pronunciation indicated that the students generally had a typical Swedish accent, and that certain factors, such as rhythm and stress, seem to have been of importance for the grades assigned.

2. Reflections

On the basis of the results obtained in this study, some reflections will now be provided on the testing of oral proficiency in English in our schools. It should be emphasized that these reflections are given in the light of the method used in the oral proficiency interviews of the Gothenburg project.
The method and the procedures in the testing and the assessment of students' oral proficiency, which is now a compulsory part of the National Tests both in the comprehensive schools and in the upper secondary schools, are to some extent different from the one used in the small-scale study of 1993. These changes include, for instance, new interactive models with students responding in pairs or groups. As for the grading, factorial ratings are not given but are commented on in the instructions for the assessment. The fact that a completely new criterion-referenced grading system was introduced in 1994 with sets of rating criteria for all school subjects (for instance for English, not for separate skills such as oral proficiency) has led to the fact that teachers now are used to assessments carried out with the help of rating criteria. Finally, the instructions of the National Tests of English used in schools now in 2003 give extensive support in terms of recommendations, instructions and, as far as foreign languages like English are concerned, commented examples of student performance.

It is my hope that the practical implications of the present study and the discussion below may be of interest to headmasters and language teachers, in particular to teachers of English but also of other foreign languages, in upper secondary schools, and partly to teachers of the upper grades in compulsory schools, when they are planning and carrying out assessments of their students’ oral proficiency. The discussion focuses on the following three issues:

1. the reasons for oral proficiency testing,
2. important factors in the construction of oral proficiency tests: the preparation, the organisation and the assessment, and
3. in-service training for teachers.

1. The results of the present study show that students with low grades in English in school tend to have, relatively speaking, a greater command of oral proficiency than of other skills. Now, in 2003, when the oral test has been made compulsory, these students, no doubt, have an opportunity to demonstrate their strengths. The students who participated in the Gothenburg project were given the opportunity to show certain skills that would of course not have been visible in written tests, for instance, skills in using intonation, stress and the pronunciation of individual sounds, as well as their competence in handling turn-taking and other discoursal features. The fact that oral tests give students the opportunity to show skills not possible to measure in other tests speaks in favour of compulsory testing of oral proficiency within the Swedish school system, as oral proficiency is one of

---

3 This model is found, for instance, in the oral part of the National Test of English A and English B of upper secondary school of 2001 and 2002.
the four overall language competence goals of the syllabi. However, this has not always been taken for granted. Before oral proficiency tests were in use, it was assumed that teachers could evaluate their students’ oral proficiency in the course of regular classroom work. The finding that students whose school grades in English were low received relatively higher grades in the oral testing than might have been expected gives further support to the system of compulsory oral tests. Another reason that might be mentioned in this context is the ‘washback effect’, i.e. the effect of testing on teaching (cf. Bachman 1990: 283 and Olsson 1996). What is tested is also given priority by the students in the learning process. Oral proficiency tests, which have been used for the last ten years, are thus a necessary compulsory part of the National Test of English.4

(2) An oral test needs to be of a certain length to provide enough material for the assessment. Students should be given a fair chance to show their competence, and, therefore, it is recommended that oral tests where students turn out to be hesitant and insecure should be lengthened in order to ensure reliable grades. One must be aware of the special needs of some students in oral proficiency tests; the lack of self-confidence and nervousness may contribute to a negative first impression of these students’ real competence. To a certain extent, flexible length of oral proficiency interviews could be used although it may complicate the organisation of the tests.

When the tasks for an oral proficiency tests are set up, there are several aspects to be considered. Oral proficiency tests are to be designed to give students the opportunity to show their oral skill, in particular students who may have a poor command of English. Oral proficiency, of course, comprises more than pronunciation and intonation, for instance, pragmatic competence (e.g. the ability to use language appropriately according to given contexts) and strategic competence (e.g. the ability to compensate for communication breakdown) (cf. 2.2. Oral Proficiency). The tasks must therefore be varied and constructed in such a way that the focus is on the student’s overall oral proficiency and those aspects of his/her competence that are not tested in the written tests. The tendency to overemphasize accuracy can also be compensated for. If, for example, the students are asked to tell a story about a person in the present tense, there may be a risk that the interviewer pays too much attention to an easily recognisable grammatical feature such as the third person singular marker. As a consequence, the student’s standard English or non-standard English use of this feature may be given a degree of weight in the overall assessment of the her/his performance that it does not deserve. This could be avoided by having the

4 The oral test, as a part of the National Test of English A or of English B of Swedish upper secondary schools, has been compulsory since 2000 (Erickson and Börjesson 2001).
student tell parts of the story in the past tense or by way of asking the student to act the role of the main character in the story. The inappropriate use of certain words, such as “false friends”, which present problems for Swedish students, may also pull down grades in a disproportionate manner. In cases where such vocabulary problems can be anticipated, the correct words could be supplied in the instructions.

The tasks should also offer opportunities both for simple storytelling and for presenting arguments and ideas in different situational contexts, in both formal and informal frameworks, for instance. A discussion of controversial topics such as those used in Task 3 in the tests of the Gothenburg project (i.e. “to argue a case”) is interesting to every student in oral tests. The purpose should be to give all students, weak and strong, opportunities to show their competence in English as fully as possible. Discussions – both between a teacher and a student, or between students – may stimulate students to show their mastery and fluency, for instance, in their use of linguistic flags characteristic of spoken English, such as hesitation devices and discourse markers in answers to the interlocutor’s questions and requests, e.g. let’s see now, I can’t find the word, let me think for a while.

What’s more, discussions regarding controversial issues (which may be highly demanding for some students) give the students the opportunity to show their strategic competence, and ideally, give all students the opportunity to do their very best.

Regarding practical procedures and principles for grading students’ oral proficiency in oral tests, it is excellent that instructions accompanying oral tests, for instance, the oral part of the National Tests of English, give practical recommendations on how to organise oral proficiency tests at school well in advance so that students are familiar with what is expected of them and how they should go about showing their oral competence. Detailed guidelines on how to prepare and carry out assessments in order to make the grades as reliable as possible and useful illustrative examples from testing procedures in use at various schools are also helpful and should constitute an obvious part of the test material. Preparations of various kinds, both for the interviewers and for the students, in all schools and especially in classes facing organised oral proficiency testing in the near future are thus recommended, as are well-known routines to ensure that previous knowledge of the student’s results does not influence the assessment (cf. 5.3. Sources of Variation). In order to reduce the influence of negative factors on the assessment, such as the possibility that the students feel nervous during the exam, oral proficiency – even if it is of course continually practised –

---

5 For illustrative examples of material which learners may use to develop their oral proficiency in English, see for instance Gustafsson and Kinrade (2002), Isitt (1998), Aagard and St. John (2002) and Trollfjärden and Svedberg (forthcoming).
should also be tested in everyday classroom activities. In this way, students’ lack of previous experience of oral tests will not affect the examination experience. Since the order of the interviews seems to have played an important role for the grades assigned in the Gothenburg project, it is essential that students be given the opportunity to participate in oral proficiency testing and to gain experience from testing situations. Furthermore, assessment by two examiners is recommended and should be encouraged whenever possible. Other ways of securing that the grades awarded to the students’ oral performance are as reliable as possible are, for instance, an exchange of examiners between schools and the recording of student performance.

(3) The results of the present study pointed to the unavoidable fact that, in many cases, different interviewers graded an individual student’s oral performance differently. It is reasonable to believe that similar results would be obtained if groups of teachers from different schools in Sweden were to take part in an experiment like the one discussed in the present study. To ensure as reliable grades as possible in oral tests, assessment of oral proficiency is thus an in-service training matter of the greatest importance. The differences observed in the grading by the three interviewer categories of the present study could be used as an interesting starting point in discussions of interpretations of rating criteria for the assessment of students’ oral performance. Further training of English teachers could include three main parts:

(1) to study developments in present-day spoken English,
(2) to discuss cases on which assessors tend to disagree, and
(3) to discuss recent methods in the testing and assessing of students’ spoken English.

