UNIVERSITY OF GOTHENBURG

Teachers' understanding and assessment of oral proficiency

A qualitative analysis of results from interviews with language teachers in Swedish lower secondary schools

Maria Frisch

2015-04-29

Teachers' understanding and assessment of oral proficiency

Teachers' understanding and assessment of oral proficiency

A qualitative analysis of results from interviews with language teachers in Swedish lower secondary schools

Maria Frisch

UNIVERSITY OF GOTHENBURG

DEPARTMENT OF EDUCATION
AND SPECIAL EDUCATION

© MARIA FRISCH, 2014

Licentiate thesis in Subject Matter Education at the Department of Education and Special Education, Faculty of Education, University of Gothenburg.

The licentiate thesis is available for full text download at Gothenburg University Publications Electronic Archive (GUPEA):
http://hdl.handle.net/2077/39226

This licentiate thesis has been carried out within the framework of the Graduate School in Foreign Language Education "De främmande språkens didaktik" (FRAM). The Graduate School, leading to a licentiate degree, is a collaboration between the universities of Gothenburg, Lund, Stockholm and Linnaeus University, and is funded by the Swedish Research Council (project number 729-2011-5277).

Abstract

Title: Teachers' Understanding and Assessment of Oral Proficiency – A Qualitative Analysis of Results from Interviews with Language Teachers in Swedish Lower Secondary Schools
Author: Maria Frisch
Language: English with a Swedish summary
Keywords: Second language didactics, English, oral proficiency, national tests, performance assessment, teacher perception, intended and perceived curriculum

In contemporary discourse on education in Sweden, there is a focus on educational efficiency and student achievement. Aspects of uniformity and equity in grading are often emphasized and the lack thereof is frequently attributed to teachers. English, as one of the most important subjects in school, according to policy documents and also according to teachers and students, has been subject to investigations by the Swedish Schools Inspectorate. English lessons have been observed and found to not always comply with what is stipulated in the curriculum. National tests have been re-rated and found to deviate too much between raters. In an effort to explore one part of the English language proficiency taught in school, this study investigates how twelve skilled language teachers define oral proficiency, how they grade the oral part of the national test for English for year 9 and what influences their grading practice.

The aim is to learn how the teachers perceive oral proficiency and the performance standards for oral proficiency in the policy documents, as well as how they organize and rate the oral subtest based on their interpretations. Furthermore, teachers' perceptions as expressed in interviews and discussions, (the perceived curriculum), is compared with what is written in the national curriculum and the syllabus, (the intended curriculum).

Empirical data to answer the research questions were collected in semi-structured interviews and in group discussions. Before discussing in groups, the informants listened to recorded examples of student interaction and graded the performances. The interviews with the informants, as well as the group discussions, were audio-recorded, transcribed and analyzed.

The analyses of the interviews and the discussions reveal a variety in perceptions of oral proficiency among the teachers. These different orientations to the phenomenon seem to be based on teachers' individual pedagogical philosophies. It is through their orientations that the informants in this study interpret the policy documents, and their orientations thus permeate their teaching as well as their assessment practice.

The analyses of the data further show that the informants are well acquainted with the current policy documents and the national test. They are positive to the test and follow guidelines and instructions for their administration and assessment. They are well aware of the complexity of oral proficiency and the test situation and take measures to ensure that every student has the best possible circumstances to show his/her ability to communicate in English. However, they feel pressed for time when it comes to assessing the tests and wish for more time for discussions on assessment.

The informants themselves also express concerns about certain local factors influencing them in in their assessment and grading of oral proficiency. They point to the group of students they are teaching as well as the community of colleagues at their schools impacting their judgment. The lack of time for preparation, assessment and discussions among colleagues is a factor hindering them in recording, listening a second time and/or asking for a second opinion on all student performances, according to the informants. As they are not unaware of the deviations in rating, the informants state that they wish for more discussions on performance standards and grading, not only at their own schools, but also with teachers from other schools, to enhance fairness and equity in grading.

The subjective judgment of an expert teacher is needed for qualitative assessment of a complex proficiency as oral communicative competence. Professional judgment introduces subjectivity into the rating process, which will be a constant dilemma in a fairness and equity perspective. Therefore extended discussions to minimize variation in grading are needed.

Table of Contents

ACKNOWLEDGEMENTS 11

CHAPTER 1: INTRODUCTION 13

1.1 Aim and Research Questions 17
1.2 Overview 17

CHAPTER 2: BACKGROUND 19

2.1 The Swedish Educational Context 19
2.1.1 School System 20
2.1.2 National Curriculum and National Testing and Assessment System 20
2.1.3 Grading 21
2.1.4 Teacher Education 22
2.2 Foreign Language Teaching 22
2.2.1 Theories on Teaching and Learning Language 23
2.2.2 Communicative Competence 25
2.2.3 The Threshold Level and the CEFR 27
2.2.4 Influences on Swedish Curricula 28
2.3 Oral Proficiency 30
2.4 National Test of English 30
2.4.1 Description 30
2.4.2 Construction and Development 31
2.4.3 Reliability and Validity 32
2.4.4 Oral Proficiency in the National Tests 32
2.5 Assessment and Grading 33

CHAPTER 3: PREVIOUS RESEARCH INTO ASSESSMENT OF ORAL PROFICIENCY 35

3.1. Research on Peer-to-Peer Oral Tests 36
3.1.1 Research on Interlocutor Variables 37
3.1.2 Research on the Co-Construction of Language 38
3.1.3 Research on the Impact of the Examiner/Rater 39
3.2. Research on Rater Cognition in Assessment of Oral Proficiency 39
3.2.1 Interpretation of Grading Guidelines 40
3.2.2 Differences in Rating Approach 41

CHAPTER 4: THEORETICAL FRAME 45
4.1 Teacher Perceptions 46
4.2 Curriculum 47
4.2.1 Intended and Perceived Curriculum 48
4.2.2 Syllabus 49
4.2.3 National Test 49
4.3 Frame Factors 50

CHAPTER 5: METHOD 53
5.1 General Description and Overview of the Study 53
5.1.1 Timeline of the Study 53
5.1.2 Context and Informants 54
5.1.3 Validity and reliability 55
5.2 Interviews 58
5.2.1 Description of Interviews 59
5.2.2 Method of Transcription 59
5.2.3 Ethical Considerations 61
5.3 Group Discussions 62
5.3.1 Description of Group Discussions 62
5.4 Method of Analysis 63
5.4.1 Words and Phrases Used by Informants 63
5.4.2 Interview Analysis 64
5.4.3 Classification of Assessment Criteria 69
5.4.4 Frame Factors 70
5.5 Limitations of Study 71

CHAPTER 6: RESULTS 73
6.1. Teachers' Understanding of Oral Proficiency 73
6.1.1 Defining Oral Proficiency 73
6.1.2 Oral Proficiency in the Classroom 80
6.2 Teachers' Understanding of Policy Documents 86
6.2.1 Informants on Policy Documents 86
6.2.2 Content, Language and Interaction in Policy Documents 89
6.3 The Oral Part of the National Test for English 92
6.4 Assessment of Oral Proficiency 96
6.4.1 Informants on Assessment and Rating of Oral Proficiency 96
6.4.2 Features of Oral Proficiency Noted 99

6.4.3 Informants Grading Oral Proficiency 102
6.4.4 Informants on Rating 105
6.4.5 Informants Reflecting on Assessing Oral Proficiency 108
6.5 Frame Factors 110
6.6 Summary 112

CHAPTER 7: DISCUSSION AND CONCLUSION 115
7.1 Main Findings 115
7.1.1 Informants' Perception of Oral Proficiency 116
7.1.2 Informants' Interpretation of Policy Documents 117
7.1.3 Informants on the National Tests 118
7.1.4 Informants on Grading Oral Proficiency 119
7.1.5 Informants on Framing Factors 120
7.2 Answers to Research Questions 121
7.3 Further Research 124
7.4 Concluding Reflections 125

SVENSK SAMMANFATTNING 127

BIBLIOGRAPHY 133

List of Figures

Figure 1. Swedish School System 20
Figure 2. Curricular Spider Web 50
Figure 3. Timeline of study 53
Figure 4. Informant interviews and group discussions 54
Figure 5. Beliefs on Teaching and Learning Oral Proficiency 68
Figure 6. Beliefs on Oral Proficiency 77
Figure 7. Beliefs on Learning and Teaching Oral Proficiency 82
Figure 8. Top 15 assessment factors classified using national test guidelines 100
Figure 9. Top 15 interactional aspects noted - matrix inspired by May (2011) 101

List of Tables

Table 1. National Tests 21
Table 2. Typology of curriculum representations 48
Table 3. Spradley's Semantic Relationships (1979, p. 111; 1980, p. 93) 65
Table 4. Domain Analysis using *'X is a part of [grading criteria for oral proficiency]'* 66
Table 5. Taxonomy using *'X is a part of [the grading criteria for oral proficiency]'* 66
Table 6. Coding Key for Supplementary Assessment Factors 69
Table 7. Coding Key Interactional Effectiveness inspired by May (2010) 70
Table 8. Defining Oral Proficiency 74
Table 9. Grading Criteria for Oral Proficiency 75
Table 10. Oral Proficiency in *Lgr11* according to informants 89
Table 11. Assessment Factors – Student Example 1 103
Table 12. Assessment Factors – Student Example 2. 104
Table 13. Further policy documents 124

List of Appendices

APPENDIX A: SYLLABUS FOR ENGLISH
APPENDIX B: ASSESSMENT GUIDELINES
APPENDIX C: INSTRUCTIONS FOR THE ORAL TEST AND WARM-UP
APPENDIX D: PART ONE OF THE ORAL TEST
APPENDIX E: PART TWO OF THE ORAL TEST
APPENDIX F: INFORMATION TO PRINCIPALS
APPENDIX G: INFORMATION TO TEACHERS
APPENDIX H: INFORMANTS
APPENDIX I: INTERVIEW GUIDE FOR FIRST INTERVIEW
APPENDIX J: INTERVIEW GUIDE FOR SECOND INTERVIEW
APPENDIX K: CONSENT FORM
APPENDIX L: ASSESSMENT GUIDELINE FROM WWW.SKOLVERKET.SE, ACCESSIBLE DURING SPRING 2013

Acknowledgements

Först och främst vill jag tacka de tolv fantastiska lärare som deltagit i den här studien. Ni gav mig av er dyrbara tid och delgav mig era personliga tankar och har varit en ständig källa till inspiration. Ni har lärt mig mycket om att vara språklärare och kommer att fortsätta påverka mig i min dagliga lärargärning.

Jag vill tacka mina härliga kollegor på Katarinaskolan i Uppsala, som under snart tre år stöttat och ställt upp på alla sätt när jag inte alltid varit helt närvarande i vardagen. Utan er stode jag mig slätt! Tack till rektor Björn Hjalmarsson och studierektor Cecilia Fröjmark för all upptänklig support med det praktiska och alla glada tillrop.

Tack även till mina kollegor på Lärarlyftskurserna på Fortbildningsavdelningen på Uppsala universitet, Maria Allström och Susan Nieland. Med schema-ändringar och glatt humör gjorde ni det möjligt! Tack också Hans Nytell, för att du trodde på mig och läste mina första texter.

Ett särskilt tack till Christina Gustafsson, Uppsala universitet, för läsning av texten och uppmuntrande kommentarer. Tack också Wieland Wermke, Uppsala universitet, för det pågående samtalet om lärares kompetens-utveckling i Sverige och Tyskland.

Forskarskolan FRAM har naturligtvis varit ett stöd i arbetet med uppsatsen och jag vill tacka alla med-licentiander för utvecklande samtal under arbetets gång. Tack särskilt till Maria HR för goda luncher, Karina PG för långa samtal och Céline RH som skapade ett digitalt forum som vi alla kunde umgås i.

Det är många som läst min text i olika skeden; Marianne Molander Beyer till planeringsseminariet, Christina Rosén och Dorte Velling-Pedersen i senare skeden. Tusen tack för värdefulla kommentarer som förde arbetet vidare.

Sist, men inte minst, ett stort tack till min huvudhandledare Britt-Marie Apelgren och min biträdande handledare Gudrun Erickson som fått läsa alldeles för många versioner av min text! Tusen tack för allt stöd och allt tålamod.

Uppsala, april 2015
Maria Frisch

Chapter 1: Introduction

In recent years, the discourse on education, in Sweden as well as in many other countries, has increasingly focused on the assessment of student achievement. In a globalized economy, the dependence on a skilled and highly educated workforce and well informed citizens has triggered intense political debate about the need for an effective school system. In Sweden, a switch to a goal- and results-orientation in school management and the decline in student results in international comparisons, have contributed to intensifying the discussion. The Swedish school system has, over the last two decades, been subject to several reforms which, in turn and combined, have also affected views on how efficiency and goal attainment in education can and should be measured.

To some extent, the role of assessment and grading in educational practices has changed. Today, assessment is often seen as an integral part of the teaching and learning process itself (Black & Wiliam, 1998; Skolverket, 2011b, p. 411)[1], rather than as a separate activity organized at the end of a unit. Formative assessment is used to enhance both learning and teaching, through feedback and so-called feed-forward (Black, Harrison, Lee, & Marshall, 2003) to learners as well as to teachers. Summative assessment at the end of a course or school year normally results in grades, showing what levels of knowledge the learner has reached at that point in time. However, awarded grades are no longer seen only as a record of what proficiency and knowledge the individual student has acquired, but are also used in attempts to achieve quality assurance in schools and to measure teachers' efforts (see e.g. Goddard, Hoy, & Hoy, 2000; Hanushek & Rivkin, 2006; Rothstein, 2008; as well as the database SALSA at Skolverket).[2]

In Sweden, student grades are also used as an instrument for measuring the efficiency of the educational system. Grades are for instance viewed in relation to the financial resources available to schools, municipalities and the nation

1 Skolverket, the National Agency of Education/NAE, will be referred to as Skolverket in the following.

2 To be found at http://siris.skolverket.se/siris/f?p=SIRIS:58:0::NO:::, retrieved 2014-12-17.

(Feldt, 2010; Skolverket, 2005, see also SKL/SALAR).[3] In addition, equal opportunities in education is a significant political goal and, consequently, nation-wide equity and equivalence in grading is of prime importance (Riksrevisionen/the Swedish National Audit Office, 2004; Skolinspektionen, 2013)[4]. The fact that the between-schools variation in grades is increasing (OECD, 2010; Östh, Andersson, & Malmberg, 2013) is worrying and intensifies efforts to secure the same standard of education throughout the country. In other words, student achievement is discussed at multiple levels, both locally and nationally. Questions on how to ensure an equal grading system and fair grades, how to verify that the grades awarded adequately mirror the proficiency and the knowledge of each individual student, how procedures and regulations regarding complaints about grades best can be organized (SOU/Swedish Government Official Reports 2010:96), are being posed. As grades are the primary means to qualify for upper secondary school, as well as higher education, equity and comparability within the school system are critically important to stakeholders as well as to society as a whole.

Efforts to reverse negative trends in education and to strengthen pedagogical development have lately resulted in a new national curriculum, *Lgr11*, with new syllabi for all school subjects, new grades and an increase in the number of national tests. Earlier reforms and curricula were heavily criticized during the first years of the 21st century and political rhetoric came to the fore, resulting in these new policy documents. The changes are aimed at enhancing the focus on subject matter knowledge (Prop./Government Bill 2008/09:87; SFS 2010:800 *Skollag*/The Education Act) as well as promoting fairness and nationwide equivalence in grading.