The first part could focus on various aspects and trends in present-day use of spoken English. It should be stressed that the Swedish students in the present study had reached such an advanced level that many factors such as meta-talk (*this word is hard to pronounce*), or-tags (*or something, or so*), abbreviated clauses (*don’t really know*), self-corrections and other charac-

---

6 Students and teachers can be informed about test formats with the help of the examples provided of National Tests of English for their respective levels.

7 Assessment by more than one assessor and the recording of students’ performance are recommended, for instance, in the instructions of the National Test of English of Compulsory Schools (AP 9). For a discussion of ways of preparing students for the tape-recording of oral proficiency tests, see Tornberg (2000: 157).

8 Another reason for an emphasis on oral proficiency in in-service training of English is that teachers appear to have increased the amount of various oral exercises in their teaching. In a study from 2001, teachers experience a movement towards more speech and oral presentations in the teaching methods of English in Sweden (Apelgren 2001: 157).

270
teristics of spoken English may have contributed positively to the grades assigned (for an introduction to spoken interaction, see for instance, Stenström 1994, and Aijmer and Olsson 1990). The students’ interactive skills could have been of influence (for examples of interactive skills, see Aijmer 2001). In-service training of teachers in modern developments in present-day English with regard to native versus non-native speakers’ mastery of spoken English in different situational contexts would therefore be of great value (for discussions of various aspects of present-day English, see for instance Crystal 2000, Hargevik 2000, Svartvik 1999, and Brutt-Griffler 2002).

The second part could focus on the following areas:

– defining students’ communicative competence, in particular oral competence,
– examples of student performance in the middle of the rating scale,
– the performance of “silent and passive” students, and
– students with an “uneven profile” (e.g. easy to understand but grammatically inaccurate), and students whose performance is of different quality in the different parts of the interview (e.g. hesitancy and frequent grammatical errors in the beginning of the interview and fluency and accurate structures towards the end).

In order to be able to assess oral proficiency reliably, assessors need to agree on its most important characteristics. Discussions regarding oral proficiency could include the significance of accuracy, compensatory strategies and speech rate in assessment. As has been discussed, the results of the present study contain examples of student production which were given very different grades by the three interviewers. These examples of student production were obviously difficult to grade according to the rating criteria provided perhaps because they were felt to contain elements that the interviewers considered to be of differing degrees of severity. Therefore, authentic tests with different profiles of student performance need to be used in discussions and interpretations of rating criteria of oral proficiency in English. 9 Bedömningsmaterial språk, Skolverket, 2001, can be used for this purpose (= Material for the Assessment of Language, The National Agency for Education, 2001) as well as the National Tests of English since 1995, which contain useful discussions and valuable comments on examples of student performance and which are free for schools and teachers to use.

The third part of the suggested further training of English teachers would

9 In the early 1990s, material for in-service training was provided by the Foreign Language Teaching and Testing Research Unit, Gothenburg University, both for compulsory schools and for upper secondary schools. These video-tapes and the accompanying study material were ordered by many schools in Sweden.
include methods in carrying out oral proficiency tests. The students felt that in the oral proficiency interviews carried out by the Native Speakers they were, to some extent, more often given the opportunity to show their competence than in the other two interviews. It was certainly of importance that two of the three Native Speakers in the Gothenburg Project had been examiners for the Cambridge Examinations in Sweden for more than twenty years. Very experienced and well-trained examiners can enable students to show their oral competence in English better than other examiners can. Training of teachers in efficient ways of carrying out oral proficiency interviews should therefore be of great importance for two reasons: firstly, students who are weak in written English should be given the opportunity to show their oral competence and, secondly, school grades should not only be based on the students’ written, but also on their oral proficiency.

It has been pointed out several times that the results of the Gothenburg project gave evidence that even the nine very experienced examiners tended to give different grades to individual students’ production. It was also interesting that some differences could be observed between the members of the interviewer categories regarding the grades they assigned and the occurrence of certain linguistic features. One such difference was between School Teachers and Native Speakers. The School Teachers seem to have paid more attention to grammatical accuracy and the Native Speakers seem to have had a notion of communicative competence where accuracy plays a less important role. This finding should be a very fruitful starting point for a discussion of principles in the assessment of oral proficiency.

Evidently, many things have changed since 1993 when the material of the Gothenburg project was collected. Today, in 2003, this material nevertheless provides useful information on the assessment of students’ oral performance in English. In the assessment of language proficiency, oral testing may be more problematic than written testing but is by no means less important. It is to be hoped that the present study will contribute to the development and discussion of useful techniques in oral proficiency testing and to the improvement of the methods used for assessing oral proficiency in schools.


274
ducation Committee, Modern Languages Division, Strasbourg. Cambridge: Cambridge University Press.


281
— 1990. Spelar grammatiken någon roll i en kommunikativ språkundervisning? 
Kommunalmaterial Lgr 80. Undervisning i främmande språk. Stockholm: 
Studium.
Takala, S. 1984. Evaluation of students’ knowledge of English vocabulary in the 
University of Jyväskylä.
report. On TESOL ’77. Teaching and learning English as a second language. 
194–203.
— 1985. Variability in interlanguage use: a study of style-shifting in morphology 
and syntax. Language Learning 35: 373–403.
versity Press.
tics 11/1: 90–105.
Taylor, D. 1988. The meaning and use of the term “competence” in linguistics and 
Thagg-Fisher, U. 1986. The sweet sound of concord – a study of Swedish learners’ 
concord problems in English. Lund Studies in English 73. Lund: Gleerups.
Thorén, B. 1957. Mål, medel och metoder vid skolans språkundervisning. Lund: 
Gleerups.
Trollfjärden, C. and M. Swedberg. (Forthcoming.) Speak up! Malmö: Gleerups.
Press.
ich.
Váradi, T. 1983. Strategies of target language learner communication: message ad-
justment. Strategies in interlanguage communication. Ed. by C. Faerch and G. 
— 1983. The whole is more than the sum of the parts – evaluation criteria in oral 
sity.
ter: University of Exeter.
sity Press.
44.


Appendix
APPENDIX 1

The Organisation of the Project: The Numbers Given to the Tests

(Red Test=School Teacher, Blue Test=University Teacher, Green Test=Native Speaker).

<table>
<thead>
<tr>
<th>Student</th>
<th>Red Test</th>
<th>Blue Test</th>
<th>Green Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>76</td>
<td>31</td>
<td>70</td>
</tr>
<tr>
<td>B</td>
<td>77</td>
<td>32</td>
<td>71</td>
</tr>
<tr>
<td>C</td>
<td>78</td>
<td>33</td>
<td>72</td>
</tr>
<tr>
<td>D</td>
<td>79</td>
<td>34</td>
<td>73</td>
</tr>
<tr>
<td>E</td>
<td>80</td>
<td>35</td>
<td>74</td>
</tr>
<tr>
<td>F</td>
<td>81</td>
<td>36</td>
<td>75</td>
</tr>
<tr>
<td>G</td>
<td>82</td>
<td>37</td>
<td>68</td>
</tr>
<tr>
<td>H</td>
<td>83</td>
<td>38</td>
<td>69</td>
</tr>
<tr>
<td>I</td>
<td>84</td>
<td>39</td>
<td>67</td>
</tr>
<tr>
<td>J</td>
<td>85</td>
<td>40</td>
<td>66</td>
</tr>
<tr>
<td>K</td>
<td>86</td>
<td>41</td>
<td>65</td>
</tr>
<tr>
<td>L</td>
<td>87</td>
<td>42</td>
<td>64</td>
</tr>
<tr>
<td>M</td>
<td>88</td>
<td>43</td>
<td>63</td>
</tr>
<tr>
<td>N</td>
<td>89</td>
<td>43</td>
<td>62</td>
</tr>
<tr>
<td>O</td>
<td>90</td>
<td>45</td>
<td>61</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>60</td>
<td>16</td>
</tr>
<tr>
<td>Q</td>
<td>2</td>
<td>59</td>
<td>17</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>58</td>
<td>18</td>
</tr>
<tr>
<td>S</td>
<td>4</td>
<td>57</td>
<td>19</td>
</tr>
<tr>
<td>T</td>
<td>5</td>
<td>56</td>
<td>20</td>
</tr>
<tr>
<td>U</td>
<td>6</td>
<td>55</td>
<td>21</td>
</tr>
<tr>
<td>V</td>
<td>7</td>
<td>54</td>
<td>22</td>
</tr>
<tr>
<td>W</td>
<td>8</td>
<td>53</td>
<td>23</td>
</tr>
<tr>
<td>X</td>
<td>9</td>
<td>52</td>
<td>24</td>
</tr>
<tr>
<td>Y</td>
<td>10</td>
<td>51</td>
<td>25</td>
</tr>
<tr>
<td>Z</td>
<td>11</td>
<td>50</td>
<td>26</td>
</tr>
<tr>
<td>XX</td>
<td>12</td>
<td>49</td>
<td>27</td>
</tr>
<tr>
<td>YY</td>
<td>13</td>
<td>48</td>
<td>28</td>
</tr>
<tr>
<td>ZZ</td>
<td>14</td>
<td>47</td>
<td>29</td>
</tr>
<tr>
<td>WW</td>
<td>15</td>
<td>46</td>
<td>30</td>
</tr>
</tbody>
</table>
APPENDIX 2

The Distribution of the Interviewers in the Tests on 8 May, 1993

(ST=School Teacher, UT=University Teacher, NS=Native Speaker).

<table>
<thead>
<tr>
<th>Student</th>
<th>Red Test</th>
<th>Blue Test</th>
<th>Green Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ST1</td>
<td>UT1</td>
<td>NS3</td>
</tr>
<tr>
<td>B</td>
<td>ST1</td>
<td>UT2</td>
<td>NS1</td>
</tr>
<tr>
<td>C</td>
<td>ST1</td>
<td>UT3</td>
<td>NS2</td>
</tr>
<tr>
<td>D</td>
<td>ST1</td>
<td>UT1</td>
<td>NS1</td>
</tr>
<tr>
<td>E</td>
<td>ST1</td>
<td>UT2</td>
<td>NS2</td>
</tr>
<tr>
<td>F</td>
<td>ST1</td>
<td>UT3</td>
<td>NS3</td>
</tr>
<tr>
<td>G</td>
<td>ST1</td>
<td>UT1</td>
<td>NS2</td>
</tr>
<tr>
<td>H</td>
<td>ST1</td>
<td>UT2</td>
<td>NS3</td>
</tr>
<tr>
<td>I</td>
<td>ST1</td>
<td>UT3</td>
<td>NS2</td>
</tr>
<tr>
<td>J</td>
<td>ST1</td>
<td>UT1</td>
<td>NS3</td>
</tr>
<tr>
<td>K</td>
<td>ST2</td>
<td>UT2</td>
<td>NS1</td>
</tr>
<tr>
<td>L</td>
<td>ST2</td>
<td>UT3</td>
<td>NS2</td>
</tr>
<tr>
<td>M</td>
<td>ST2</td>
<td>UT1</td>
<td>NS3</td>
</tr>
<tr>
<td>N</td>
<td>ST2</td>
<td>UT2</td>
<td>NS2</td>
</tr>
<tr>
<td>O</td>
<td>ST2</td>
<td>UT3</td>
<td>NS3</td>
</tr>
<tr>
<td>P</td>
<td>ST2</td>
<td>UT1</td>
<td>NS1</td>
</tr>
<tr>
<td>Q</td>
<td>ST2</td>
<td>UT2</td>
<td>NS2</td>
</tr>
<tr>
<td>R</td>
<td>ST2</td>
<td>UT3</td>
<td>NS1</td>
</tr>
<tr>
<td>S</td>
<td>ST2</td>
<td>UT1</td>
<td>NS2</td>
</tr>
<tr>
<td>T</td>
<td>ST2</td>
<td>UT2</td>
<td>NS3</td>
</tr>
<tr>
<td>U</td>
<td>ST3</td>
<td>UT3</td>
<td>NS2</td>
</tr>
<tr>
<td>V</td>
<td>ST3</td>
<td>UT1</td>
<td>NS3</td>
</tr>
<tr>
<td>W</td>
<td>ST3</td>
<td>UT2</td>
<td>NS1</td>
</tr>
<tr>
<td>X</td>
<td>ST3</td>
<td>UT3</td>
<td>NS1</td>
</tr>
<tr>
<td>Y</td>
<td>ST3</td>
<td>UT1</td>
<td>NS2</td>
</tr>
<tr>
<td>Z</td>
<td>ST3</td>
<td>UT2</td>
<td>NS2</td>
</tr>
<tr>
<td>XX</td>
<td>ST3</td>
<td>UT3</td>
<td>NS3</td>
</tr>
<tr>
<td>YY</td>
<td>ST3</td>
<td>UT1</td>
<td>NS1</td>
</tr>
<tr>
<td>ZZ</td>
<td>ST3</td>
<td>UT2</td>
<td>NS3</td>
</tr>
<tr>
<td>WW</td>
<td>ST3</td>
<td>UT3</td>
<td>NS1</td>
</tr>
</tbody>
</table>
APPENDIX 3

The Organisation of the Tests on 8 May, 1993

<table>
<thead>
<tr>
<th>Time</th>
<th>Red1</th>
<th>Red2</th>
<th>Red3</th>
<th>Blue1</th>
<th>Blue2</th>
<th>Blue3</th>
<th>Green1</th>
<th>Green2</th>
<th>Green3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.00</td>
<td>A</td>
<td>K</td>
<td>U</td>
<td>Y</td>
<td>Z</td>
<td>C</td>
<td>W</td>
<td>E</td>
<td>M</td>
</tr>
<tr>
<td>10.20</td>
<td>B</td>
<td>L</td>
<td>V</td>
<td>D</td>
<td>N</td>
<td>X</td>
<td>WW</td>
<td>S</td>
<td>XX</td>
</tr>
<tr>
<td>10.40</td>
<td>C</td>
<td>M</td>
<td>W</td>
<td>A</td>
<td>E</td>
<td>U</td>
<td>K</td>
<td>Y</td>
<td>T</td>
</tr>
<tr>
<td>11.00</td>
<td>D</td>
<td>N</td>
<td>X</td>
<td>V</td>
<td>B</td>
<td>L</td>
<td>YY</td>
<td>Q</td>
<td>H</td>
</tr>
<tr>
<td>11.20</td>
<td>E</td>
<td>O</td>
<td>Y</td>
<td>G</td>
<td>K</td>
<td>R</td>
<td>I</td>
<td>C</td>
<td>ZZ</td>
</tr>
<tr>
<td>11.40</td>
<td>BREAK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.00</td>
<td>F</td>
<td>P</td>
<td>Z</td>
<td>YY</td>
<td>ZZ</td>
<td>WW</td>
<td>B</td>
<td>L</td>
<td>J</td>
</tr>
<tr>
<td>12.20</td>
<td>G</td>
<td>Q</td>
<td>XX</td>
<td>S</td>
<td>T</td>
<td>I</td>
<td>D</td>
<td>N</td>
<td>V</td>
</tr>
<tr>
<td>12.40</td>
<td>H</td>
<td>R</td>
<td>YY</td>
<td>J</td>
<td>W</td>
<td>F</td>
<td>P</td>
<td>U</td>
<td>O</td>
</tr>
<tr>
<td>13.00</td>
<td>I</td>
<td>S</td>
<td>ZZ</td>
<td>M</td>
<td>Q</td>
<td>XX</td>
<td>X</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>13.20</td>
<td>J</td>
<td>T</td>
<td>WW</td>
<td>P</td>
<td>H</td>
<td>O</td>
<td>R</td>
<td>Z</td>
<td>F</td>
</tr>
</tbody>
</table>

The tests are over at 13.40.