The importance of English is stressed in the curriculum. In Swedish compulsory school, foreign languages as school subjects have been mandatory since the 1960s. English is the first foreign language taught, starting in school year 1-3,[5] and a second foreign language is introduced in school year 6. Language proficiency is considered vital for communication across borders in a globalized world. Being functionally proficient in languages other than the

3 SKL/Swedish Association of Local Authorities and Regions/SALAR http://skl.se/4.409b7ad7144f9a5c5aeb1df1.html

4 Skolinspektionen, the Swedish Schools Inspectorate/SSI, will be referred to as Skolinspektionen in the following.

5 The Local Education Authority/LEA of every municipality is free to decide on the time for introducing English in compulsory school, year 3 being the last year possible for the introduction of English as a school subject.

mother tongue is valuable for the individual as well as for society. In the national curriculum for the comprehensive school, *Lgr11*, it is thus pointed out that:

> The English language surrounds us in our daily lives and is used in such diverse areas as politics, education and economics. Knowledge of English thus increases the individual's opportunities to participate in different social and cultural contexts, as well as in international studies and working life. (Skolverket, 2011c, p. 32: *official translation*)

This is in line with international trends and EU-recommendations, as proficiency in languages other than the mother tongue is the second of eight key competences for lifelong learning[6] and for 'personal fulfillment and development; active citizenship; social inclusion; and employment' according to the European Commission[7].

When it comes to English language proficiency, Swedish students are generally doing well, according to national standards (Erickson, 2010). They learn English both inside and outside the classroom, since they encounter and use English outside the educational setting in their everyday lives in society at large (Sundqvist, 2009; Sylvén & Sundqvist, 2012). In international comparisons, they outperform students in many countries (Skolverket, 2004, 2012a), contrary to their declining results in tests of other school subjects, such as Mathematics (TIMSS)[8] and Natural Sciences (PISA).[9] Students' attitudes to English are very positive and have been so over many years (Oscarson & Apelgren, 2005; Skolinspektionen, 2011). Spoken language and oral communication are central to communicative competence and students, as well as teachers, rate oral proficiency as the most important skill to gain through their English studies (Erickson, 2010).

Oral proficiency has over the last 40 years gradually been promoted to become part of the dominant competences in language ability described in curricula and syllabi, as *knowledge about* language has given way to *knowledge how* to use language (Apelgren, 2013). This is clear both in the curriculum and in the syllabus for English. The second overall knowledge goal in the Swedish

6 Recommendation 2006/962/EC of the European Parliament and the Council of 18 December 2006 on key competences for lifelong learning, OJ L 394, 30.12.2006.
7 European Commission/EACEA/Eurydice, 2012. Developing Key Competences at School in Europe: Challenges and Opportunities for Policy. Eurydice Report. Luxembourg: Publications Office of the European Union.
8 Trends in International Mathematics and Science Study, http://timssandpirls.bc.edu/#
9 Programme for International Student Assessment, http://www.oecd.org/pisa/

school is to ensure that each student can communicate in English in both spoken and written form (Skolverket, 2011c, p. 15).

The 2011 syllabus, with a specified subject content and clearer goals and objectives than the previous syllabus, frames the planning, teaching and assessing of English. In addition, there are annual national tests for English in school years 6 and 9. To assist teachers in the rating of students' language proficiency, teacher information and guidelines with supplementary assessment factors are provided with the tests. However, oral communicative proficiency is not easy to define and assess, as oral interaction is:

> …dynamic rather than /…/ static /…/ It depends on the negotiation of meaning between two or more persons. /… [It] is context specific. Communication takes place in an infinite variety of situations, and success in a particular role depends on one's understanding of the context and on prior experience of a similar kind. (Savignon, 1983, pp. 8-9).

To rate this dynamic competence requires the professional skill and experience of the individual teacher/rater. Human raters bring subjectivity into the process, which has to be addressed to warrant fair grading of student achievement. The factors that influence raters in their assessment decisions need to be explored, in order to better understand and minimize potential variability (Davison, 2004).

The competence of Swedish teachers assessing and grading student performances in national tests has been studied and questioned by Skolinspektionen. During 2010-2013, the inspectorate organized re-assessment of national tests of English, Swedish and Mathematics to investigate rater agreement. According to Skolinspektionen, the results were not positive, as the inter-rater consistency was considered too low (Skolinspektionen, 2013). However, the re-assessment included only written material and showed the least inter-rater variation, and a higher degree of concurrence, for the re-rating of the English tests.

The reports on the results of the investigation have been heavily publicized and have added to a general distrust in teachers' grading practices. The methods used and the conclusions drawn by Skolinspektionen have, however, also been criticized (J.-E. Gustafsson & Erickson, 2013). In their article Gustafsson and Erickson question the design used in connection with the inferences drawn and demonstrate that there are alternative explanations to the results from the investigation.

The re-rating of national tests can be seen as yet another way of trying to secure equity and comparability in grading of student efforts between schools and throughout the country. It can also be seen as a way of controlling teachers' actions and practices. The general aim of this study is to shed some light on current teacher practices by investigating how twelve teachers of English in year 9 perceive oral proficiency and how they organize and grade the oral part of the national test in English.

1.1 Aim and Research Questions

Communicative competence has been at the center of the English syllabi in the Swedish national curriculum for many years. Oral communicative proficiency is seen as a vital part of language proficiency in all languages. Thus, it is of interest to establish how this productive and interactive proficiency is understood and rated by teachers.

The aim of this study is to investigate how the oral part of the national test in English for year 9 is perceived, assessed and graded by a number of teachers of English.

The following questions will be further explored:

- How do the informants define oral proficiency?
- How do the informants describe the organization of the oral part of the national test of English in year 9 at their respective schools?
- What influences how the informants rate the oral proficiency tests?

In relation to these questions, it is of interest to explore the relationship between teachers' perceptions of the curriculum and syllabus, and what is stipulated in these documents.

1.2 Overview

In Chapter 2 of this text, a background to the study will be provided. Chapter 3 gives a brief overview of some relevant research into peer-to-peer testing of oral proficiency and teachers' grading practice. In the fourth chapter the theoretical frame is outlined. Chapter 5 describes the different methods used in the study, and in Chapter 6 the results are presented. The last chapter discusses the most significant results, and ends by making suggestions for further research, as well as looking at some implications of the findings from the study.

Chapter 2: Background

A focus on accountability and measurement in general, together with a call for efficiency in education, have contributed to a debate on teachers' grading practices and uniformity in grading (see e.g. Skolinspektionen, 2010; SOU 2010:96). In the following, a general background to language teaching and assessment in the Swedish context will be provided.

Firstly, a brief outline of the Swedish school system is presented. Then an overview of some influential theories on foreign language learning and international frameworks for language teaching, as well as of communicative competence, and how they have influenced policy documents, is given. Thirdly, an attempt at defining oral proficiency in the Swedish context is provided. Thereafter the national test for English is described and finally, assessment and grading in a Swedish context is presented.

2.1 The Swedish Educational Context

Swedish children normally start compulsory education in year 1 at the age of 7 and have the right to finish after completion of grade 9. The nine years of mandatory schooling are preceded by an optional pre-school year. After compulsory school, a voluntary three-year upper secondary education, with both vocational programs and programs preparing for higher education, is provided free of charge. Students are admitted to upper secondary school based on their grade point average from compulsory school, where they are assigned final grades by their subject teachers, as there are no external exams in the Swedish school system.

2.1.1 School System

Figure 1. Swedish School System

The nine-year compulsory school in Sweden is a unified and un-streamed education. 98 percent of Swedish students start upper secondary school after their nine years of compulsory school.[10]

2.1.2 National Curriculum and National Testing and Assessment System

A national curriculum and a syllabus for each subject regulate Swedish compulsory education. The syllabi describe aim, core content and knowledge requirements (performance standards)[11] for the different subjects. The syllabus for English stipulates a communicative approach, but no specific teaching methods. In the educational system there is, further, an extensive program of national tests and diagnostic materials supporting and guiding teachers in grading and assessment. Nationwide tests have a long tradition in Sweden. They were initially optional tools developed to support the teachers and have always been well received within the teacher community (Erickson, 1991, 2010; Lundahl, 2006; Marklund, 1987). The current national tests are obligatory and organized annually in compulsory school, as well as in upper secondary school.

Every test comes with teacher instructions, guidelines and benchmarks to serve as support for rating and assessment. National tests are carried out in year 3, year 6 and in year 9 in compulsory school.

10 Retrieved from Skolverket http://www.jmftal.artisan.se/nyheter.aspx 2014-12-08

11 The term 'knowledge requirements' is used in the official translation of the syllabus for English in Lgr11 and will therefore be used in the following.

Table 1. National Tests

School subjects tested per school year		
School year 3	**School year 6**	**School year 9**
Swedish/Swedish as a Second Language	Swedish/Swedish as a Second Language	Swedish/Swedish as a Second Language
Mathematics	Mathematics	Mathematics
	English	English
		History/Geography/Religion or Civics[12]
		Biology/Chemistry or Physics[13]

The aims of the national tests are to support fair and equal grading and to enable an analysis of to what extent the performance standards are reached. They can also be seen as a concretization of goals and criteria.[14]

In addition, the tests can be used for formative purposes. They are obligatory, but at the same time advisory, not decisive, for final grading. For fair and equal grading it is recommended that groups of teachers meet to discuss their ratings of the tests.

2.1.3 Grading

A new grading scale was introduced in 2011. It is goal referenced, as was the previous one, but has six grade levels instead of the previous three: F (fail), E (pass), D, C, B and A.

> According to the 2010 Education Act, Swedish grades are awarded on a scale from A to F. Pass grades are designated A, B, C, D or E, with A as the highest grade and E as the lowest. A fail grade is designated F.
>
> In each course, there are a set of national requirements that need to be satisfied for each grade. There are defined requirements for grades A, C and E.[15]

12 In year 9 one of the subjects History, Geography, Religion or Civics is tested each year. The schools do not know in advance which test they will be required to organize. The subjects are divided up between different regions in Sweden, changing every year. Tests for year 6 are provided according to the same principles, but from 2015 they are optional.

13 In year 9 one of the subjects Biology, Chemistry and Physics is tested each year. The schools do not know in advance which test they will be required to organize. The subjects are divided up between different regions in Sweden, changing every year. Tests for year 6 are provided according to the same principles, but from 2015 they are optional.

14 http://www.skolverket.se/bedomning/nationella-prov, retrieved 2015-01-04, my translation.

Grades D and B are to be awarded for performances that reach a majority of the requirements for C (D) or A (B), but fail to reach all of them.

The national curriculum and the subject syllabi thus specify the goals, objectives, core content and performance standards for each subject and constitute the basis for fair and equal grading.

2.1.4 Teacher Education

Swedish teacher education has been reorganized a number of times during the last three decades. However, for language teachers teaching in year 7-9 and upper secondary school it has, for many years, been mandatory to study the target language at a university language department and to study didactics and curriculum theory at an educational department. The didactics studied have been consistent with language learning theories of the time, as well as with contemporary methods and policy documents, and have thus varied with the period of study. Today (2014), 60 percent of the teachers of English in years 7-9 are trained and certified[16] English teachers.[17]

2.2 Foreign Language Teaching

When discussing English language learning in the following, the term foreign language (instead of second language) will be used, as English is not an official language in Sweden, nor is English commonly used for everyday communication within the country. Furthermore, language learning, instead of language acquisition, will be used to avoid any misunderstandings, as some theories on foreign language learning make a distinction between the two (see Section 2.2.1).

Pedagogical ideas on how best to teach and learn a foreign language have varied over time, and the conflict between formal grammar teaching and functional use of a language has influenced how languages have been taught in educational settings for centuries. Language learning and teaching are,

15 http://www.skolverket.se/om-skolverket/andra-sprak-och-lattlast/in-english/2.7806/swedish-grades-and-how-to-interpret-them-1.208902

16 In 2011, certification of teachers and preschool teachers was introduced in Sweden. The purpose of the reform is to raise the level of skills among teachers and preschool teachers so as to improve the quality of educational services. The Swedish National Agency for Education takes decisions on certification of teachers and preschool teachers. Certification requires a degree in education or in preschool education.

17 Skolverket, http://www.skolverket.se/publikationer?id=3312, retrieved 2014-11-20

however, multi-faceted notions, not just a question of grammar or functional use.

Policy documents regulating the school system frame teachers' actions and thereby also their assessment. Since educational policy documents are influenced by research and pedagogical discourse, the following section aims to introduce some influential theories and how they are reflected in a Swedish curricular context.

2.2.1 Theories on Teaching and Learning Language

It has been argued that ideas on how to teach have a tendency to prevail in schools as institutions long after they have been replaced by new theories in the wider educational world. Teachers' own experiences as students, together with their training and the community of teachers they become a part of at their work place seem to contribute to preserving 'traditions' (Cuban, 1990). New pedagogical ideas are therefore not always readily established, even if they are well known and prescribed in policy documents.

Contrary to language pedagogy before the Second World War, language teaching in school settings in the Western World in the 1950s and 1960s often had a focus on communicative functional language. The gradual shift from formal language skills to functional language use was due to new demands for language proficiency in society (Richards, 2001; Tornberg, 2005). The audio-lingual method, partly based on ideas from behaviorism (Skinner, 1957), was developed to meet these demands. Learning by imitation and the formation of habits were in focus.

Chomsky's theory on language learning was introduced in reaction to Skinner's behavioristic ideas and criticized the thought that only what the learner is exposed to will be learnt. Human beings have an innate language ability, a 'universal grammar', he argued (Chomsky, 1965). He saw 'competence' and 'performance' as two separate entities, where language/linguistic competence (a native speaker's knowledge of the language system) allows him/her to produce language, referred to as 'performance'. 'Competence' is then perceived as the ideal inner language system, connected to universal grammar, whereas the 'performance' is the language system used in actual communication, disrupted by outer distractions as well as inner disturbances, such as memory limitations. (Chomsky himself made no claims about any implications of his theory for foreign language teaching.)

Krashen's (1981) 'monitor model', inspired by Chomsky's idea of a universal grammar, has five hypotheses. The first introduces the contrast between acquisition and learning; the *acquisition* of a second language happens unconsciously, as the learners are exposed to understandable fragments of the new language. The learners *learn* when they, consciously, pay attention to form and rules. There is no connection between the two, according to Krashen. In other words, *learning* about language structure and grammar does not directly affect the actual output of language.

The second hypothesis of Krashen's model is the 'monitor hypothesis', stating that the *acquired* system enables the learner's spontaneous use of language, whereas the *learned* system is used to monitor what is produced. The *learned* system, however, needs time and enough knowledge to function properly. In the 'natural order hypothesis', the order of features *acquired* is defined. In the 'input hypothesis' Krashen states that language is *acquired* when the acquirer is exposed to comprehensible input, the $i + 1$ (i representing current language level and $+ 1$ indicating input just one step above the current level).

Targeted instruction is then, according to Krashen, not very useful and will not impact acquisition, as implicit and explicit knowledge are seen as separate systems with no transfer between them. Motivation, needs, attitudes or feelings are an affective filter, hindering or supporting learning/acquisition in the 'affective filter hypothesis' of the model (Krashen, 1982).

The 'processability theory' presented by Pienemann (1995) also focuses on what is learnable and teachable. Pienemann argues that it is of no avail teaching too far above (or below) the learner's current level of knowledge. "It is important to know what is learnable at what point in time." (p. 4). Thus, Krashen and Pienemann agree that learners cannot process and appropriately use linguistic information that is not on the right developmental level.

Krashen's 'monitor model' has been criticized in several ways, e.g. for not properly describing the input hypothesis (White, 1987) and for campaigning against formal language training and not empirically validating the claims of the 'comprehensible input hypothesis' (Ellis, 1985). However, the model and its ideas were very influential at a time when views on foreign language teaching and learning were changing from emphasizing drills and imitation to stressing meaning and communication, and are still referred to today. Pienemann's 'processability theory' continues to be a topic in pedagogical discourse and research (Bardel & Falk, 2007; Ellis, 2002).

Other theories inspiring foreign language teaching are, for instance, the 'interaction hypothesis' (Long, 1985) and the 'comprehensible output hypothesis' (Swain & Lapkin, 1995). Long argues, along with Krashen, that comprehensible input is essential for language learning. The negotiation of meaning in interaction is emphasized and seen as promoting linguistic development.

In the 'comprehensible output hypothesis' Swain (2000) argues that output pushes learners to process language more deeply.