Tasks in the tests:
Task 1 (to entertain): E1, E2, E3
Task 2 (to inform): I1, I2, I3
Task 3 (to argue a case): A1, A2, A3

Tasks in oral tests assessed by School Teachers (Red): E1, I1 and A1
Tasks in oral tests assessed by University Teachers (Blue): E2, I2, A2
Tasks in oral tests assessed by Native Speakers (Green): E3, I3, A3
APPENDIX 4A

The Instructions and Tasks of the Students in Tests with School Teachers

INSTRUCTIONS FOR THE STUDENT

You are going to take part in a test in which you are expected to speak English. The test has **three parts**. You have about ten minutes to prepare for what you are going to talk about in the **first two** parts.

**Important! You should do most of the talking! Don’t wait for your teacher to ask questions. What you say will be recorded on a cassette player.**

**Part 1** Look at the six pictures on the next page. You will be asked to tell your teacher about the car mechanic, Mr Brown.

– Describe what happened to Mr Brown last Friday.
– How did he manage to get out of the cinema? Try to work out a solution!

**Part 2** You have been given three topics. Choose **one**! You will be asked to tell your teacher (who represents a person who does not know very much about the subject) as much as you can about it. Try to make as interesting as possible!

*Please also read the information below about Part 3.*

**Part 3** Your teacher will present three topics to you. You must choose one and argue for it as best you can. Your teacher will play the role of a journalist. Remember that you should do most of the talking.
UPPGIFT 2

1. Midsummer in Sweden
2. An important meeting in my life
3. A book I would like to read again

UPPGIFT 3

1. It’s necessary to have marks at school
   or
   it’s not necessary to have marks at school
2. The age for getting a driving license should be lowered
   or
   The age for getting a driving license should be raised
3. Using animals in scientific experiments is necessary for progress
   or
   Using animals in scientific experiments should be prohibited.
APPENDIX 4B
The Instructions and Tasks of the Students in Tests with University Teachers

INSTRUCTIONS FOR THE STUDENT

You are going to take part in a test in which you are expected to speak English. The test has three parts. You have about ten minutes to prepare for what you are going to talk about in the first two parts.

Important! You should do most of the talking! Don’t wait for your teacher to ask questions. What you say will be recorded on a cassette player.

Part 1 You have friends in England – a family of four - who have decided to rent a furnished summer house in Sweden for two weeks. You have promised to help them find a house and you have two suggestions. You are expecting a phone call from one of these friends who wants to learn about what you have found.

Your interviewer is acting as your Englishfriend. Your task is to describe the two houses and the surroundings with the help of the drawings on the next page. You have been to see both houses.

– First you describe the modern house: the inside of the house itself, the garden, the surroundings, the cost etc.
– Then you do the same thing with the old house.

When you have met the interviewer, the telephone conversation begins!
You start the conversation.

Part 2 You have been given three topics. Choose one! You will be asked to tell your teacher (who represents a person who does not know very much about the subject) as much as you can about it. Try to make as interesting as possible!

Please also read the information below about Part 3.

Part 3 Your teacher will present three topics to you. You must choose one and argue for it as best you can. Your teacher will play the role of a journalist. Remember that you should do most of the talking.
The modern house

3900 Swedish kronor a week

The old house

1600 Swedish kronor a week
UPPGIFT 2

1. New Year celebrations in Sweden/in my homeland
2. An elderly person that I like
3. A memorable journey

UPPGIFT 3

1. It is better to start a family early in life
   or
   It is better to start a family when you are older

2. The lower age-limit for inter-railing should be twenty-one
   or
   The lower age limit for inter-railing should be sixteen

3. Adults have the same ability to learn new things as young people have
   or
   Adults don't have the same ability to learn new things as young people have
APPENDIX 4C

The Instructions and Tasks of the Students in Tests with Native Speakers.

INSTRUCTIONS FOR THE STUDENT

You are going to take part in a test in which you are expected to speak English. The test has three parts. You have about ten minutes to prepare for what you are going to talk about in the first two parts.

Important! You should do most of the talking! Don’t wait for your teacher to ask questions. What you say will be recorded on a cassette player.

Part 1

Look at the nine pictures on the next page. You will be asked to tell your interviewer about the two boys, Pete and Dave.

– Describe what happened to the two boys last Saturday.

Part 2

You have been given three topics. Choose one! You will be asked to tell the interviewer as much as you can about it. The interviewer does not know very much about the subject. Try to make it as interesting as possible!

Please also read the information below about Part 3.

Part 3

Your teacher will present three topics to you. You must choose one and argue for it as best you can. Remember that you should do most of the talking.
A Dangerous Adventure

1

2

3

4
UPPGIFT 2

1. The way we celebrate a wedding in Sweden/my homeland
2. My hobby
3. What my home area has to offer a tourist

UPPGIFT 3

1. Keeping up with new trends in clothing styles is important
   or
   Keeping up with new trends in clothing styles is not important

2. Equality between men and women has come a long way, especially in Sweden
   or
   Equality between men and women has not made very much progress at all, not even in Sweden

3. Sweden would gain a lot by being host to the Olympic Games
   or
   Sweden would not gain a lot by being host to the Olympic Games
INSTRUCTIONS FOR THE INTERVIEWER

You are going to lead an exam of oral proficiency in English. The aim of the exam is to make the student speak as much as possible and speak more or less freely.

There are three parts in the exam:

Part 1 With the help of some pictures the student is to tell a story or take part in a telephone conversation about two houses.

Part 2 The student is to inform about something you are supposed to know very little about. The student chooses one out of three topics.

Part 3 The student is to argue a case. The student chooses one out of three cases and you are to help the student by taking the other standpoint and have a discussion.

Try to make the student speak as much as possible so that you have enough language to be able to assess the student’s oral proficiency.

The exam should take approximately 10 minutes and not more than 15 minutes.

Immediately after the exam you should make your assessment on the Oral Mark Sheet.
APPENDIX 6

The Questionnaire to the Interviewers

1. What is your opinion in general about the tasks the students were given in this oral proficiency interview? Did they get the opportunity to show – for better or for worse – their proficiency in English?

2. Did one of the three tasks in the interview turn out to give more information for the assessment than the others? If so, which one? Do you think that the tasks could have been changed to the better in some way? Could one of the tasks have been left out?

3. What is your reaction to the rating criteria? Are they clear and useful?

4. What aspects of the English language and what linguistic variables played the most important role when you made your assessment of the oral proficiency interviews? What was of importance in the student’s oral proficiency when you gave a good mark? Are there any crucial linguistic variables that lead to a good assessment? Try to describe, for example, five aspects that you noticed and that you think are important.

5. What is especially disturbing in some students’ spoken English? What is there in particular in the language that makes you give a low rating? Are there any strategies which are very characteristic of low-achievers? Are there any linguistic errors that you noticed and that you think are significant for the assessment?
APPENDIX 7

The Questionnaire to the Students

(The questions were in Swedish in the hand-out to the students; my translation.)

1. Name and class at your school?
2. Boy or girl? What year were you born?
3. Programme studied at gymnasieskolan?
4. Your three test numbers during the day?
5. Your grade in English at school?
6. Your grades in Swedish at school?
7. Other languages you study now? Grades?
8. How many weeks/months have you been in English-speaking countries (including holidays, language courses, school journeys)?
9. How many weeks/months have you been in other countries abroad than Denmark, Norway, Finland and Iceland?
10. Have you participated in a summer language course arranged by for instance EF, STS, SIS or other organisation?
11. Have you ever read an English book in your spare time?
12. Do you have the possibility to watch an English or American TV-channel such as CNN or BBC at home?
13. How many times a month on average do you think that you have seen a film (TV, video or cinema) in English the last year?
14. Do you think that you were given the opportunity to show what you know in the conversations today?
15. In the conversations today there were three tasks, the first one was intended to be the easiest one, the third task was intended to be the most difficult one. Was one of the tasks good or bad?
16. Was one of the interviewers “easy to talk to” and made you do justice to yourself? Was one of the interviewers “difficult to talk to” which made you say less that you had been able to do?
17. Other ideas and opinions on oral proficiency testing and the structure of the tests used today that you would like us to know?
The Rating Criteria

1. Global Evaluation:
This is the first and overall impression of the language produced by the learner. The assessment is given on a five-step scale.
A – Distinction: A fluent and overall correct use of English.
B – Very Good: Very good production with minor inaccuracies.
C – Pass: Acceptable language in spite of errors.
D – Not enough: Not showing enough to be acceptable.
E – Poor: Poor production of English in many respects.

2. Factorial Grading:
Three different factors are to be considered in the assessment, namely
2a Communication and Fluency
2b Grammar and Vocabulary
2c Pronunciation

<table>
<thead>
<tr>
<th>2a Communication and Fluency</th>
<th>2b Grammar and Vocabulary</th>
<th>2c Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C Easy to understand. Takes some initiatives. Sometimes produces longer and fairly coherent flow of language.</td>
<td>Some inaccuracies and a few noticeable basic errors but in general a good command of structure and vocabulary. Range of vocabulary adequate.</td>
<td>Good pronunciation of most of the individual sounds and acceptable intonation. Foreign accent evident.</td>
</tr>
<tr>
<td>D Initiatives are rare and contributes little. Needs help.</td>
<td>Structures limited in range. Basic errors are noticeable. Dependent on paraphrase of simplification. Shows a limited vocabulary resource.</td>
<td>Individual sounds poorly articulated. Some utterances difficult to understand, putting strain on the listener.</td>
</tr>
<tr>
<td>E Frequent hesitations disrupt the flow of language. Speech can be difficult to follow. Constantly needs help.</td>
<td>Makes many kind of mistakes. Very limited range of vocabulary. Frequent basic errors. Sometimes uses Swedish words.</td>
<td>Difficult to understand due to pronunciation. First language intonation. Poor articulation of individual sounds.</td>
</tr>
</tbody>
</table>
APPENDIX 9
The Oral Mark Sheet
ORAL MARK SHEET

Candidate’s name: ………………………………….
Candidate’s number: ………………………………
Marker’s name: …………………………………….

Do you know the candidate? Yes …. No ….

1. Global Evaluation A   B   C   D   E
   A – Distinction
   B – Very Good
   C – Pass
   D – Not Enough
   E – Poor

This assessment is to be given during or directly after the oral examination.

………………………………………………………………………………

2. Factorial Evaluation

This assessment can be made during and/ or after the oral examination
according to the rating criteria given.

a) Communication and Fluency: A   B   C   D   E
b) Grammar and Vocabulary: A   B   C   D   E
c) Pronunciation: A   B   C   D   E

Comments: (e.g. body language):
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
APPENDIX 10

The Length of the Tests

<table>
<thead>
<tr>
<th>Student</th>
<th>Red tests</th>
<th></th>
<th>Blue tests</th>
<th></th>
<th>Green tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test Length</td>
<td>Test Length</td>
<td>Test Length</td>
<td>Test Length</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>76 11.10</td>
<td>31 9.20</td>
<td>70 10.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>77 13.55</td>
<td>32 *</td>
<td>71 10.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>78 9.40</td>
<td>33 9.50</td>
<td>72 17.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>79 11.45</td>
<td>34 9.30</td>
<td>73 11.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>80 10.30</td>
<td>35 *</td>
<td>74 11.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>81 8.45</td>
<td>36 11.55</td>
<td>75 12.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>82 9.20</td>
<td>37 10.35</td>
<td>68 13.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>83 10.25</td>
<td>38 16.00</td>
<td>69 10.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>84 9.55</td>
<td>39 13.10</td>
<td>67 11.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>85 18.00</td>
<td>40 10.15</td>
<td>66 11.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>86 9.45</td>
<td>41 *</td>
<td>65 11.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>87 9.50</td>
<td>42 11.00</td>
<td>64 13.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>88 9.10</td>
<td>43 10.20</td>
<td>63 9.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>89 11.40</td>
<td>44 *</td>
<td>62 16.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>90 11.30</td>
<td>45 15.55</td>
<td>61 12.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>92 12.30</td>
<td>59 17.55</td>
<td>17 17.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>93 10.25</td>
<td>58 10.50</td>
<td>18 10.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>94 8.45</td>
<td>57 9.10</td>
<td>19 12.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>95 14.55</td>
<td>56 12.15</td>
<td>20 12.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>96 11.00</td>
<td>55 11.20</td>
<td>21 15.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>97 12.10</td>
<td>54 11.55</td>
<td>22 12.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>98 8.20</td>
<td>53 11.20</td>
<td>23 10.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>99 11.15</td>
<td>52 9.35</td>
<td>24 12.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>10 12.45</td>
<td>51 9.57</td>
<td>25 18.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>11 10.35</td>
<td>50 *</td>
<td>26 13.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>12 12.10</td>
<td>49 11.55</td>
<td>27 10.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZZ</td>
<td>13 12.00</td>
<td>48 11.00</td>
<td>28 10.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WW</td>
<td>14 11.50</td>
<td>47 20.25</td>
<td>29 11.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = The test was not recorded due to technical reasons.
APPENDIX 11