> …it is dialogue that constructs linguistic knowledge. /…/ It is where language use and language learning can co-occur. It is language use mediating language learning. It is cognitive activity and it is social activity. /…/ …this external speech facilitates the appropriation of both strategic processes and linguistic knowledge." (p. 97)

There is a strong focus on oral interaction in both these hypotheses, stressing a functional view of language teaching.

Contemporary foreign language teaching has been influenced by the language theories of the last decades as well as by the more general socio-cultural theory of knowledge presented by Vygotsky (1978). Vygotsky's view on human activity as mediated by semiotic means like language, became influential in the 1980s. It was, however, originally conceived in the 1920s and reflects the interest in the social context of speech during that time (cf. Saussure, 1916/1970). Semiotic tools, of which language is one, facilitate the co-construction of knowledge in social interaction, according to Vygotsky. This co-constructed knowledge is, in time, internalized by the individual. Vygotsky thus claims that internal mental activity has its origins in external communicative activity and views speaking and thinking as tightly knit processes. Learning cannot be separated from language, social context and social interaction (Vygotskij & Kozulin, 1986). The impact of socio-cultural theory on curriculum, instruction and assessment is visible in the communicative and interactional aims for foreign language teaching, with a focus on meaningful interaction as the basis for language learning.

2.2.2 Communicative Competence

The concept of communicative competence has been central to foreign language teaching, learning and research for decades. Some of the various

definitions of communicative competence, as they have been put forward in different models or theoretical paradigms, will be presented below.

Communicative competence, as a term, was defined by Hymes (1972) in reaction to what he found to be inadequate explanations of 'competence' and 'performance' (Chomsky, 1965). Hymes took a sociolinguistic perspective and stated that knowing whether something is *possible*, *feasible* and *appropriate* to say in a certain situation, and whether something actually is said in a particular *context* is essential in defining language practices (Hymes, 1972, p. 281). As the language that is perceived as appropriate varies across speech communities, the *social rules for language use* "without which the rules of grammar would be useless" (p. 278) have to be taken into account in the learning, teaching and assessment of language proficiency. Hymes' ideas were part of new radical thoughts on a more democratic society and a communicative turn in linguistics which was gradually introduced during the 1960s and 1970s (Kramsch, 1986).

The concept of communicative competence was further developed by Canale and Swain (1980) as they divided the notion into:

- *linguistic competence* (grammar rules, spelling, pronunciation, etc.)
- *sociolinguistic competence* (social rules, differences in language use)
- *discourse competence* (being able to combine meaning and grammatical form to produce various kinds of comprehensible oral and written texts)
- *strategic competence* (ability to use different strategies to support communication).

A definition of communicative language ability (CLA) was presented by Bachman (1990), who thereby renamed communicative competence. In his definition the processes of interaction between the components of communicative ability and the context are added. The components of CLA were defined as:

- *language competence*,
- *strategic competence* and
- *psychophysical mechanisms*.

Language competence comprises "… specific knowledge components that are utilized in communication via language" (Bachman, 1990, p. 84). Strategic competence is "… the mental capacity for implementing the components of language competence in contextualized communicative language use" (Ibid.). Psychophysical mechanisms are neurological and psychological procedures in the course of using language, like sound and articulation (Ibid.). The

importance of context in language use is highlighted, as well as the dynamic (non-static) interaction between the context and discourse (Bachman, 1990). Bachman and Palmer (2010) extended this further as they introduced

- *language use*

as an aspect of communicative competence. Language use is described as:

> …the creation or interpretation of intended meanings in discourse by an individual, or as the dynamic and interactive negotiation of intended meanings between two or more individuals in a particular situation (Bachman & Palmer, 2010, p. 34).

They also point to a number of factors such as personal attributes, topical knowledge, affective schemata and cognitive strategies that may have a major influence on language use and language performance. These factors are of importance when the ability to communicate in a foreign language is to be assessed, an aspect which will be further developed in Section 2.5.

Contemporary discourse on communicative competence sometimes touches on the need for yet another expansion of the definition. Due to changes in a society that is more multi-lingual and multi-cultural than previously, due to the varieties of English used globally, and due to new ways of interacting and communicating in various media, it has been suggested that the term communicative competence, as it stands today, does not adequately describe the true nature of contemporary co-constructed communication in social interaction (Kramsch, 2006; Leung, 2005). Thus, new attempts at defining communicative competence are to be expected.

2.2.3 The Threshold Level and the CEFR

In the mid-1970s, a joint effort within the European Union resulted in the Threshold Level for languages (Ek, 1975). It described a basic level for learners of a foreign language in terms of notions and functions, based on the abilities required for using language in communication. This was a new way of defining the goals for language learning. The Threshold Level was continually improved and extended and ultimately led to the development of a common framework for language levels, the CEFR, the Common European Framework of Reference for Languages (Council of Europe, 2001; North, 1995). The CEFR describes six separate proficiency levels on a scale. The levels define competences and sub-competences, using descriptors for each level in a progression from A1 to C2. The CEFR has become very influential,

leaving its mark in many language syllabi around the world (*The Common European Framework of Reference: The globalisation of language education policy*, 2012; North, 2014).

Research into foreign language learning and the development of different pedagogical methods as well as international framework, have affected policy documents for the teaching of new languages. In curricula and syllabi as well as in classrooms, the different methods and findings from research bring about change and contribute to develop activities. The mediating agents are the teachers in the classrooms, with their individual perceptions and understanding of language teaching and language learning.

2.2.4 Influences on Swedish Curricula

Based on the presentation in the previous section, the aim below is to explore how these theories and ideas have impacted steering documents for English in the Swedish school and, consequently, teacher perception and teacher action.

In the 1960s the influence of the audio-lingual method led to curricula prescribing the use of the target language for instruction, as well as the practice of micro-dialogues and repetition. In the syllabus of *Lgr62,* it was pointed out that learning grammar was not a goal as such, but a means to better understand and use the language (Skolöverstyrelsen, 1962). In *Lgr69,* the use of the target language, combined with drills and replication, was heavily emphasized (Skolöverstyrelsen, 1969a, 1969b). The actual communicative use of the language was something for the future (Ferm & Malmberg, 2001; Skolöverstyrelsen, 1990; Tornberg, 2005).

The theories of Chomsky (1965), Hymes (1962) and Krashen (1981), as well as the 'threshold level' (Ek, 1975), and the reasoning behind it, influenced the English syllabus in the curriculum of 1980, *Lgr80* (Skolöverstyrelsen, 1990). There was a move towards a communicative approach, as well as a curricular shift towards a focus on the learner, stressing psychological, emotional and social aspects. As a result, so called affective goals were introduced in the syllabus for English:

> The instruction is further to lead to pupils wanting to and feeling confident enough to use English… (Skolöverstyrelsen, 1980, p. 77, *my translation*)

The orientation was holistic, including emotional, social and psychological as well as a cross-curricular perspectives. The target language was to be used in contexts meaningful to the learners. There were no recommendations or

prescribed methods for teaching; the syllabus only stipulated goals to be reached and the primary goal was to learn oral skills (Skolöverstyrelsen, 1980, p. 77). Objectives described were, for instance, "to enhance students' ability to obtain and give information, as well as to express wishes, feelings and opinions." (Ibid., p. 79, *my translation*), and the language used in oral exercises was to be "natural and realistic" (Ibid., p.77, *my translation*), reflecting a socio-linguistic perspective on language use and communicative competence. The change in the curriculum from *Lgr69* to *Lgr80* is described as a shift from a focus on teaching to a focus on learning, from teaching of certain prescribed elements to learning of certain functions and notions (Skolöverstyrelsen, 1990).

From 1994, communicative competence is at the heart of the syllabus for English. The ability to use a language to interact and communicate was described as consisting of receptive, productive and interactive skills (cf. CEFR). The use of communicative strategies was emphasized. Metacognition, learning how to learn by reflecting on one's own learning, was stressed. In the commentary to the language syllabi the ideas of Hymes (1972), Krashen (1985, 1993), Pienemann (1984) and Vygotsky (1995), among others, are referred to (Skolverket, 2001). There are further signs of the sub-competences of communicative competence of Canale and Swain (1980) in the text (cf. Apelgren, 2013). This curriculum also introduced a goal-and-criterion referenced orientation, as well as a new criterion referenced grading system in Sweden. Specific methods for teaching were, however, not included. The syllabi in *Lpo94* were revised in year 2000.

In the current syllabus for English, the importance of language as a tool for learning is emphasized: "Language is the primary tool human beings use for thinking, communicating and learning" (Skolverket, 2011c, p. 32), making a socio-cultural approach explicit. Communicative interactional proficiency dominates the syllabus and the harmonization with the CEFR (Council of Europe, 2001) is apparent. Since year 2000, the syllabi for languages for compulsory and upper secondary school include seven steps, which all correspond to CEFR-levels, e.g. the pass level (grade E) in English in school year 6 equals A2.1 in the CEFR and the pass level in school year 9 equals B1.1 (Skolverket, 2011a, p. 7). The general impact of CEFR-descriptors for the various competences and sub-competences is explained in the commentary to the syllabus for English (Skolverket, 2011a).

The different theories and frameworks presented in the beginning of this section can, accordingly, be said to be reflected in the Swedish curricula and syllabi over time.

2.3 Oral Proficiency

Numerous definitions of speaking proficiency have been suggested, but they differ among researchers, making it difficult to arrive at a definite specification (Iwashita, 2010; McNamara, 1996). There are multiple characteristics of the skill to take into consideration, such as grammar, fluency, pronunciation, vocabulary and comprehensibility, as well as interaction with a partner. These are normally aspects in a final global rating of the proficiency. In curricula and syllabi, as well as in grading criteria, attempts at defining both the proficiency and its different components are made.

In the Swedish syllabus for English, speaking and writing are mostly mentioned simultaneously, e.g. it is stated that the learners are to be given the opportunity to develop their ability to "express themselves and communicate in speech and writing" (Skolverket, 2011c, p. 32, *official translation*). However, in the core content more specific traits of oral proficiency are presented (see Appendix A). In the oral subtest of the national test and in the guidelines for the test, as well as in the supplementary assessment factors, a concretization of the definitions of speaking is presented, (see Appendix B).

2.4 National Test of English

The national test of English is part of the national testing and assessment system. It is how teachers perceive and rate the oral part of this test that is under scrutiny in this study. A description of the test is given below.

2.4.1 Description

Language tests have a long tradition in the Swedish school system. The first nationwide tests of foreign languages for upper secondary schools were held in 1864 (T. Lindblad, 1991). At the end of the 1950s, English was included in a package of annual standardized national tests (Erickson & Börjesson, 2001). In 1987 the tests became mandatory and in 1998 the oral subtest was added (Erickson, 2009).

The national test of English for year 9 is a multi-dimensional proficiency test, divided into several parts and covering a broad representation of the construct of language proficiency. Productive and receptive as well as interactive language skills are tested. Included in the testing materials distributed on paper to all schools are instructions for the organization and administration of the tests, as well as supplementary assessment factors and commented examples of authentic student production/interaction to serve as benchmarks and guidelines for assessment. These are there also

> …to serve as in-service training within the field of language assessment in a wide sense. Hence, the materials have a measurement purpose, as well as a pedagogical purpose. (Erickson, 2010, p. 2).

The results of the tests are collected by Skolverket each year and statistics are made available to teachers, schools and the general public.

In general, test results have been excellent; around 95 percent of the students reach the pass level according to national standards (Erickson, 2010; Velling Pedersen, 2013). Stakeholders' reactions are mainly positive; both teachers and students approve of the tests. During the last ten years, around 95 percent of the teachers have been very positive in their evaluations. Negative feedback gathered in the annual questionnaires typically comment on the work load or lack of time for marking, not on the tests themselves or what they are testing (Erickson, 2010).

2.4.2 Construction and Development

The national test for English is developed in an on-going collaborative process involving teachers, teacher educators, test developers and researchers from different disciplines, as well as students. Current research, together with national and international experiences from the field of language testing, is taken into consideration (Erickson, 2010, 2012). After its construction, an initial succession of pilot tests and then a major pre-test is carried out in a number of randomly chosen classes and schools all over Sweden. The ambition is to let 400-500 students try out the tasks before the compilation and distribution of the final version (Erickson, 2012; Skolverket, 2011a).

2.4.3 Reliability and Validity

National tests need to be reliable and valid tools for measuring student achievement, and test scores need to be fair and dependable as well as useful for their intended purpose (Luoma, 2004). Reliability, i.e. the consistency of test scores over time and internal consistency (that individual raters are consistent in their own rating), is strengthened by well-defined criteria and bench marks. Validity, i.e. the meaningfulness of test scores, has been described as

> …an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* and *actions* based on test scores… (Messick, 1993, p. 13, italics in original)

Reliability and validity have also been discussed in terms of trustworthiness, authenticity and 'thick descriptions' of context and participants (Guba & Lincoln, 1989); this tallies with the way opinions of the hundreds of test-users influence the decisions made in modifying the final tests for English.

The national tests for English are repeatedly evaluated (see e.g. Naeslund, 2004; Velling Pedersen, 2013), showing that teachers find the tests well aligned with the curriculum and that they appreciate guidelines and benchmarks. Furthermore, the test results, year by year, consistently demonstrate a high degree of correspondence with final grades,[18] indicating agreement between performance standards, teacher perceptions, test specifications and student performance.

Studies on inter-rater agreement and consistency in the English tests have reported high degrees of concurrence (Erickson, 2012; Velling Pedersen, 2013), contrary to reports from Skolinspektionen (Skolinspektionen, 2013). In 2009 the inter-rater agreement for the oral interaction and production part of the test, which is the focus of this study, had been over .90 for three consecutive years, based on data from the internal development process (Erickson, 2009).

2.4.4 Oral Proficiency in the National Tests

The oral subtest is a peer-to-peer test where the students are divided into pairs or groups and instructed to interact and keep the conversation going according to the directions in the test. The test is divided into three parts and

[18] http://siris.skolverket.se/siris/f?p=SIRIS:1:0::NO::: retrieved 2015-02-08

lasts for 15-25 minutes (see Appendices C, D and E). The teacher is present during the conversation and acts as a coach and an instructor at the start of the interaction. After that s/he is a quiet observer and examiner. The test is assessed holistically, with the support of a number of analytical aspects defined in the supplementary assessment factors in the guidelines, (see Appendix B), as well as commented benchmarks. The guidelines strongly recommend that the interaction be audio-recorded to enable a second listening as well as a discussion on the test performances among colleagues.

2.5 Assessment and Grading

Assessment is an integral part of the didactic process, carried out for different purposes and in different formats on a daily basis. However, it has been suggested that the teacher profession, contrary to reform intentions, has lost some of its former assessment competence over recent years (Lundahl, 2009, 2011). In a study, 50 percent of Swedish teacher students claimed to have had no instruction at all on assessment (Lundgren & Nihlfors, 2005). According to a recent OECD-report, Swedish teachers take part in professional development to a lesser extent than teachers in other countries and report that they feel a need for more training, especially when it comes to assessment and grading. They also report that they need more information on the new curriculum (Skolverket, 2013). In spite of extensive information and training during the implementation of *Lgr11* and the new grades, teachers in the above report expressed a need for further professional development on these topics. This suggests uncertainty and could be detrimental to the full implementation of the new curriculum and to fairness of grading. Policy documents with high expectations, clear objectives and assessment of results are often appreciated by teachers, if they are not too rigid and do not restrain the pedagogical creativity of the teachers (OECD, 2009). To award reliable and valid grades to student performances, teachers need clear performance standards but also room for professional judgment based on pedagogical experience and didactical knowledge.

Chapter 3: Previous Research into Assessment of Oral Proficiency

The focus of this study is teacher perceptions of oral proficiency and the rating of the oral part of the national test for English. In the following, relevant previous research on assessment of oral proficiency, on peer-to-peer orals and on rater cognition will be presented. As there are few studies involving younger Swedish or Scandinavian students and the rating of their proficiency to be found, most studies mentioned here are from a non-Swedish context.

To speak and interact in a foreign language requires planning, thinking, retrieving from memory, adjusting to the situation and speaking partner/-s, as well as orally formulating and verbally uttering what you intend to say, all at the same time. It is, in other words, a demanding and complicated task: "it is not static and dense but mobile and intricate", like dancing (Halliday & Matthiessen, 2004, p. xxiii). The demands on the speaker of a foreign language are extensive:

> Learners must simultaneously attend to content, morphosyntax and lexis, discourse and information structuring, and the sound system and prosody, as well as appropriate register and pragmalinguistic features. (Hinkel, 2006, p. 114)

It has been suggested that attending to one aspect of oral interaction (e.g. complexity of language, fluency or accuracy) may hinder the ability to attend to other aspects (Krashen, 1992; Skehan, 1998). As speaking, especially in interaction with other speakers, offers limited time for planning, the vocabulary used tends to be vaguer and more generic than in writing. Also the grammar of spoken language is simpler, which has to be taken into account when assessing (Luoma, 2004).

Oral proficiency in interaction is a collaborative action. To reach a level of effective interaction, interlocutors need to share knowledge of the world, references to some external context and the co-construction of an internal context (Kramsch, 1986). A large part of oral communicative proficiency is in the joint effort, the turn-taking and adjustments made to reach the speaking partner.

> "… what people say and understand in real communications with other people is co-constructed by virtue of the interactive nature of such communications…" (de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012, p. 10).

Speaking partners thus influence one another and the co-constructed performance.

The type of interchange also affects the verbal outcome. An interaction based on familiar topics usually results in more accuracy and fluency but simpler language, whereas an exchange requiring that the interlocutors explain and justify a standpoint often results in the use of more complex language and possibly in less correctness (Skehan & Foster, 1997).

The "observation of free oral communication" was found to be one of the three most commonly used forms of assessment of language proficiency among Swedish language teachers in a study reported in 2011 (Oscarson & Apelgren, 2011, p. 7). This informal and often formative way of assessing speaking is thus probably the most common way of assessing oral proficiency.

The formal assessment of speaking proficiency in an international context is often carried out through performance assessment where the learners demonstrate their proficiency by interacting with a partner (face-to-face, on the telephone or digitally), in the format of an oral proficiency interview (OPI) or in a peer-to-peer conversation. In both cases, the aim is to elicit extended talk from the test taker. In the OPI, a trained examiner is conducting an interview of the candidate. In the peer-to-peer test, two or more candidates interact with one another. The talk elicited varies with the different formats. How much of real life reflection, of target language use and of domain reflection (Bachman & Palmer, 1996) is needed to establish the level of oral proficiency has to be decided in each case of performance testing.

3.1. Research on Peer-to-Peer Oral Tests

The assessment of oral proficiency is by necessity a complex task (de Jong et al., 2012; Kramsch, 2006). When the use of group interaction as a means of testing was introduced in the 1970s, it was met with suspicion because of the intricacy of the test format (Fulcher, 1996). Today, peer-to-peer oral interaction in paired or group tests is used alongside the OPI, which was formerly the standard procedure. In paired or group interactions between peers, the test-takers move between listening, speaking and co-constructing

the dialogue, using a broad variety of communicative skills (Brooks, 2009; May, 2011). It has been suggested that candidates demonstrate more facets of interactional proficiency in peer-to-peer tests than in the OPIs, where the built-in imbalance between interviewer and interviewee may restrain the elicitation of a candidate's full range of skills (Brooks, 2009).

There are numerous issues to consider when the level of a learner's individual oral proficiency is to be established in the joint construction of discourse in a group or paired test. Defining the construct of effective interaction lies at the heart of the matter, as well as using appropriate measures to capture this construct. The peer-to-peer test format is seen as similar to a realistic communicative situation and has been shown to provide test takers with the opportunity to demonstrate a large part of their real-life-abilities (Gan, Davison, & Hamp-Lyons, 2009). However, attempting to make the test situation as authentic as possible comes with the risk of considerable variation, which may jeopardize fairness and validity and make it difficult to balance authenticity and validity when comparing test results (Brown, 2003). What is gained with a format of perceived "true" communication may be threatened by difficulties in ensuring equal opportunity and fair judgment for the test takers.

3.1.1 Research on Interlocutor Variables

Paired or group oral interaction includes the interlocutor/speaking partner (-s) as a variable that may affect not only the joint construction of a conversation but also the individual test-taker's performance. The potential interlocutor effect on performances has been demonstrated in several studies (Brown, 2003; Chalhoub-Deville, 2003; Davis, 2009). Personal characteristics such as extraversion and assertiveness have been shown to affect the performance and the scores of the individual test-taker as well as the joint construction of the interaction (Nakatsuhara, 2011; Ockey, 2009). Furthermore, the candidates' level of acquaintanceship with one another has been found to influence the outcome (Ikeda, 1998).

The general level of language proficiency of the speaking partners has also been reported to have an impact (Gan, 2010), but not always on the scores awarded (Davis, 2009; Iwashita, 1999). Davis further discovered that test-takers at a lower proficiency level tended to produce more words in interactions with speaking partners with a higher level of proficiency, than in

interactions with other lower proficiency students, while high-scoring test-takers produced a large number of words in both constellations. However, in pairs of a high proficiency student and a student of lower proficiency, the weaker student tended to be more passive during the conversation (Davis, 2009, pp. 386-388). It has also been observed that raters find it especially difficult to award scores for individual performances in asymmetric pairs or groups, where they feel a test taker might be disadvantaged because of a mismatch (May, 2009). The influence of the interlocutor variables on the co-constructed interaction and on the rating of the performances is thus not straightforward and research shows mixed results. (It should be remembered that the OPI situation also involves an interlocutor, i.e. the interviewer, who has, as shown in other studies, (e.g. Brown, 2003; Nakatsuhara, 2008), an impact on the interaction.)

3.1.2 Research on the Co-Construction of Language

The co-construction of discourse in the candidate-to-candidate interaction complicates the assessment of individual performances. Whose proficiency is being assessed (Brooks, 2009; Ducasse, 2009; May, 2009)? The definition of individual proficiency, as opposed to co-constructed interaction has been identified as somewhat problematic (Gan, 2010; May, 2009, 2011). A shared grade for the actual interaction has been suggested (May, 2009).

Oral proficiency is not only the production of speech, but also includes interactional skills, the ability to listen to one another, to encourage and support one another, as well as to include everyone in the conversation (Galaczi, 2008); all these skills are therefore normally incorporated in the construct, and tested. In her research, Galaczi found four interactional patterns for co-construction of discourse in peer-to-peer speaking tests: *collaborative*, *parallel*, *asymmetric* and *blended* interaction. In a *collaborative* interaction both speaking partners introduce and develop their own and one another's topics. A *parallel* interaction is characterized by two solos; both speakers are focused on their own production and not really listening or reacting to the partner. In an *asymmetric* interaction one of the speakers is taking all the initiatives and doing most of the talking. The *blended* interaction is described as an interaction where the speakers use two or more of the above patterns in their conversation (Galaczi, 2008).

3.1.3 Research on the Impact of the Examiner/Rater

In the peer-to-peer tests the examiner/rater is often also the instructor. The role of the examiner is then radically different from that in the OPI, where the interviewer leads and designs the dialogue, and where variation in interviewer behavior has been shown to influence the interaction and the opportunity for the candidate to show his/her oral skills (Brown, 2003; Nakatsuhara, 2008). In the case of the peer-to-peer interaction, the instructor/examiner is not to intervene in the conversation. S/he is to first give instructions to the candidates and then to listen and rate the efforts of the participants according to specified grading criteria. However, there is, of course, the possibility that the instructor/examiner impacts the test situation and the test takers in this test format as well, whether intending to or not, as discussed, with explicit reference to the Swedish national test of English, by Sandlund and Sundqvist (2011).

3.2. Research on Rater Cognition in Assessment of Oral Proficiency

The complexity of assessing peer-to-peer interaction leads to multiple sources for possible variation in rater behavior, leading to potential differences in how students' performances are interpreted by individual raters (Brown, 2000). Rater effects are eternal and universal and may take many different forms and can be hidden in most parts of an assessment practice. Variation in test scores associated with rater factors is extensively reported in research. The differences in rater behavior are often attributed to a general harshness or leniency in judgment by individual raters. Teachers' assessment decisions have been found to be based on their beliefs and teaching experience, not only on prescribed criteria and benchmarks. In a review of 25 years of research into language teacher thinking and practice, Borg summarizes:

> Teachers are active, thinking decision-makers who make instructional choices by drawing on complex, practically-oriented, personalised, and context-sensitive networks of knowledge, thoughts, and beliefs. (Borg, 2003, p. 81)

Assessment practices as part of instruction have thus been shown to be individually as well as contextually and culturally embedded. For instance, the general level of language proficiency, as well as the task and the prompts for

the oral interaction, had an impact on whether the content of the interaction was seen as just a vehicle to show linguistic skills or a grading criteria per se in a study by Brown, Iwashita and McNamara (2005). Other studies have shown that when student performances are weak, raters have a tendency to rely on linguistic features instead of content (McNamara, 1996), and vice versa; there seems to be a stronger focus on content than on accuracy when the language level is higher (Ang-Aw & Goh, 2011).

3.2.1 Interpretation of Grading Guidelines

Deviations from prescribed grading rubrics are not uncommon, even though several studies show a strong influence of government policies on assessment of student performances in national tests (Rea-Dickins in Hedge, 2001; Leung & Teasdale, 1997). The problems of fairness in judging student performances according to criteria defined in a grading rubric may be considerable. The criteria may be interpreted and weighted differently by raters in a holistic judgment, or ordered differently in a hierarchy. For example, self-correction in a paired oral has been shown to be regarded as positive by some raters, adding to and clarifying the interaction, whereas others perceive this to be negative and disturbing the interaction (Brown et al., 2005).

Further, raters may also understand and interpret the guidelines and scoring criteria differently. They might agree on the quality of the performance, but vary in how they interpret the rating scale and disagree on which grade to award the performance (Brown, 2000; Orr, 2002).

Features raters find salient in a student's oral performance might not be available in the rating criteria for the test in question. In several studies researchers have found criteria in addition to those prescribed are being included in the assessment of paired orals. For instance, aspects like personality, body language, culture and demonstrated assertiveness may affect the assessment (Ducasse, 2009; May, 2009; Sandlund & Sundqvist, 2011). Ducasse (2009) found that raters added listening, supportive listening and listening for comprehension, as criteria for successful interaction. Another category of criteria, not included in the official rating rubric for the study but used by raters, was interactional management: how turn-taking and topic cohesion were managed or mismanaged in the conversations. Aspects of "communicative skills" were also added as criteria by raters in a study reported by Brown (2000). Eye contact, gestures and listening were

considered supportive of effective interaction by raters in a study conducted by May (2011). In other words, raters sometimes pay attention to features not included in the guidelines for assessment, which they, however, subsume under a given category or simply add to the other criteria because they prioritize them (Brown, 2000; Douglas, 1994; Orr, 2002). This is in line with other research stating that teachers often use "implicit constructs" (Rea-Dickins, 2004; Teasdale & Leung, 2000) that are their own internal quality standards when assessing student performance. The risk of these individual constructs not agreeing with the prescribed grading schemes and thus creating a gap between what is actually assessed and what is to be assessed has been found to be substantial (Leung & Teasdale, 1997). Teachers may also include construct-irrelevant factors outside of criteria in their ratings, such as students' behavior, effort or improvement (Brookhart, 1991, 1993; Oscarson & Apelgren, 2011).

How deeply raters motivate their scoring decisions differs between individuals. Several of the raters in a study by Joe, Harmes and Hickerson (2011) never consulted the full rubric for scoring when making their decisions. Especially among experienced raters, a personal framework for assessing, which can present a mismatch with guidelines and therefore introduce a threat to consistency in rating, was obviously used. This seems to suggest that experience does not automatically add to the consistency of judgment of levels of oral proficiency. According to Joe, Harmes and Hickerson (2011), the experienced raters in their study relied on holistic, intuitive evaluations to a greater extent than did inexperienced raters, who tended to follow the guidelines more closely (see also Orr, 2002).

Independently of how the guidelines are interpreted, there are studies indicating that raters pay attention to a limited number of features during an assessment session (Joe, Harmes, & Hickerson, 2011). When guidelines include too many criteria for a rater to pay attention to during the rating process, they may compromise the fairness and accuracy of grading.

3.2.2 Differences in Rating Approach

Raters have different approaches depending on their experiences and on their general pedagogical philosophy. Some teachers have a general holistic approach when rating and form an initial impression, and then go back checking various aspects of the performance. Other teachers have an

'objective' approach and look for reactions to each prompt of the test separately and then arrive at a final score (Ang-Aw & Goh, 2011; Pollitt & Murray, 1996). A mixed approach is another possibility, in which both prompt reactions and specific aspects are considered simultaneously (Brown, 2000). Raters are, as mentioned above, also often divided into individual rater profiles, e.g. harsh/lenient, which are referred to their experiences, perceptions and beliefs (Brown, 1995; 2000; McNamara, 1996).

Comparing Performances

Comparing students' performances, instead of measuring student performances in relation to the appropriate rating scale threatens to turn the test into a norm-referenced, instead of a criterion-referenced, test. According to some researchers, comparison between performances occurs when the descriptors are not clear enough for the raters to use them properly (Orr, 2002). Others claim this happens because comparing is cognitively less demanding (Bejar, 2012). This can lead to varying or unclear definitions of the construct for different levels of performance, and to raters expressing that they "feel" which level is right (Ibid.).

Test Organization

The practical organization and administration of the test itself may impact the rating. Rating many tests in succession may lead to a comparison between student performances instead of measuring achievements according to a scoring rubric. Also, fatigue after long sessions of assessing many performances, may have an impact on the grades awarded and the consistency of scoring (Harik et al., 2009; Puhan, 2008).

Bejar further claims that there are anecdotal reports that assessing many performances consecutively can lead to raters avoiding the highest scores as well as the lowest, feeling it is inappropriate to award too many extreme scores (2012).

Influence from a Community of Practice

Assessment is context-embedded in various ways and teachers' decisions have been shown to be highly influenced by the assessment culture of the school or a local community of colleagues (Davison, 2004; Hall & Harding, 2002). In environments where group discussions on rating of student achievement and an on-going exchange of ideas and thoughts on formative and summative

assessment are made possible and encouraged, and where a common language for pedagogical judgment and assessments makes the implicit and tacit individual criteria more transparent and explicit, teacher assessment is positively influenced, according to research reported (Davison, 2004; Hall & Harding, 2002; Wiliam, 2007). Another recent study showed that Swedish and German teachers greatly appreciate assessment discussions with colleagues (Forsberg & Wermke, 2012) and find them to be excellent examples of informal professional development. It has been reported that inter-rater-reliability is raised through discussions on students' results (Erickson, 2012), which would then strengthen fairness in grading. However, strong professional communities need to be learning communities open to challenge and critical discussion, as well as being supportive, in order to promote professional development (e.g. Borko, 2004). Teacher collaboration can also be conformist and represent group think (Hargreaves, 1994), constraints and the preservation of particular local traditions and routines (Munthe, 2003).

In the following, the perceptions and assessment practice of a number of skilled English teachers concerning oral proficiency have been investigated. Teacher statements will be analyzed to find out how teachers perceive speaking proficiency, how they organize the oral tests, how they rate oral proficiency and what might influence their perceptions and practice. The theoretical frame for the investigation will be presented in the next chapter.

Chapter 4: Theoretical Frame

This study aims at investigating teachers' perceptions and assessments of oral proficiency. It focuses on perception over practice and involves interviews and group discussions. It is argued that teachers' verbalized perceptions and actions reflect their understanding and intentions and therefore inform about the relation between what is stipulated in policy documents about oral proficiency, the intended curriculum, and how this is understood by teachers, the perceived curriculum.

The approach of this study is constructivist/interpretative. Thus knowledge is not seen as something passively acquired, but as constructed by the learner in social interaction with the world. The learners' own experiences are the basis for possible meanings given to the constructed knowledge. Thus, there are multiple socially constructed realities. The goal is to understand the various meanings expressed by the informants (Lincoln & Guba, 2000). They are therefore given the opportunity to articulate these meanings in interviews and discussions. The constructivist approach applied in the present study views knowledge as constructed by those active in the research process, i.e. the researcher is seen as an individual also involved in the process. The multiple meanings made apparent in the course of the study may be in conflict with each other, and perceptions of reality may change throughout the process of the study.