The Number of Students’ Words in the Tests

<table>
<thead>
<tr>
<th>Student</th>
<th>Red tests</th>
<th>Blue tests</th>
<th>Green tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Words</td>
<td>Test</td>
</tr>
<tr>
<td>A</td>
<td>76</td>
<td>740</td>
<td>31</td>
</tr>
<tr>
<td>B</td>
<td>77</td>
<td>860</td>
<td>32</td>
</tr>
<tr>
<td>C</td>
<td>78</td>
<td>910</td>
<td>33</td>
</tr>
<tr>
<td>D</td>
<td>79</td>
<td>720</td>
<td>34</td>
</tr>
<tr>
<td>E</td>
<td>80</td>
<td>970</td>
<td>35</td>
</tr>
<tr>
<td>F</td>
<td>81</td>
<td>560</td>
<td>36</td>
</tr>
<tr>
<td>G</td>
<td>82</td>
<td>830</td>
<td>37</td>
</tr>
<tr>
<td>H</td>
<td>83</td>
<td>700</td>
<td>38</td>
</tr>
<tr>
<td>I</td>
<td>84</td>
<td>570</td>
<td>39</td>
</tr>
<tr>
<td>J</td>
<td>85</td>
<td>1100</td>
<td>40</td>
</tr>
<tr>
<td>K</td>
<td>86</td>
<td>970</td>
<td>41</td>
</tr>
<tr>
<td>L</td>
<td>87</td>
<td>790</td>
<td>42</td>
</tr>
<tr>
<td>M</td>
<td>88</td>
<td>1030</td>
<td>43</td>
</tr>
<tr>
<td>N</td>
<td>89</td>
<td>1040</td>
<td>44</td>
</tr>
<tr>
<td>O</td>
<td>90</td>
<td>1220</td>
<td>45</td>
</tr>
<tr>
<td>Q</td>
<td>92</td>
<td>850</td>
<td>59</td>
</tr>
<tr>
<td>R</td>
<td>93</td>
<td>790</td>
<td>58</td>
</tr>
<tr>
<td>S</td>
<td>94</td>
<td>620</td>
<td>57</td>
</tr>
<tr>
<td>T</td>
<td>95</td>
<td>1400</td>
<td>56</td>
</tr>
<tr>
<td>U</td>
<td>96</td>
<td>910</td>
<td>55</td>
</tr>
<tr>
<td>V</td>
<td>97</td>
<td>1330</td>
<td>54</td>
</tr>
<tr>
<td>W</td>
<td>98</td>
<td>840</td>
<td>53</td>
</tr>
<tr>
<td>X</td>
<td>99</td>
<td>850</td>
<td>52</td>
</tr>
<tr>
<td>Y</td>
<td>100</td>
<td>1600</td>
<td>51</td>
</tr>
<tr>
<td>Z</td>
<td>101</td>
<td>1060</td>
<td>50</td>
</tr>
<tr>
<td>XX</td>
<td>102</td>
<td>930</td>
<td>49</td>
</tr>
<tr>
<td>YY</td>
<td>103</td>
<td>1130</td>
<td>48</td>
</tr>
<tr>
<td>ZZ</td>
<td>104</td>
<td>650</td>
<td>47</td>
</tr>
<tr>
<td>WW</td>
<td>105</td>
<td>1090</td>
<td>46</td>
</tr>
</tbody>
</table>

* = The test was not recorded due to technical reasons.
All figures are rounded to tens.
APPENDIX 12

Illustrative Samples of Three Oral Tests

In this appendix the texts from three oral tests are provided for illustration. The three tests were selected according to the following criteria.

– Girls’ and boys’ performance: Tests M63 and U6 are with girls and test C33 with a boy.
– Different ratings: Test M63 was assigned a high Global rating (A), test U6 was assigned an average Global rating (C) and test C33 was assigned a low Global rating (D).
– The three interviewer categories: Test U6 is with a School Teacher, test C33 is with a University Teacher and M63 is with a Native Speaker.
– The quality of the tape recordings: the three tests were recorded with good sound quality.
– The length of the tests: the length of the three tests is close to the stipulated 10 minutes in the instructions (9.50, 9.55 and 11 minutes long).

Test M63

Interviewer: Would you like to come in?
Student: Thank you.
I: will you leave this for the next one? You can bring that if you want, yes yes okay right would you like to sit in that chair?
S: yeah thank you.
I: excuse me I’m just going through okay now sit down.
S: yeah thank you.
I: I’m Peter, what’s your name?
S: Anna-Karin
I: Anna-Karin right okay now you’ve read the instructions okay?
S: yes
I: and you are
S: Green
I: eh yes green sixty
S: sixty-three.
I: sixty-three. So you are number sixty-three, I’ll simply sign this and
S: yeah
I: then you know you’ve been here
S: Yes
I: okay I put Peter on there, so shall we start with telling me about the pictures.
S: alright
I: here they are
S: yeah we have then Tom and Pete
I: okay
S: who were planning eh a day on the beach with picnic
I: hm
S: and they got to the beach and they sat down they ate eh had a lot of fun and suddenly they saw this toy airplane aeroplane on a shelf on the
I: aha
S: on a rock big rock
I: yes
S: and they thought well why don’t we go and get it
I: Mm
S: so they climbed up and they got hold of the plane and then they realized that they weren’t able to get down
I: aha
S: and this isn’t a good place to stay
I: I see
S: So they screamed for help and eh there were some other people on the beach and they saw them and a lady in a
I: yeah turn over
S: a fat lady (laughter) in a black swimsuit she saw them and she went to a phone and she called for help she called the probably police or coast guard
I: yes
S: something
I: yes
S: and they came with a helicopter and they eh took the boys down from the shelf
I: aha
S: and then well so they are grateful says thanks a lot to the man who rescued them and to the lady who eh had put on a dress over her
I: oh yes
S: yeah
I: good well I think you’ve told me everything that’s fine that’s right now
S: now
I: number two what are you going to tell me about it
S: as wha what my home area has to offer
I: okay
S: a tourist at all
I: yes
S: yeah well my home area I see that as Gothenburg and the archipelago
I: archi yes yes
S: and eh Gothenburg is the second largest city in Sweden so when it comes to stores and eh amusement all that entertainment
I: right
S: we’ve got quite a lot of that
I: mm
S: quite a few theatres to go to and eh there are the big amusement park Liseberg
I: aha
S: of course who’s quite famous in Europe it’s quite a big park
I: can you tell me a bit about it
S: yes it’s a sort of like a Disney concept in a on small scale
I: yes
S: we have these little green and pink rabbits instead
I: do you mean people walking around
S: yeah people walk around as rabbits sort like they do it in Disney too
I: yes
S: they have people dressed up as Mickey Mouse
I: Yes
S: sort like a logo
I: aha
S: and there are of course roller-coaster rides
I: yes
S: there are eh eh what can you see more you can eh fortune wheels you can sort of play bets
I: oh I see right
S: and then you can see a number comes up
I: right
S: win chocolate or little bunnies
I: what
S: little bunnies or something like that and eh there are also a big stage there where usually eh it’s like groups comes music groups famous in Sweden
I: yes pop groups
S: pop groups and eh also I think they usually have some paly for eh little amusements eh for older people not just it’s not just pop acts it’s
I: no
S: they sort of like tend to try to get to something to suit everyone
I: yes yes
S: and eh it’s quite beautiful you can walk in the park it’s not just amuse-
ment they have sort like flowers and a garden
I: restaurant?
S: a restaurant too yes
I: yes
S: so it’s a little bit for everyone
I: do you know what it costs to go in?
S: I don’t
I: roughly
S: I think it is last year I think it is twenty-five Swedish crowns maybe they raised it or it is thirty-five now
I: yeah if I just got one day in Gothenburg what would you say I really must do?
S: eh I would take you to a big park of ours the only really big one called Slottsskogen
I: aha
S: and I think that’s a very beautiful place it’s eh for free
I: right
S: you can can the entrance is is for free you could just walk around with eh a lot of trees a lot of say animals you got animals too that’s fun for the kids
I: is a zoo is it?
S: yeah it’s a zoo too
I: yeah
S: it’s sort like a big park you can just walk around you could take a some play baseballs with your friends or something and then they are this there is a zoo for children but I like to see it I’m not I’m not that young but I like to see it too it’s the seals it is the main attraction there
I: (laughter) they are rather nice aren’t they?
S: okay they are rather nice
I: okay let’s leave a bit of time
S: yeah
I: for you to eh do the last thing ah you don’t know what that’s it do you
S: no
I: let me see if I’ve got it here eh somewhere eh that’s number two here we are now you got to choose one of these
S: yeah
I: and then we have a discussion about it
S: alright
I: but you choose which one you want to do
S: let’s see now I think I’ll take number eight
I: you take number eight
S: equali
I: equality between men and women are you going to say it has come a long way or it hasn’t
S: I think I will say it hasn’t it hasn’t made very much progress
I: okay okay you tell me a bit about it
S: you are always told that it has made a big progress
I: yeah
S: and maybe in Sweden it’s bigger than in some other places but
I: bigger than England yeah
S: yeah but still it’s quite a lot of difference between men and women the way they are treated in society
I: give me some examples
S: if you say school for example that’s a good one
I: oh yes that’s a good one
S: eh that’s the one I know about (laughter)
I: right
S: and eh well boys boys and girls they are different from the start in a way of acting girls is sort like they are expected to be a bit more quiet and behave themselves a bit more and that could be because girls eh they mature a bit faster than boys
I: do you think that’s true
S: yeah I think that that is true
I: yeah yeah okay
S: I mean we all catch up
I: yes
S: but when we talk about like seven-year-olds and eight-year-olds I mean then it’s it’s a quite a big difference
I: yes yeah
S: and boys are allowed to be a bit more mischievous they can play around they can eh scream a bit more without being treated as there are eh that as problem but if girls do that they are there are problem at once you know
I: I see
S: you are not supposed to behave like that
I: right yeah what do you think has still to be done then?
S: I think eh
I: from what you say this problem starts at home
S: yes it starts at home but it’s eh I think I think it’s not just at home it’s at the school system too because they allow boys to be a bit more outspoken and eh a bit more loud
I: they who is they you have two sorts of teachers men teachers and women teachers
S: no it’s it’s just the
I: no difference
S: no it’s I don’t think it’s a difference at that age it’s everyone sort like says it’s okay and that there is the problem later on when boys tend to continue to do that
I: mm
S: and girls sort like have to they don’t get the same attention
I: no right
S: and I think that’s why boys can advertise themselves more later when it comes to jobs and they get a bigger pay for the same work as the
I: do they is that allowed?
S: yes
I: not in Sweden surely
S: no if you say it’s eh an what do you say a boss a big boss or a company
I: yeah
S: eh if that is a man it tend to be the men who have more pay
I: I see
S: for that post and that that is either because they are regarded more highly but that isn’t very likely because it’s almost the same job
I: yes
S: or they are better to advertise themselves
I: yes
S: and then we have the problem with women they someone has to take care of the children when they get ill but that’s usually the woman
I: it’s still the woman is it even in Sweden
S: yes even in Sweden
I: right I’m going to stop there because although it’s very interesting thank you very much we’ve run out of time
S: yes
I: right nice to talk to you
S: yeah
I: what was it Anna-Karin or?
S: Anna-Karin
I: Anna-Karin right thanks for coming
S: thank you
I: now then you are not taking you’re not taking anything which got to stay in here are you
S: no
I: no that’s okay
S: not that one
I: good bye
S: good bye