In qualitative research of the present kind, trustworthiness is suggested as a better term to use than validity (Lincoln & Guba, 1985). Trustworthiness is sought in the multiple understandings of the phenomena among the participants, and in the presentation of these various views in authentic quotations. The findings cannot be claimed to be generalizable, but the data from the interviews and discussions will demonstrate the varied perceptions of the phenomena in a group of skilled English teachers, and may generate hypotheses and concepts that can be used by other researchers exploring similar phenomena.

4.1 Teacher Perceptions

Teachers' perceptions of learning processes and the subject matter they teach have an impact on how they interpret curriculum, how they stage instruction in the classroom and how they assess. Perceptions are shaped and re-shaped in reciprocal interaction between training, experiences, external influences such as policy documents, and beliefs (Brookhart & Freeman, 1992; Pajares, 1992).

Experiences are, in the following, seen as personal, social as well as contextual in the way that it is the individual who decides what is relevant; in that the experiences are informed by social interaction, and in that time and context influence how experiences are interpreted by individuals (Apelgren, 2001).

Beliefs are understood as implicit personal ideas and theories strongly held by individuals, influencing the interpretation of training, experience and policy documents (Pajares, 1992). Teachers' beliefs are seen as having a significant influence on their professional pedagogical practice, as well as on assessment decisions (Black & Wiliam, 1998; McMillan, 2003). To study and understand teachers' actions and perceptions, the personal aspect has to be taken into account (Magnusson, 1998).

In educational research three inter-linked key dimensions of teacher knowledge have been distinguished: content knowledge, pedagogical content knowledge, and generic pedagogical knowledge (Shulman, 1986). These are connected in that subject matter knowledge is the basis for teachers' content knowledge, which, combined with pedagogical knowledge, is pedagogical content knowledge (PCK), defined as "...the most important part of the knowledge base of teaching…" (Gudmundsdottir, 1991, p. 411). The definitions of exactly what constitutes PCK have changed somewhat over the years (Hashweh, 2005), but it is still considered the 'most important part of teaching' and research still confirms the importance of PCK for effective teaching and better student achievement (Baumert et al., 2010; Ruohotie-Lyhty, 2013). Reorganizing specialist content knowledge into pedagogical and teachable units seems to involve adapting a disciplinary orientation, i.e. focusing on one or the other approach to the subject. "Having a point of view probably plays a major role in transforming content knowledge into pedagogical content knowledge." (Gudmundsdottir & Shulman, 1987, p. 67). The "beliefs component" in PCK is considered to be strong and might even

affect the quality of the PCK developed (Hashweh, 2005, p. 287). This 'point of view' is seen as related to the individual's perception of the subject matter and affecting all aspects of teaching the subject (Gudmundsdottir, 1991) and also the interpretation of policy documents and assessment (Hyltegren, 2014).

Additionally, the context, such as local circumstances, frames pedagogy in praxis, i.e. teachers work in places where temporary alliances and negotiations among individuals in vulnerable positions are predominant (Carlgren & Lindblad, 1991, p. 513). Thus, the local community of practice and traditions, i.e. the school culture further influences teacher perception and teacher action (Apelgren, 2001; Cuban, 2013; Hargreaves, 1994).

In this study, teachers are regarded as agents acting within certain structures, but not determined by them. Instead, they are viewed as interacting with these and forming new structures or affirming existing ones (Archer, 1995). Teachers' perceptions inspire these actions. The idea that the world is made real through people's actions and thoughts is essential (Chesebro & Borisoff, 2007), and the exploration of the meaning construed and expressed in interaction and dialogue between informants or between informant and researcher is the object of this study. The teacher as the story-teller is the basis for the investigation (Pope & Denicolo, 2000). Data collected include the feelings of the informants and the interpretations of what they have expressed, both explicitly and tacitly (Young & Tardif, 1992). The study thus has a natural setting and a mixed method, as well as a constructivist/interpretative approach.

4.2 Curriculum

The relevant policy documents form the basis for the curriculum implemented in the classroom, together with a number of different components such as time available, the composition and size of the group of learners and the material resources at hand. Another component in transforming curriculum from the intention in the documents to implemented curriculum is the teachers themselves, with their respective experiences, training and beliefs about education and the world at large, which contribute to forming their perception of the intended curriculum. A prerequisite for teachers interpreting and implementing the intended curriculum according to intentions is that it has been properly communicated and that the goals and

objectives are accepted as appropriate (cf. Riksrevisonen, 2004; Selghed, 2004; Tholin, 2006).

4.2.1 Intended and Perceived Curriculum

Curriculum outlines the setting for education. The word 'curriculum' traditionally defines the goals and the content of teaching and learning and describes the organization of education at various levels. Evaluation of education as well as assessment are now also included in the concept, stipulating what is valid knowledge and how it is to be measured (Broadfoot, 1996). In the current study the policy documents, i.e. *Lgr11* containing the syllabus for English and the national test for English with guidelines, are seen as the intended curriculum. The assessment of the oral test is seen as a manifestation of the intended curriculum.

The national curriculum can thus be defined as the intended curriculum, whereas the student results can be described as the attained curriculum. The level linking the two is the implemented curriculum, which in turn can be viewed as the perceived curriculum and the operational curriculum (Van den Akker, Kuiper, & Hameyer, 2003).

Table 2. Typology of curriculum representations

Intended curriculum	Ideal	Vision (rationale or basic philosophy underlying a curriculum)
	Formal/Written	Intentions as specified in curriculum documents and/or materials
Implemented	Perceived	Curriculum as interpreted by its users (especially teachers)
	Operational	Actual process of teaching and learning (curriculum-in-action)
Attained	Experiential	Learning experiences as perceived by learners
	Learned	Resulting learning outcomes of learners

(Van den Akker et al., 2003, p. 3)

How the outlines, laid down in policy documents like national curricula and subject syllabi, i.e. the steering documents, are perceived by teachers affects the school system and what students learn, as teachers mediate and transfer between the intended and the attained curriculum.

Curricula can be seen as content-focused, process-focused or results-focused (Sundberg & Wahlström, 2012). *Lgr11* is mainly a results-focused curriculum describing standardized levels of knowledge, decided in advance, to be measured, evaluated and compared over years, between units, regionally and nationally for reasons of fairness and quality. The political level decides on the objectives and the profession is responsible for the results. However, the political level supervises through quality controls and national testing, as well as through a number of regulations and guidelines for schools and teachers to abide by (Sundberg & Wahlström, 2012). In *Lgr11* traces of content-driven and process-driven designs, remnants from previous curricula can be found. This gives it a hybrid quality, which may lead to divergent interpretations of the intended curriculum. An expanding system of national tests and guidelines can therefore be viewed as a natural part of the current curriculum.

4.2.2 Syllabus

The national curriculum contains general educational goals as well as subject syllabi for each individual subject. In the 'Aim' in the syllabus for each subject, general values concerning that subject are communicated, the subject itself as a school subject is described and long term goals are defined. In the 'Core Content' it is stipulated <u>what</u> is to be taught, but not how. The 'Knowledge Requirements' describe the criteria to be reached for each grade level, i.e. they are the performance standards for the specific subject. The expected student results are aligned with the general curriculum and the core content stipulated.

4.2.3 National Test

The national test for English, with guidelines, is here seen as part of the policy documents, as they come with an extensive commentary including instructions and guidelines, thus contributing to concretizing the curriculum and the syllabus.

In this study, the relationship between teachers' understanding, the perceived curriculum, and the intended curriculum, as manifested in the national test of oral proficiency, will be explored.

4.3 Frame Factors

Teachers are responsible for what goes on in the classroom, they are the agents mediating curriculum. Teachers also shape curriculum (Clandinin & Connelly, 1992). Their perceptions and beliefs influence several curriculum components, which are all interconnected. Van den Akker (Van den Akker et al., 2003) has demonstrated this in a spider's web diagram, see Figure 2 below.

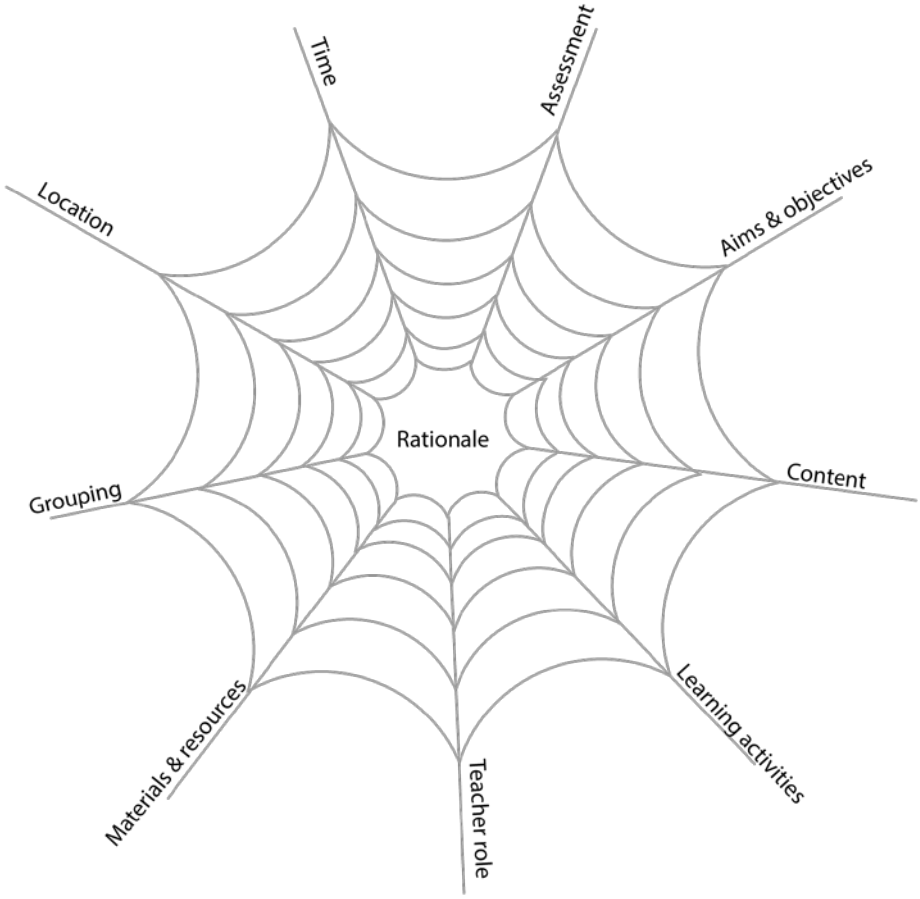


Figure 2. Curricular Spider Web

(Based on figure by van den Akker in Van den Akker et al., 2003, Fig.1)

The image above is to be seen with the 'rationale' in the middle and the different arms being interconnected in many places and leading from the middle. An alteration in one of the arms, making it longer or shorter, will influence the other arms and the connections between them. The structure is thus flexible, as a spider's web.

The web illustrates how the different components contribute to the whole. Teachers' beliefs and perceptions influence the decisions made concerning

several of these components, such as aims, objectives and content, the organization of learning activities and their teacher role. There are, however, a number of factors constraining their freedom of choice. Teachers do not normally decide on student groupings, locations, time or resources.

Seeing teaching as a process framed by certain external circumstances, such as time, resources and the characteristics of the group of students, can help explain the relation between the framework for teaching and the outcome of teaching (Dahllöf, 1967), as well as the decisions teachers make in the classroom (Lundgren, 1972, 1979). Frame factors will be employed in an effort to clarify why decisions made by teachers are to be found within certain limits. The factors may also point to why the actors, here teachers, interpret or re-interpret the (intended) curriculum in a certain way.

The theory of frame factors has been criticized for limiting the view on teacher thinking and teacher action by focusing too much on external logic and neglecting internal logic and teachers' practical reasoning (Carlgren & Lindblad, 1991; S. Lindblad, Linde, & Naeslund, 1999). There are, however, also suggestions proposing that frame factors might be discovered when teachers are asked to describe their life in school and to identify factors framing their practice (C. Gustafsson, 1999).

Frame factors will be used to explain some of the processes described by teachers. How oral proficiency, as defined in the curriculum and syllabus for English, is implemented in assessment of oral proficiency can be seen in relation to probable frame factors. Knowledge about these factors may shed some light on implementation processes (perceived and operational curriculum) and on teachers' perception of policy documents as well as the decisions teachers make when they assess students' oral performances.

Chapter 5: Method

In order to seek answers to the broader question on the relationship between teachers' understanding and the curriculum, and to find answers to the research questions on how teachers perceive, assess and grade the oral part of the national test for English, interviews with twelve informants were held. Six of these informants also took part in an assessment activity and discussed their grading of student interaction. Furthermore, relevant policy documents were read for comparison with the statements of the informants.

5.1 General Description and Overview of the Study

The informants were first contacted during the fall of 2012 and the last interviews were held during the spring of 2014.

Due to the introduction of the new national curriculum and the new syllabus for English as well as new grades, the 2013 national test for English was slightly different from earlier tests. (New policy documents and grades were issued in the fall of 2011, but students in year 9 were to receive grades according to the previous grading scale during the school year of 2011-2012, i.e. the national test for the spring of 2013 was the first in accordance with the new documents and grades.)

5.1.1 Timeline of the Study

The study was carried out between the late spring of 2013 and the late fall of 2014 as shown below in Figure 3.

Figure 3. Timeline of study

Firstly, during the spring term of 2013 an introductory qualitative semi-structured research interview was conducted with twelve teachers (Kvale,

1997). Secondly, based on the assumption that it is difficult for the informants to be detailed and accurate enough in their accounts of how they grade the oral tests, recorded examples of students' peer-to-peer interaction in the oral subtest of the national test were provided for the teachers to assess and then discuss in groups. All informants were invited to take part in this activity and six of the twelve informants agreed to do so. Meetings were then arranged in September 2013. Thirdly, during the late fall of 2013 and the spring of 2014, a second interview was conducted with the six teachers who had taken part in the assessment and group discussions. The process is illustrated in Figure 4 below.

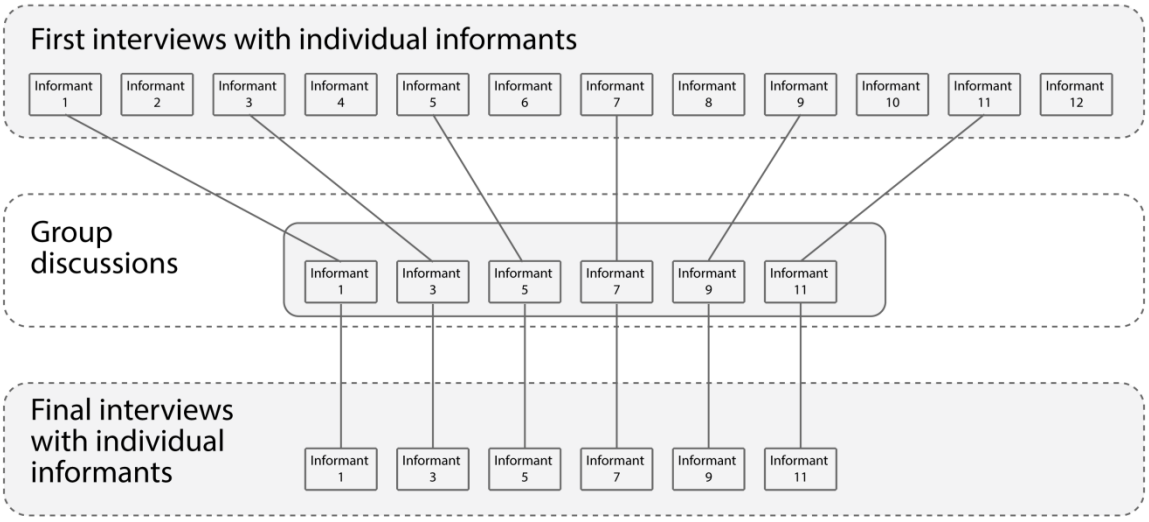


Figure 4. Informant interviews and group discussions

All twelve informants took part in interview 1, six informants joined the group discussions and were then interviewed a second time, as shown above.

5.1.2 Context and Informants

Letters (Appendix F) were sent (e-mailed) to principals of 24 schools in two different geographical areas in Sweden. Two separate locations were chosen to create variation in the sample. The 23 municipal schools with year 9 that were contacted were chosen from different parts within the two regions. One independent school[19] was contacted.

[19] Independent compulsory schools are open to all and the education should correspond to that provided in municipal compulsory schools. They have a different organizer/owner compared to municipal schools. The organizer may be a company, a foundation or an association. Independent compulsory schools are approved and inspected by the Swedish Schools Inspectorate.

Teachers from eight of the contacted schools volunteered to take part in the study. Two of the schools are located in a bigger city, one in a medium-sized town, three schools in suburban areas, and two in municipalities outside a bigger city. Both smaller and larger schools, as well as schools in different kinds of neighborhoods are included, creating a further variety of schools and students.

The contacted principals were asked to forward the information received to their teachers, for further communication with the researcher and possible participation. A first selection of participants was thus managed by the principals, who may have chosen to forward the information to their most positive and interested staff members. This selection process is thus a potential limitation in the study.

Ultimately, ten teachers from six of the contacted schools volunteered to take part in the study. Two further teachers were recommended by a fellow doctoral student, making the total number of informants in this study twelve and the number of schools eight.

All of the informants volunteered to take part in the study and expressed an interest in oral proficiency and how this skill is assessed. They self-selected into the study after having first received information from their principals (or in two cases directly from the researcher), and then from the researcher (Appendix G). They thus represent a purposive selection of informants with views and opinions on the phenomena in question (Teddlie & Yu, 2007).

The twelve teachers (ten women and two men) had been teaching English for 5-35 years at the time of the study. They were all qualified English teachers and had organized and graded the national tests of English several times. Seven of the informants were from one region and five from the other (see Appendix H). None of the informants had had any special training in assessment and grading.

5.1.3 Validity and reliability

Research needs to show both reliability and validity as scientific proof. Reliability emphasizes that the investigation needs to be sufficiently stable and robust, and so affected as little as possible by coincidence. Qualitative research in social sciences involving human judgment is difficult to repeat and is therefore sometimes claimed to be less reliable. Reliability is, further, a

prerequisite for validity, which requires that the methods used are actually measuring what they intend to measure in a study.

> In modern social science, the concepts of validity, reliability, and generalization have obtained the status of a scientific holy trinity. (Kvale, 1995, p. 20)

As validity, reliability and generalizability are crucial dimensions in all research, this becomes somewhat problematic in qualitative studies. However, Lincoln and Guba (1985) talk instead about the trustworthiness, credibility, dependability and confirmability of findings in qualitative research. This signals a perspective of multiple truths and multiple ways of knowing, as in a constructivist stance. It is not a question of total relativism, but rather a moderate position, where specific local, individual and community forms of truth are considered possible (Kvale, 1997).

For qualitative research, where the subjectivity of informants contributes to a degree of bias, validity can be seen as a matter of degree (Cohen, Manion, & Morrison, 2011). It is then a question of striving to minimize invalidity and maximize validity through constant checking and probing during the entire research process. It is also a matter of examining the questions posed to the data and then querying the findings (Kvale, 1997; Yin, 2009).

An interview is more reliable the more structured it is. On the other hand, the complexity and open-endedness of social interaction is not structured and therefore not easily captured in a strictly controlled and structured interview (Scheurich, 1995, pp. 241-249). Even if the wording of each question is the same, there is no guarantee that each informant understands it the same way (Silverman, 2006). An interview cannot be replicated and in that sense data collection is not repeatable. Different interviewers will carry out their interviews differently and then analyze data in ways that are impacted by their own beliefs. In this case semi-structured interviews have been used.

In qualitative studies involving a limited number of participants and often involving the researcher, the level of generalizability is limited. On the other hand, the findings from this kind of research contribute to a deeper understanding of the phenomena investigated.

Constructivist researchers advocate the notion of *trustworthiness* rather than validity and reliability (Guba & Lincoln, 1989), and, to a certain extent, this is also the approach taken in this study. Credibility, authenticity and generalizability as different aspects of trustworthiness will be explored.

Research is *credible* when the multiple experiences, perceptions and beliefs of the informants are adequately presented (Lincoln & Guba, 1985). In Chapter 6 of this study, the various statements of the individual informants are presented in quotations and summaries to illustrate the variation in perception, thus adding to the credibility of the findings. As every researcher takes a starting point in his/her individual experiences, perceptions and beliefs, the credibility of the researcher is also of importance. The researcher's individual beliefs are brought into the investigation and affect his/her positioning as well as the results of the study (Patton, 1990, p. 472). Personal and professional information about the researcher is therefore provided where applicable.

Authenticity can be strengthened through frequent use of direct quotations from informant narratives in the report on the findings. Ethical aspects, such as the protection of the informants' privacy and integrity during data gathering and in the compilation of the findings also impact the authenticity of an investigation. In this case, data were gathered in two separate interviews and in group discussions between the two interviews. The informants were not explicitly asked to comment on their own previous statements when interviewed a second time, but were given ample opportunities to further explore the same issues as in the first interview and in the group discussion. The presentation of the results in the next chapter includes extensive quotations from the interviews and group discussions but efforts have been made to avoid exposure of individual informants.

The informants of the present study represent a non-probability sample and a purposive sampling. Informants volunteering to take part and sharing an interest in the topic to be studied offer an opportunity for more depth in the study, but usually at the same time less breadth, because of the limited number of participants normally involved (Patton, 1990; Teddlie & Yu, 2007). Deep and rich data on the perceptions and beliefs of the informants are the focus of this study. Thus, the data are unique and not *generalizable*. On the other hand, findings from interviews with a group of skilled English-teachers who share a set of common characteristics and work conditions can serve as an example or illustration of a certain category of English-teachers (Hammersley, 1984; Ragin & Becker, 1992; Yin, 2009). The findings from a case study of this type can thus be said to potentially be transferable to other contexts (Lincoln & Guba, 1985).

5.2 Interviews

One aim of this study is to better understand the informants' perceptions, to allow conclusions to be drawn about what may influence teachers' assessment of oral proficiency. A qualitative approach has been taken. A small group of teachers has been interviewed to collect data, which will make it possible to see things from the interviewees' point of view.

An interview guide was used to help the interviewer keep track of topics and questions specified in advance (see Appendices I and J), allowing her to introduce them in a flexible order adapted to the interviewees, thus permitting them to freely elaborate on each topic at their own pace. The guide helped to make the data collection comprehensive and helped in the analysis of the data (Denscombe, 2007; Kvale, 1997; Patton, 1990).

The general purpose of the interviews was to encourage the teachers to elaborate on each topic in a spontaneous way, so as to capture the unique experiences of each individual. The main priority was to seek each informant's descriptions and interpretations of the phenomena. The topics for discussion were introduced by the questions or statements in the interview guide, and the idea was to try to elicit information during a conversation, rather than asking each question separately.

Using interviews as a method of collecting data requires the researcher to take certain factors into consideration. Seeing the interview not merely as an exchange of information, but as a social encounter demands that measures are taken to ensure a balanced interaction. Typically, the interviewer has an advantage as s/he usually decides what topics are to be discussed and how the interview is to be carried out. As the interviewer and the informant create and construct the interview together, the role of the interviewer has to be examined, making it evident what impact s/he has on the data elicited/co-constructed (Kvale, 1997, p. 183). Factors such as age, gender and ethnicity as well as general attitude, behavior and atmosphere, may have an impact on how the interviewee responds. The interviewees are under scrutiny and may therefore feel they want, or need, to comply with or please the interviewer. They may want to avoid certain questions or topics for private reasons. These aspects have to be openly disclosed in the recount of the interview and taken into consideration in the analysis, so that the research process is made transparent.

In this study the interviewer/researcher is an experienced senior language teacher who shares the same working conditions and professional frame of reference as the informants. This implies a shared reality, but may also lead to subjective interpretations and hasty conclusions. There is a danger of the interviewer seeing the interviewee in her own image and of looking for answers that support preconceived notions.

The informants were interviewed in the described way to obtain information on their experiences, notions and perception of oral proficiency, the oral part of the national test and the assessment and grading of this subtest (see Appendices I and J). The interviews were audio-recorded and then transcribed verbatim and have been analyzed to find patterns, similarities and differences between the different statements.

5.2.1 Description of Interviews

The interviews were conducted at a time that was convenient to the interviewed teacher. They took place at the interviewee's school to minimize the extra effort for the informant, but also to find a setting in which the teachers could feel at ease while they were sharing their thoughts and reflections. Each interview lasted about an hour. The first interview focused on the oral part of the national test, oral proficiency and the assessment of oral proficiency. The second was a follow-up interview after the group discussion, allowing the participants to comment on the group discussion and also to confirm their views and perceptions from the first interview.

To diminish the asymmetric power situation in the interviews, date, time and place for the interview were decided by the informants (Kvale, 1997; Young & Tardif, 1992). Leaving these decisions to the interviewees sometimes changed the power dynamics, as the teachers themselves were in charge at their respective work places and at times invited colleagues into the interview or involved themselves in other activities during the interview session. These circumstances add to the atmosphere and context of the oral interaction during the interviews and may have influenced the information elicited.

5.2.2 Method of Transcription

An interview is here seen as a contextual social interaction co-constructed by the interviewer and the interviewee, and the transcription of an interview is viewed as an interpretation of this interaction. During the transcription, oral

language, with its specific rules and format, is transferred into written language with a different set of rules and forms. Decisions have to be made on how to best transfer the recorded material into writing. A transcription model transferring what was said, with repetitions, hesitations, pauses, sighs and laughter was chosen. Hesitation or pausing might underline uncertainty or carefulness in verbalizing an idea or a belief, giving a nuance to what is said. Sighing may indicate problems or difficulties. Laughter here often signals a joke or irony in the interaction.

A decision was made to write down all the words uttered, using a simple but correct written form of Swedish. A choice had to be made about how to break a stream of spoken utterances into sentences, but when sentences were not complete or where there were 'false starts', nothing was added or omitted. Pauses were inserted when a silence occurred that was long for that particular informant. In the transcription, the researcher/transcriber strived to preserve each informant's style of talking. The result of the transcription is a hybrid form of text, not fully agreeing with an oral or a written format, as seen in the example below.

> Then naturally it is important that you have old recordings or that you try to meet to talk about assessment. Because it felt like … /*pause*/ eehh… /*hesitation*/ Well, it feels as if you need to, need to mull over what we are doing. Most of the time we landed in the same… We were relatively close, but… (Andrea)

The interviews as well as the group discussions were transcribed by the researcher/interviewer. In the transcriptions, the informants have been assigned alias names to anonymize quotations used. The audio-recordings varied in quality, not all words and phrases could be heard and are therefore not included in the transcripts. A second transcriber has listened to all the recordings and checked the transcripts to ensure that they are as accurate as possible. The second reader is a trained language teacher and former translator, now working in a different field of work. Minor deviations between the researcher's and the second reader's interpretations were discussed. When applicable, changes in transcripts were made.

Every transcription represents an interpretation of the audio-recording transcribed. No interpretation/transcription is more objective than the other – they are merely different written constructions of a recording of an oral interaction. It has been suggested that data and the relationship between meaning and language are contextually situated; they are unstable, changing

and capable of endless reinterpretation (Mishler, 1991, p. 260). In this case the transcriber (researcher) has tried to capture the ideas and thoughts of the informants as they were expressed in the audio-recorded conversations. The quotations in this text are translated from Swedish into English by the transcriber/researcher. This translation represents a second step of interpretation of what was uttered in the interviews, which has to be taken into consideration.

5.2.3 Ethical Considerations

How to phrase questions and how to word comments during an interview is a delicate task. It is important that the interviewees feel comfortable and willing to share their thoughts, beliefs and experiences, and that they feel respected as individuals and professionals.

The fact that the interviewer belonged to the same professional community as the informants helped establishing an atmosphere of trust. On the other hand, for the same reason, a distance must be kept to avoid too strong an identification between interviewer and informant. The balance between collegial recognition and a formal distance reflecting the interview situation was sometimes challenging. The overall aim was to identify the specific characteristics, beliefs and thoughts of each participant (Lincoln & Guba, 1985). As a consequence, it is of vital importance that the identities of the participants are kept secret and that all data remain anonymized.

In the e-mail initially sent to the teachers, the information about the study was presented (Appendix G). Additionally, a consent form was sent to the participants prior to the first interview (Appendix K). It contained the information about the study once again. This was done to give the interviewees a chance to re-read the information and again make an active decision to take part in the study. A copy of the signed consent form for each teacher taking part in the group discussion was handed out at the beginning of the activity, to once again review the aim of the study and give the informants a chance to decline further participation.

The data from interviews and group discussions have been analyzed in order to identify possible answers to the research questions and will be presented in the next chapter. The identity of the individual teachers and their schools will not be revealed or in any way made recognizable. All informants

were first given number codes during the analyses and later given alias names to simplify the presentation of the results of the analysis.

5.3 Group Discussions

After the first interviews, meetings for rating and group discussions were organized. The purpose of this activity was to explore what features of oral proficiency the informants noted and what rating criteria they referred them to when rating student interaction, and how they exchanged opinions on those features, criteria and ratings in a group discussion.

5.3.1 Description of Group Discussions

Six teachers from six different schools gathered to discuss and assess two examples of student performances. Two different group meetings were held, one in each geographical location. Both meetings were held outside of the respective schools (in one case at a university department and in the other case at a municipal office). The informants were asked to be prepared to assess and grade students in the oral sub-test, but were informed that no extra preparations were needed prior to the meeting. When the session started they were provided with the guidelines for the oral part of the national test, to be used as reference in their assessment and group discussion. They were also given time to study this material before they listened to the student examples.

Two audio-recorded examples of student interactions were played and the participants were asked to assess and grade the student efforts in the examples individually, in the same way they usually assess and grade the proficiency of their own students during this test.

The researcher was present in the room during the listening part and the silent individual rating part of the activity to make sure the technical equipment functioned properly. As soon as the recordings had been played and the teachers had finished rating, the researcher left the room.

After the informants had completed their individual assessment and rating of the student examples, they discussed the student performances and their individual ratings of these performances. The discussions were audio-recorded and later transcribed in the same manner as the interviews (see Section 5.2.2 above).

The material handed out to the participants was the actual test (Appendices C, D and E) as well as the guidelines for assessment of the oral

part of the national test (Appendix B). In addition to the knowledge requirements in the syllabus for English in the national curriculum, supplementary assessment factors to support the rating of the student performances are provided in the guidelines.[20] These supplementary assessment factors are a concretization of the long term goals and the knowledge requirements for oral proficiency in the syllabus.

5.4 Method of Analysis

Transcriptions of the interviews and the group discussions have been analyzed to find differences, similarities and patterns in the way the informants describe oral proficiency and how they organize and assess the oral part of the national test, as well as how they interpret the policy documents. Firstly, a *concordance-analysis* was carried out (see Section 5.4.1). Secondly, four *steps of Spradley's Developmental Research Sequence*, DRS, (Spradley, 1979) were used to uncover more implicit themes in the data (see Sections 5.4.2). Thirdly, the transcriptions of the group discussions were examined using two different *coding schemes* as outlined in Section 5.4.3. Finally, some possible *frame factors* influencing teachers' interpretation and implementation of the intended curriculum were investigated (see Section 5.4.4).

5.4.1 Words and Phrases Used by Informants

The first stage in analyzing the data was to compare the actual words the teachers used when they described oral proficiency and when they specified grading criteria in interviews and group discussions with those of the policy documents. This was done using AntConc,[21] a tool for analysis of concordance.

Correspondence and discrepancies in what words were used in the interviews and the discussions were investigated. The occurrences of specific words were compared between speaking situations, groups and with the wording in policy documents. The intention was to try to better understand the informants' perceptions of oral proficiency as well as their interpretations of the definition of oral proficiency in the policy documents.