Test U6

Interviewer: hello Nadja
Student: hello
I: this is student number six okay go ahead
S: yes I should eh tell you about the pictures
I: yes
S: well the first one is an Mrs eh Mr brown I think he is a car mechanic
I: can you speak up a bit?
S: he is a car mechanic
I: yeah
S: and he works eh with his partner Gary and this Friday there isn’t much to do so he fix an engine and eh then he goes home yes
I: mm
S: and eh picture number two is when Mr Brown is out running he likes to keep in shape and eh it’s early and he goes out in the forest to run a bit and eh then Gary doesn’t really know what to do he goes out in the town and there he sees a poster which shows a film Al Capone II and he has seen the first part of it so he likes he he think that he should see the second one too eh picture number four is when he is in the saloon at the cinema and it’s very exciting but he is very tired because had worked late all the week long and ja he is a bit tired but he then he fell asleep
I: that’s right
S: and eh yes the people the film finished and the people go home walk out and he doesn’t really see that so he continue to sleep and when he suddenly wakes up of a strange noise there is a eh cockroach at his eh doing some sound he wakes up he is terrified and when he sees that all the people has gone he really gets scared and he had he noticed also that he is locked in eh
I: that’s right so what happens
S: he eh screams really loud and there is a security guard that come along and wonder what what’s that noise is and a security guard enters the cinema saloon and eh he see the terrified Mr Brown on the floor crying and eh so he brings eh Mr Brown poor
I: end of story?
S: yes
I: yeah we have a solution to this problem another solution
S: yes
I: it’s gone in that he finds a cellular phone you know what that is?
S: mm
I: and he phones home or the police or whatever and that’s that’s is the solution so what do you think about these solutions which one is the best?
S: my solution of course (laughter)
I: why
S: eh eh then he can that eh this shows how eh how nice people are around that people help one another and eh that you can rely on a on every every one around you and that’s a nice thought I think
I: yeah okay let’s go into the next subject you have read it haven’t you
S: yes
I: so you either are to talk about the Midsummer in Sweden or an important meeting in my life or a book I would like to read again which one do you choose?
S: I think I’ll choose number three a book I would like to read again
I: mm
S: eh it’s a book called Castaneda and Castaneda is the writer and he is a eh professor in social anthropology I think it’s called and he was in Mexico to study about eh eh magicians old Indian magicians magicians
I: mm
S: and eh it isn’t just one book eh there are three three or four books I think
I: mm
S: or more I don’t really know and this is they are very exciting and you can follow his eh ex experiences and adventure in all these books eh eh Castaneda meets magician called Don Juan it isn’t his real name but they call him so in the book and he has a lot of power he eh knows how to become a magician and he takes Callows eh Castaneda as a pupil or whatever you can say eh and Castaneda doesn’t know anything about this world about this magic world that exists really and he learns very much but also loses very much because he he have to do many things that he isn’t used to and his family also rejects him a bit because they don’t like this kind of studies and eh eh it’s very eh interesting to read about this because I think it’s based eh on really happenings but eh but there I don’t think that it’s all reality in the book it eh fantasy mixed with reality and yes so it isn’t a real story but eh eh it becomes very real to you when you read it and there are many things eh in eh in the book that you can eh try that you can think of that exist in your real life
I: mm
S: and that touched many subjects and happenings you can find in your own world but not described that they are described differently in the book
I: mm yes
S: mm
I: yes I’ve never heard about it let’s have another subject somewhere you know about the third subject that you are going to talk about?
S: no
I: eh what we are going to discuss now is this you can read it through
S: mm
I: which one would you like to talk about
S: the first one
I: the first one okay
S: eh I don’t think it’s necessary to have marks in en school while eh it eh they are not really justified
I: mm
S: eh you can’t really note a person I think because give him a mark that you are you are like this you are or eh people are people are individuals and they have many sides that that are good or bad or different
I: mm
S: and you can’t say that you are you are this really you are four five eh
I: but don’t you think people will work less do less if they don’t
S: no I think you should have three marks eh qualified or not qualified or very good qualified
I: mm
S: you can say and eh you do what you have to do in the school and then you get qualified
I: mm
S: and if not you don’t get qualified and you can’t continue and everyone has to work for being qualified
I: mm
S: and when everyone are that it’s good there isn’t any bad about it and eh
I: so you think that is enough? You don’t think that people will do less and that they won’t do enough?
S: no I think that everyone will do as much as they do now
I: you think so?
S: yes and what what the good thing about it is that eh eh pupils maybe pupils will be happier and not think about the grades so much and eh that eh there won’t be anger and eh disappointment in in the school they see they just have to do they just have to do qualified and then they can do whatever they want and eh I don’t mean that they won’t study they would of course study it’s interesting to study for them and they can choose whatever they want to give their energy to and put put energy to
I: mm
S: and
I: yeah okay I think this enough Nadja
S: mm
I: ah thank you very much
S: yes