---

[20] There are also commented, assessed and graded examples of student interaction (benchmarks) in the teacher information material. They were however not made available to the informants in this case for reasons of time.

[21] http://www.antlab.sci.waseda.ac.jp/software.html

5.4.2 Interview Analysis

The interviews captured numerous aspects of teaching, learning and assessment of oral proficiency in English. The transcripts of the interviews had to be scrutinized several times and the recorded sessions listened to repeatedly, in order to establish what the informants actually expressed. Listening to recordings, reading transcripts, organizing and analyzing data became an iterative process.

In analyzing the statements of the informants, parts of Spradley's ethnographic interview analysis, the Developmental Research Sequence/DRS, (1979, 1980) were used. Steps from the DRS have been used as analytical tools to uncover and understand implicit subject-specific conceptualizations held by informants in several other case studies in educational research (e.g., Carlone, 2004; Carlone & Johnson, 2007; Lee, Nargund-Joshi, & Dennis, 2010). The DRS has a total of 12 steps and was originally developed to clarify semantic, cultural knowledge shared by a community of individuals and used primarily for ethnographic studies. In this study it has, however, been used in an adapted version, similar to that developed by Lee et al. (2010), to show the different conceptualizations of the participants. Aligned with Lee et al., this study only uses four steps and the units of analysis have been moved from the level of culture to the level of individuals; yet, the term 'cultural theme' is kept as a notion for certain possible orientations among language teachers.

Spradley has been criticized for a positivist stance in the DRS. Using the DRS as a tool and at the same time acknowledging the interpretative engagement of the researcher is however possible, as long as the researcher engagement does not overpower the meanings the participants communicate (Lee et al., 2010, p. 45). The analysis made in this study is first and foremost based on the statements made by informants. Quotations, which frequently are used in the study, let the voices of the participants come to the fore. In this context, the DRS is used as a tool to help clarify and categorize the informants' stated opinions.

Another comment on the DRS has been that the distinction between individual and culture is not made clear. In their development of parts of the DRS, Lee et al. use Giddens' structuration theory (Giddens, 1979) to get around this problem. Their suggestion is thus that "[…] the study of individual conceptualizations is interpreted as a recursive expression of both

individuals and culture." (Lee et al., 2010, p. 46) This innovation is used in the present study as well.

Lee et al. claim that "[…] Spradley's analysis provides one of the more comprehensive set of strategies for understanding the linguistic attributes of participants' lived and talked about experiences."(2010, p. 47) At the same time they conclude that a limitation of the method is that it is complicated and time consuming. In the present study, the method is combined with the concordance analysis, coding schemes for grading criteria and the investigation into frame factors.

Firstly, semantic relationships (Spradley, 1979, 1980) were used to organize and categorize the great number of diverse, complex statements made by the informants. This tool helped sort the data, and create a more distinct image of how the statements were interrelated. Semantic relationships can be of different kinds, as shown in Table 3.

Table 3. Spradley's Semantic Relationships (1979, p. 111; 1980, p. 93)

1. X is a kind of Y	4. X is a reason for doing Y	7. X is a way to do Y
2. X is a place in Y, is a part of Y	5. X is a place for doing Y	8. X is a step/stage in Y
3. X is a result of Y, is a cause of Y	6. X is used for Y	9. X is a characteristic of Y

When the data were viewed in the light of semantic relationships between different items (X) and categories (Y), as seen in Table 3, links that otherwise were less noticeable were uncovered. Organizing data according to these semantic relationships involved gathering items with the same relation to a 'cover term' (i.e. name of category) in an analysis. When items, (i.e. statements by informants), were grouped according to their semantic relationship to different 'cover terms', specific 'domains' (larger units of knowledge/categories) were revealed.

During the process of looking at the different items, what 'cover terms' they had a semantic relation to and if they in turn included further items or could be grouped together, the researcher discovered various new combinations. This is the first step, the domain analysis, of the DRS.

In Table 4 an example of a domain analysis is shown. Statements on *grading criteria for oral proficiency* were first put into a domain analysis using 'X is a part of Y'.

Table 4. Domain Analysis using *'X is a part of [grading criteria for oral proficiency]'*

Included terms	Semantic relationship	Cover Term
Using nice vocabulary and phrases	*is a part of*	*grading criteria for oral proficiency.*
Asking questions		
Content		
Using ok grammar		
Assertiveness		
Does not get stuck		
Speaks without too much hesitation		
Using strategies to keep talking		
Has a rich vocabulary		
Using idiomatic expressions		
Pronunciation… etc.		

All the 'included terms' in Table 4 have the same semantic relationship to the 'cover term', i.e. they are all *parts of* the grading criteria for oral proficiency.

Secondly, the different 'included terms' gathered were tested to see if they in turn included further sub-categories or could be grouped together. For instance, some of the 'included terms' in Table 4 above could be grouped in a 'domain analysis' according to 'X is a part of *fluency*', creating a sub-domain to grading criteria for oral proficiency. Further sub-domains were found and could be ordered in a taxonomy as shown in Table 5 below.

Table 5. Taxonomy using *'X is a part of [the grading criteria for oral proficiency]'*

Taxonomic Name (Y)	Item (X) (sub-domains)	
Grading criteria for oral proficiency in NT for year 9	Vocabulary	…
	Interaction	…
	Fluency	Handles difficulties such as not finding a word
		Does not get stuck
		Speaks without too much hesitation
		Uses strategies to keep on talking
		Assertiveness…etc.
	Grammar	…
	Variation	
	Pronunciation	…
	Content	…

Thus, a number of various definitions of a phenomenon were made apparent. All items (Xs) in Table 5 have included terms, as shown here with the item *fluency*, grouping the different statements by the informants, but these are left out here for reasons of space. The taxonomies in Chapter 6 are displayed without the included terms in the tables.

Thirdly, further parts of the DRS were used to extract more information from the data and explore the informants' more implicit and tacit individual conceptualizations of oral proficiency and of instruction of oral proficiency. The data were analyzed in an attempt to establish what the informants regarded as the main points of discussion ('domains') when they talked about a topic (in the example below: *the beliefs on teaching and learning oral communicative proficiency*). To identify the main points, the researcher sought to establish what items the informants frequently referred to, or returned to, on the topics in the interviews, and how these items were inter-related. Within each 'domain' a taxonomic analysis was carried out to find the relations between different items in that specific domain and then these were organized in sets showing different aspects of the domain (see the example of taxonomic analysis in Figure 5 below). When the researcher was probing the data to find the internal structure within the domains, major aspects or dichotomies within each 'domain' were uncovered. The next step was to sort and group the various items into 'dimensions of contrasts'. This step of the DRS can be used to establish what a phenomenon is NOT, according to informants, in order to more precisely define/understand the meanings of the participants. In this case the opposing views are usually held by different informants. In a last step, the relationships between the original 'domains' and how they are linked to the topic (in the example: *beliefs on teaching and learning communicative oral proficiency*) were sought.

Figure 5 illustrates the analysis of the 'domain' *social factors*. In the taxonomic analysis several aspects of the 'domain' are contrasted, resulting in two main aspects. The componential analysis identified two sets of divergent beliefs on teaching and learning. Based on this, two separate cultural themes can be derived, as seen below.

Step in Spradley's analytic strategy	Teachers' beliefs on teaching and learning oral communicative proficiency
Domain analysis: A search for the larger units of cultural knowledge	Identified domains: ***Social factors*** Communication Correctness Content
Taxonomic analysis: A search for internal structure of domains that leads to identifying different aspects or contrasting sets.	***Social factors:*** • Students are afraid to speak English in class • Students gladly speak English in class
Componential analysis: Process of searching for contrasts, sorting them out, grouping some together as dimensions of the contrast and entering all this information onto a paradigm.	***Social factors:*** Motivating and encouraging students to speak is important. Therefore: • Instruction partially in Swedish for clarity • Reminding students to speak English all the time • Emphasis on individual and pair work (avoiding large groups) vs. Creating opportunities for students to develop their resp. levels of oral proficiency is important. Therefore: • Instruction in English only (speaking English also outside of classroom) • Variation in interactive oral activities and methods
Cultural themes: A search for the relationships among domains and how they are linked to the culture as a whole.	***Social factors:*** A safe and friendly atmosphere with clear instructions (when needed in Swedish) and basic linguistic tools provided encourage students to speak English. Student-friendly themes and general topical issues as well as interactive and challenging activities encourage students to develop and stretch their oral proficiency in negotiation of meaning.

Figure 5. Beliefs on Teaching and Learning Oral Proficiency

When the researcher explored the above domain through the modified steps of Spradley's DRS as seen in Figure 5, different orientations and perceptions

became visible. This second tool for analysis, inspired by Spradley (1979, 1980) and the modifications by Lee et al. (2010), was used for parts of the data collected in order to answer the research questions.

5.4.3 Classification of Assessment Criteria

To further investigate how teachers assess and rate students' peer-to-peer interaction in the oral part of the national test, the transcripts of the group discussions (see Section 5.3) were coded to enable an analysis of the elements of oral interaction mentioned in the discussions. One set of codes used for the classification followed the supplementary assessment factors closely, as shown in Table 6 (see also Appendix B).

Table 6. Coding Key for Supplementary Assessment Factors

Codes for content	Explanation	Codes for language and expressiveness	Explanation
CO:INT-CLA	Intelligibility and clarity	LA:CS-CC	*Communicative strategies*: to develop and carry on conversation
CO:RIC-VAR	Richness and variation: examples and perspective	LA:CS-SP	*Communicative strategies*: to solve linguistic problems through rephrasing, explaining and clarifying, etc.
CO:COH-STR	Coherence and structure	LA:FLU	Fluency and ease
CO:ADA	Adaption to purpose, recipients and contexts	LA:VOC	*Richness, variation, clarity and assertiveness*: vocabulary, phraseology and idiomatic expressions
		LA:PRO	*Richness, variation, clarity and assertiveness:* pronunciation and intonation
		LA:GRAM	*Richness, variation, clarity and assertiveness:* grammatical structure
		LA:ADA	Adaption to purpose, recipients and contexts

The supplementary assessment factors are divided into *Content* and *Language and expressiveness* and the interactional features are embedded in these two assessment categories. There are therefore only two main categories in the table above.

In order to uncover what specific interactional features were salient to the informants, 15 of the categories from a coding scheme by May were tested as an extra coding of the same transcripts (May, 2010), displayed in Table 7.

Table 7. Coding Key Interactional Effectiveness inspired by May (2010)

Codes for interactional effectiveness	Explanation
UND	understands interlocutor's message
LIS	listens to interlocutor
RES	able to respond to interlocutor/build on interlocutor's ideas
UCS	uses communicative strategies
EXP	able to express ideas and opinions
CLA	asks for clarification, confirmation or clarifies own ideas, concept checks
QUA	quality of the interaction
ASK	asks for partner's opinion
MAN	manages/controls interaction- usually mentioned positively
DOM	dominates the discussion- usually mentioned negatively
ASS	assertiveness, demonstrated through communication
WOR	working together cooperatively
HEL	helps partner out
INT	intelligible
EFF	effectiveness, in general

The second classification scheme (Table 7 above), inspired by May (2010), was introduced to contrast the results from the analysis based on classifications from the supplementary assessment factors from the national test guidelines. The idea was to study in more detail what features raters noticed and how they ascribed these features to rating criteria.

During the coding of the data, the researcher had to take several aspects into account. In both groups, the participants were strangers to one another. Because of this, open conflicts or strong arguments between participants were not to be expected. The groups listened to and assessed the same examples, but differed in constellation of participants concerning experience, age, education and gender.

When counting the number of remarks made on the student efforts, nods or verbal agreements like "Mmm…" were not included, which needs to be remembered. Still, parallels and variations in what was salient to the informants and how that may impact their ratings could be detected. Deviations and similarities exposed in the analysis will be explored in Chapter 6 of this text.

5.4.4 Frame Factors

In order to explore what may affect how the informants perceive and enact the intended curriculum, frame factors have been used as a tool. These factors

are here used to explain why certain educational processes occur within certain limits, as suggested by Gustafsson (C. Gustafsson, 1999).

There are multiple factors framing pedagogic processes. The focus here is on the "traditional" frame factors, such as time, resources and student grouping, as well as additional factors exposed by the informants in the interviews (C. Gustafsson, 1999).

In a first step, a number of plausible frame factors were chosen and investigated, including factors referred to by the informants. Then, in a second step, these factors were tested through a reverse process of scrutinizing the transcripts of the informants' statements about their teaching and assessment processes in interviews and discussions, and testing the statements against the plausible factors previously identified.

In a third and final step, the factors acknowledged in step two were recognized as the factors possibly influencing the interpretation and implementation of the intended curriculum, transferring it to the perceived curriculum (see Table 2, Section 4.2.1).

The above methods have been used to analyze the data and answer the research questions. The results of the analyses will be presented below in Chapter 6.

5.5 Limitations of Study

There are limitations that need to be considered when conclusions are drawn from the findings in this study.

First of all, it is important to point to the limited number of informants. Further, the participants have chosen to take part because of their interest in the subject. They are, in other words, confident sharing their opinions and concerns regarding assessment and oral proficiency. Therefore, they represent themselves, i.e. confident, expert teachers of English, not English teachers in general. The limited number of participants and the interest they express in oral proficiency give an opportunity for depth in the data gathered, but will cause a lack in breadth (Patton, 1990; Teddlie & Yu, 2007). This is in line with the constructivist/interpretative frame of the study, which emphasizes multiple understandings of the specific phenomena.

Orientations with regard to oral proficiency, as presented here, are abstractions of possible orientations. They are collective descriptions of teacher orientations, and individual aspects are therefore lost in the attempt to

simplify and reduce the data. However, it is argued here that the presented orientations can portray possible orientations held by teachers, based on their varying perceptions of oral proficiency (Hammersley, 1984; Ragin & Becker, 1992; Yin, 2009).

A further limitation is that all data input is self-reported by the informants. It reflects their perceptions and their descriptions of their actions. Data were gathered in interviews and group discussions only. No observations are included in this study. Using interviews for data collection involves the researcher as an actor in the process. As the researcher in this case is an experienced English teacher herself, there is a strength in her knowing the informants' context, but it also represents a weakness in that it may create a bias (Kvale, 1997; Patton, 1990). In the attempt to stay true to the informants' narratives, quotations from the interviews and the group discussions have been used extensively in the report on the findings.

The policy documents referred to in this study are limited to the national curriculum, the syllabus for English and the national test for English with guidelines. These documents were chosen because they are assumed to be the most well known and most consulted texts on the subject of assessment of oral English. There are further regulations, diagnostic materials and advisory materials supporting teachers of English that are not included here and that may also have an impact on the assessment practice of English teachers.

These limitations need to be kept in mind when reading about the results from the investigations in the next chapter of this thesis.

Chapter 6: Results

An analysis of the data collected should help answer the research questions on how the informants perceive, manage and grade the oral part of the national test, and to what extent this reflects the intended curriculum. This section presents the findings grouped into major themes, which emerge from the analysis of the interviews and group discussions.

A main theme is how the informants perceive (1) oral proficiency. Conclusions on how the informants understand the proficiency have been drawn from their statements on the nature of oral proficiency, what criteria they believe capture it and how they describe their own instruction aimed at developing their students' oral proficiency.

Another theme relates to (2) policy documents, how they are viewed by the informants and what is perceived to be stipulated in them on oral proficiency. The documents are then compared with the statements of the informants.

A third theme is how the informants perceive and organize (3) the oral part of the national test. How the informants (4) assess and grade oral proficiency in the national tests is a further theme emerging from the data.

The last theme investigates the influence of (5) frame factors on the assessment practice of the teachers.

6.1. Teachers' Understanding of Oral Proficiency

Oral proficiency is difficult to pinpoint. It involves many different aspects which makes it challenging to define. The element of interaction and co-construction adds to the complexity of the phenomenon.