Test C33
Student: sit down here or
Interviewer: yes have a seat here
S: okay yes madam thank you
I: right sit there no that seat there you might
S: okay
I: so that good picture of me when perhaps a little bit more
S: yes
I: well it’s very nice of you to come here on a Saturday I think and sacrifice yourself for us
S: oh I will try to do my best
I: (laughter) okay nice to see you anyway ehm yeah you are the one to do the talking okay so it’s right
S: okay
I: I’m going to try to be as quiet as possible ahh
S: okay
I: well should we do the first thing then
S: yes
I: that you could tell eh eh something about houses
S: yes
I: here in Sweden and it’s also would be a telephone and I will ring and the conversation will begin so I can be in the telephone now alright ring ring off you go
S: hello it’s Ulf here
I: hello this is Moira calling from England have you have you found us any houses?
S: yes I have found two eh interesting houses in Sweden it’s eh quite different from each other and eh I think some of you some of them will suit you and the first one is a big house in Stockholm ehh near the big city and eh well it’s it’s a perfect eh eh
I: yes
S: läge oh
I: perhaps situation
S: oh situated yes eh it’s quite near a big market and you can see the city from there and you can go there and shopping and so on and in your spare time you can go bathing because it’s just around the sea and so on the house is also very nice it’s eh got all that you can ever wants in the house it’s television and so on and eh yeah it’s near the city it perhaps it can be lots of cars outside and so on but ja it was the best I could find near a city
I: aha but we could go bathing there too
S: yes yes
I: it’s both in the city and near bathing so it’s very good
S: yes
I: okay what what does it cost?
S: eh it cost three thousand nine hundred Swedish crowns a week
I: three thousand nine hundred
S: yes
I: Swedish crowns

314
S: I found a lot cheaper one

I: oh I’m glad

S: yes but his house is eh eh quite a bit far away from Stockholm if you can say so it’s eh it’s in the middle of Sweden and it’s situated with with a lake at eh a very typical lake for Sweden with a boat you eh you can have a boat and go out fishing and so on and it’s two houses one eh one small one and a little bit bigger

I: mm

S: eh you have not eh the same eh so much things there you can have to go to take your own water and the bathroom is outside you can say

I: outside

S: yes

I: thank God it’s no winter time when I’ll come there

S: no but it’s summer in Sweden it can be good

I: aha

S: and I think it’s in eh a good area

I: aha

S: with eh lots lots of woods around and so on

I: aha right what did you say it cost

S: it cost one thousand six hundred Swedish crowns a week

I: well that sounds a lot more reasonable doesn’t it?

S: yes can al… you can also take the car to Stockholm if you want that car trip it will perhaps stay three four hours so you can se both

I: yeah okay I think the old house sounds pretty good

S: yeah

I: it’s about what we can afford so thanks very much then for all your work on my behalf so if you can fix that one for us would make us happy thank you

S: yes I will try to do that

I: okay great oh thank you (laughter) for that part eh fine and the if we now go on to the next eh task

S: yes maybe

I: it is for you to describe something to me

S: yes

I: what did you

S: eh I have chosen number six

I: number six ehm

S: and I will try to describe what I did last summer when I got on Interrail

I: ahm

S: and I travelled around Europe eh in a month time and I start to go by if you go by Interrail you eh buy a card in Sweden there you can go free in all train in all train ways trainways all over Sweden eller all over Europe
I: *ahm*
S: free and eh I did it with my best friend and we were out to in a month time and first we went to Paris eh when we got on further down to Riviera France Riviera and where we stayed eh some days eh we eh visit eh eh visit eh eh Cha… the Chagall Museum this was very nice eh and we went on go to Venedig and stayed there some days and then we travelled all through Italy and from Italy we took a boat over to Greece and in Greece we were in Athen at the first and we went doing everything a tourist should do in Athen then we took a boat out far away out in the sea out to a little island called eh Santorini a very nice little island with not so much tourists and so on but very nice
I: *ahm*
S: and this have taken approximately three weeks to get all the way through Europe and eh we began to run out of money we have to began to travel back and eh we did so in a two days time eller or two and half day and night just going by train all the way home
I: *ahm*
S: but that was one of the most lovely thing I’ve ever done in my whole life
I: *oh sounds absolutely marvellous*
S: yes
I: *did you bring enough of money with you so you could afford to eat as well*
S: yes we we eat a lot
I: *sounds as if you ran out of money in southern Europe, you have starved (laughter)*
S: no that was eh not so difficult because when we ran out of money we had okay lots so much money that we could get home alright
I: *alright good yeah sounds like a marvellous trip*
S: yes
I: *very good*
S: yes
I: *and in a way that connects to maybe one of the questions that you could be eh want to discuss eh in the third part of your task have you*
S: yes
I: *and which are we are going to argue or discuss rather not argue (laughter)*
S: mm
I: *and eh it the questions suggested are here so either we could talk about eh it’s better to start a family early in life or you know which one you prefer up there*
S: okay
I: *or the lower age limit for inter-railing should be twenty-one or the lower age limit for inter-railing should be sixteen and with all your experience*
you have something to say about that adults have the same ability to learn new things as young people have or adults don’t have the same ability to learn new things as young people have so what you have to do is to choose one of these

S: I think I choose the first one
I: the first one
S: I prefer that
I: have you got lots of lots of about that okay you think it’s better to start a family early in life
S: I think it’s best to do something in the middle
I: ah
S: eh because I I think it’s good when you are perhaps twenty-five not too early and not too late
I: mm
S: when you are older you have to live your life and you travel a lot in the beginning and then you can make your education
I: ahm
S: and then I think it’s better to settle down and take it easy and get kids and so on
I: I wonder if you are somewhere in between I wonder if that is early in life or later in life or in between eh you know in Britain they think if you if you are not finished with your qualifications in the university when you are twenty-one then you are really pretty old the problem you think it’s better to wait until you are a little bit older okay
S: yeah
I: eh on the other hand if you are waiting till you are a bit older then perhaps you are older when your children eh you know sort of argue
S: yeah
I: wouldn’t you find that a bit tiring and so on so if you imagine when you are sort of about forty years old and you’ve got teenagers around the house
S: I don’t think that will be any trouble
I: you don’t think so
S: no I don’t think so
I: you you think teenagers are are you know they are not so hard to deal with
S: yeah yes I think it’s hard to be with but I think it’s sort of fun and so on and can ja
I: mm mm
S: it’s cou… give you a kick I think
I: ahm keep you young perhaps
S: what?
I: keep you young perhaps
S: yes
I: *when you are bit older (laughter) okay so what will be you think there are some disadvantages of having a family too early*
S: *eh*
I: *in what wha... wha... what would that I mean if you had children about twenty, twenty-one*
S: *if I say for my sake I will first go on travel and then I will get on studying an I think it’s will be hard to do both at the same time*
I: *ahm*
S: *and yeah and get a family too*
I: *mm*
S: *it would be a lot so I think it’s better for me and for the children I that I have a got a job a work*
I: *mm*
S: *something when I get children and so on*
I: *mm right okay thanks very much I think perhaps I forgot to say that you are number 33*
S: *okay*
I: *right thank you*
S: *mm thank you very much*
proficiency 57
progressive 46, 144–147, 175–182, 246–247
rating criteria 107
reliability 61
– inter-rater reliability 63
– intra-rater reliability 63
spoken language 51
standard English 94
stress 166, 221–223, 254
students' attitudes 43, 139–140
third person singular marker 44–46, 143–144, 168–175, 245–246
validity 63
– concept validity 64
– empirical validity 64
– face validity 64
– prognostic validity 64
vocabulary 46–48, 147–148
– modifiers of adjectives 150–151, 194–199, 249–250
word coinage 155
written language 51
ACTA UNIVERSITATIS UPSALIENSIS
Studia Anglistica Upsaliensia
Editores: Rolf Lundén, Merja Kytö & Monica Correa Fryckstedt

112. Anna Uddén: Veils of Irony: The Development of Narrative Technique in Women’s Novels of the 1790s. 2000.