6.1.1 Defining Oral Proficiency

To examine how the informants view oral proficiency, their different statements on the topic were grouped in taxonomies inspired by Spradley's (1979) semantic relationships (see Section 5.4.2). Many informants focused on the ability to express ideas, on fluency and on a rich vocabulary when they

described good oral proficiency for students in year 9, as shown in Table 8 below.

Table 8. Defining Oral Proficiency

Taxonomy Developed Using 'X is a characteristic of [good oral proficiency in year 9]'		
Taxonomic Name	**Item (X)**	**Informant**
Good oral proficiency in year 9	Being able to express oneself	Andrea, Eva, Gabriele, Kari, Kim, Linda
	Speaking fluently	Chris, Hanna, Ingrid, Kari, Linda
	Using a rich vocabulary	Andrea, Ingrid, Jennifer, Kari, Linda
	Being able to carry on a conversation	Hanna, Ingrid, Jennifer, Linda
	Making few mistakes (that do not destroy communication)	Andrea, Eva, Kari, Linda
	Being able to adapt	Eva, Gabriele, Kari
	Using ways to get around problems	Eva, Hanna, Kari
	Wanting to interact	Ingrid, Kari, Kim

(Sub-categories are not shown here, see Section 5.4.2, Table 5, p.68.) The statements are translated from Swedish by the researcher.

The expressions used by the informants to characterize oral proficiency are general and descriptive of oral communicative ability in a wider sense, as can be seen in the table above. The wording is often close to that in the policy documents (see Appendix A). Some teachers also include the *willingness* to interact and share ideas in their descriptions. This can be seen as an echo from earlier syllabi that include *wanting to use English* as one of the goals for English instruction (see Section 2.2.4).

When the informants reflected on the criteria used for assessing and grading oral proficiency they shifted emphasis to more easily measureable entities, as shown in Table 9 below. 'Vocabulary' is exemplified as 'phrases' and 'idiomatic expressions' (instead of 'using a rich vocabulary' as in Table 8 above) and 'fluency' is broken down into 'assertiveness' and 'getting around word problems' (instead of 'speaking fluently'), for instance.

Table 9. Grading Criteria for Oral Proficiency

Taxonomy Developed Using 'X is a part of the [grading criteria for oral proficiency in year 9]'		
Taxonomic Name	**Item (X)**	**Informant**
Grading criteria for oral proficiency in year 9	Vocabulary (phrases, idiomatic expressions)	All informants
	Fluency (assertiveness, getting around word problems)	All informants
	Interaction (dialogue, conversational skills, being an active, supportive speaking partner)	All informants except Hanna
	Grammar	Andrea, Chris, Eva, Ingrid, Kari, Kim, Linda
	Variation	Andrea, Chris, Dominique, Eva, Hanna, Jennifer
	Pronunciation (intonation, prosody)	Chris, Dominique, Hanna, Ingrid, Kari
	WHAT they talk about (content)	Dominique, Eva, Jennifer, Kari, Robin

(Sub-categories not shown, see Section 5.4.2, Table 5, p.68.)

The three top criteria listed in Table 9 are 'vocabulary', 'fluency' and 'interaction'. The ability to 'express one-self' is not included here. In their reflections on grading criteria, the informants again stay close to the wording in the policy documents, indicating that they are familiar with the syllabus and the knowledge requirements, as well as the assessment factors for the oral subtest.

Despite the stress on 'interaction' as a criterion for oral proficiency and the closeness to policy documents, the word 'strategy' is rarely used by the informants. It appears to have been seen as a new notion added to the criteria in the new syllabus, although it was previously included in the commentary to the syllabus in *Lgr80* and included as a goal in the syllabus in *Lpo94* (Skolverket, 2000a, 2000b; Skolöverstyrelsen, 1990; Utbildningsdepartementet, 1994). Some of the informants found it difficult to assess and to properly comprehend.

> Q1: Well, what's new are all those strategies, I guess, that they talk about. And so… /*hesitation*/ And they are pretty hard to assess, I think… (Eva)

> Q2: And then you try to find something that could count as a strategy. (Jennifer)

On the other hand, most of the informants described how they expected their students to be able to work around problems in the interaction by rephrasing,

describing, using synonyms, explaining, and asking in order to make themselves understood or to support the interaction. 'To keep the conversation going' was another expression used, indicating that the concept of a language strategy is not unknown to the informants and that they are familiar with the definition of communicative strategies in the supplementary assessment factors for the oral test (see Appendix B).

The perceptions of the informants appear in explicit as well as implicit ways. In Figure 6 below, major beliefs on oral proficiency that were detected when a modified version of Spradley's DRS was used (see Section 5.4.2) are shown:

Step in Spradley's analytic strategy	Teachers' beliefs on oral proficiency
Domain analysis: A search for the larger units of cultural knowledge	Identified domains: • Content • Interaction • Correctness • Speaking
Taxonomic analysis: A search for internal structure of domains that leads to identifying aspects and/or contrasting sets.	• To elaborate on matters outside the immediate personal sphere ←→ to express personal ideas and feelings • To interact in conversations, discussions and debates ←→ to make oneself understood • To speak clearly and understandably ←→ to be able to speak correctly • To speak fluently ←→ to be willing to talk and share information
Componential analysis: Process of searching for contrasts, sorting them out, grouping some together as dimensions of the contrast, and entering all this information onto a paradigm.	Interaction, collaborative efforts and content are in focus (authentic negotiation of meaning) vs Individual linguistic efforts in varied interactions are in focus.
Cultural themes: A search for the relationships among domains and how they are linked to the culture as a whole.	• Oral proficiency is being able to make oneself understood • Oral proficiency is using the target language orally in an acceptably correct manner. • Oral proficiency is being able to take part in discussions and negotiations of meaning with other individuals sharing one's own opinions and reflections on general topical issues.

Figure 6. Beliefs on Oral Proficiency
(←→ indicate contrasting sets or aspects of a domain).

In Figure 6, the informants' statements about their beliefs on oral proficiency were first grouped into four main domains; content, interaction, correctness and speaking. In a second step of the DRS (see Section 5.4.3), a taxonomic analysis was carried out within each domain. The major aspects for each domain are shown above in Figure 6 under 'Taxonomic analysis'.

A third step was to group the aspects into two opposing dimensions of beliefs among the informants, as shown in Figure 6 under 'Componential

analysis'. In the last step, the relationships between the original domains and how they are linked to the topic ('beliefs on oral proficiency') were uncovered. The informants differ in how they emphasize interaction, linguistic accuracy and negotiation of meaning, respectively, in their definitions of oral proficiency.

Commenting on the first domain, 'content', most of the informants maintained that students need to be able to convey an opinion, whereas some found it acceptable that students can communicate some sort of message in a broader sense. Four of the informants stated they find it important that the students are able to leave their own private sphere and engage in more general topics in their conversations, in order to show an adequate level of oral proficiency.

> Q3: Then they must have moved focus from this circle around themselves to the wider circle and the world. (Andrea)

A second domain apparent in the statements was interaction. To ask questions, to comment, to explain and to encourage a speaking partner to keep a conversation going are some of the communicative features mentioned by most informants. For higher grades, the demands were said to be higher:

> Q4: But for those who want up to a C or an A, then I want you to kind of be the one pushing in discussions. (Robin)

Some informants seemed to be satisfied if the students are *confident enough* to speak and to make themselves understood in a conversation.

> Q5: It is important that you are confident enough to speak and that you dare make yourself understood. /…/ to dare show what you are able to do so, not being afraid of trying different ways, even if you can't, trying to express yourself in a different way… (Hanna)

To dare and to not be afraid, *to be confident enough* to speak English, was touched upon by all informants as being vital. This is however not mentioned as a criterion in the current policy documents.

Wanting to speak is also mentioned by several informants.

> Q6: If you look at them, the criteria for E for the oral part here in the guidelines, for example, if you WANT to … interact with another person, but may find it difficult – I don't think that is proficiency, that you want to. I think there is a difference… that you should, you should have a higher level of proficiency to reach an E than we have today. (Ingrid)

To *want to* interact is not part of the supplementary assessment factors or the current knowledge requirements (i.e. not a "criterion for E"), but is mentioned in an introductory paragraph in the guidelines describing the functional view of language (see Appendix B). It is also mentioned in a descriptive way in the commented examples of oral interaction in the guidelines. However, 'the *willingness* and ability to take part in a conversation' were stressed as foci for assessment of oral interaction in the instructions for older examples of oral national tests as late as spring 2013.

> Oral proficiency is assessed with a stress on *the willingness and the ability to take part in a conversation and on the ability to convey content in an understandable way*. An oral performance showing many linguistic deficiencies can be passed if the message is still conveyed and the communication functions. (Skolverket, National Tests, Example task, Assessment guideline, italics in original, *my translation*)[22]

As mentioned above, to *want to* and *to feel confident enough to* speak English were goals in the syllabus for English in the curriculum from 1980, *Lgr80* (Section 2.2.4). It thus seems as if previous descriptions of oral proficiency in policy documents prevail in the perceptions of some of the teachers in the study.

A third domain addressed by the informants was correctness. Several of the informants pointed out that they were not focusing on mistakes, but rather on what the students manage to express in their interactions.

> Q7: …because you should look positively at what you do with the language, you know. (Kim)

However, some of the teachers talked about the kind of mistakes that disturb communication and make the interaction difficult or impossible.

> Q8: They [the mistakes] destroy so much that /*pause*/… /…/ it [the performance] doesn't really become… It doesn't become informative and it doesn't become communicative either. (Kari)

Virtually all informants referred to grammatical accuracy when speaking about grading criteria. Most of them stated that correctness was not and should not be emphasized, but that correctness can make a difference to clarity in a message and for the higher grades awarded, grammatical accuracy at a certain level was taken for granted.

[22] Retrieved from www.skolverket.se in May 2013, (see Appendix L).

A fourth domain was speaking in itself. A few participants found it difficult to encourage their students to speak English in the classroom. Several of the informants had difficulties with one or two students not wanting to speak during lessons. This was of great concern and a frustration for these teachers, who tried all sorts of activities and measures to motivate their students and encourage them to practice and demonstrate their proficiency. However, a majority of the students in the informants' classes were said to be very good at speaking English.

> Q9: Often they are very good orally. That is usually not a problem. They enjoy speaking [English]. (Linda)

Most teachers appreciated their students' fearlessness and readiness to discuss and share ideas in English during lessons and also outside the classroom in the school hallways and the school cafeteria. The informants found they spend less time teaching pronunciation than they used to and are often astonished by the wide vocabulary of their students. However, insecurities in pronunciation and limitations in vocabulary can be hidden in the readiness to talk incessantly, according to several informants.

To summarize: The informants hold different views on oral proficiency and what constitutes the essence of the skill. There is an emphasis on different aspects, but the ability to make oneself understood is a basis for all. Some informants stress basic linguistic knowledge and correctness as a first step towards good oral proficiency, whereas others view oral proficiency as a social activity and expect the students to be able to engage in conversations, sharing ideas and opinions on a wide variety of topics using the language at hand, as a first step on the way to reaching good linguistic skills and accuracy.

6.1.2 Oral Proficiency in the Classroom

In order for the researcher to reach a deeper understanding of how they perceive oral proficiency, the informants were asked to reflect on their own teaching aimed at developing oral proficiency.

All informants mentioned 'speech' and 'oral interaction' as two separate parts of oral proficiency. The prepared speech is a way of demonstrating one part of oral proficiency, production, but it has little to do with communicative interaction, according to the informants. All teachers engage their students in both activities during lessons.

Concerns about how to encourage students to speak as much English as possible and create authentic situations for the exchange and negotiation of meaning during lessons were expressed by a majority of the participants when they talked about activities in the classroom. Creating a positive classroom environment to motivate all students to feel at ease and confident when speaking the target language was stressed by all informants. Some of the teachers used English only in class, as well as during breaks. Some of the informants stated that they translate what they say into Swedish to help all students understand and speed up information. Most teachers said they switched to Swedish when teaching grammar.

All informants stated that their students, in general, have good English speaking skills and that their oral proficiency is better than other language skills. Five teachers found their students eager and willing to speak English "all the time", both to one another and to teachers. Several of the informants expressed their amazement at how competent their students are when they engage in projects in class. One teacher expressed his/her joy when listening to his/her students working:

> Q10: And there are those who basically, well, they really speak very well. /…/You know, I [have them] do some exercises and I find myself smiling stupidly. You kind of go: Ah, shit are they good or what! (Robin)

Still, there are students who are not confident enough to speak English in the classroom, in presentations or in dialogues with a partner. One teacher found it difficult to inspire the students to speak English at all during lessons. Another teacher said it is difficult to prompt the students to speak English in front of the class or even in a smaller group. Some of the informants mentioned one or two quiet students with whom they had to particularly struggle to make them speak. For these students, the oral tests become especially important since these tests give them one of few opportunities to demonstrate their oral proficiency, as they are too inhibited to interact in a classroom situation. All informants mentioned an extensive vocabulary as a criterion for good oral proficiency, but few mentioned dedicated exercises intended to make students extend their vocabulary. Two of the informants stated that vocabulary is better acquired through reading and writing. In the interviews the informants readily talked about and gave numerous examples of oral exercises and projects that are carried out during lessons.

Using the analysis inspired by the modified version of Spradley's DRS (see Section 5.4.2), two main categories of teaching orientations can be traced in the comments and reflections on the topics in the interviews, as shown in the figure below.

Step in Spradley's analytic strategy	Beliefs on Learning and Teaching Oral Proficiency
Domain analysis: A search for the larger units of cultural knowledge	Identified domains: • Social factors • Communication • Correctness • Content
Taxonomic analysis: A search for internal structure of domains that leads to identifying contrasting sets.	• Students are afraid to speak English ←→ gladly speak English • Emphasis on oral exercises ←→ focus on oral interaction • Grammar as a necessary basic tool ←→ correctness to further amplify content conveyed • Focus on language as content ←→ focus on ideas, thoughts and views as content
Componential analysis: Process of searching for contrasts, sorting them out, grouping some together as dimensions of the contrast, and entering all this information onto a paradigm.	Focus on basic vocabulary and grammar, language accuracy, to enable safe communication about ideas, thoughts and views in an encouraging and supportive classroom atmosphere. (HOW it is said in focus) vs. Focus on communicating ideas, thoughts and views in interaction in a creative and positive atmosphere. Linguistics viewed as a means to clarify and illuminate message delivered in interaction. (WHAT is said in focus)
Cultural themes: A search for the relationships among domains and how they are linked to the culture as a whole.	Creating an atmosphere where every student feels at ease to express him/herself is the responsibility of the language teacher. To facilitate this, o some teachers express that students need formal language basics to feel safe to engage in exchange of ideas and, o other teachers state that students need engaging topics to exchange ideas and opinions about.

Figure 7. Beliefs on Learning and Teaching Oral Proficiency
(←→ indicate contrasting sets or aspects of a domain).

Firstly, as can be seen in Figure 7, four main domains were found in an analysis of the various statements made by the informants on learning and teaching oral proficiency: social factors in the classroom, communication in the target language, correctness of language, and content of interaction. These themes were mentioned by all the teachers, and they are important in different ways to their teaching aimed at developing oral proficiency. Secondly, in the taxonomic analysis different aspects, sometimes contrasting, other times complementary, became apparent. Thirdly, grouping the aspects in the componential analysis showed basically two different dimensions of emphasis among the informants.

Finally, in the last step of the analysis, the relationships between the original domains and how they are linked to the 'beliefs on teaching and learning communicative oral proficiency' were investigated. This investigation showed the main themes among the beliefs of the informants, as illustrated in Figure 7 above. The analysis indicates that the teachers in the study either emphasize the need to supply students with linguistic tools and for them to develop confidence to speak English through basic linguistic skills (HOW it is said, grammatical correctness, vocabulary, idioms, etc.), or they emphasize the need to supply students with interesting topics to motivate them to develop rich content (WHAT is being said, being able to express opinions, etc.).

Most informants emphasized the importance of a helpful and friendly atmosphere among students. Some were concerned about their students' unwillingness to speak freely in the classroom, others mentioned quiet and introvert students who lack the confidence to speak during lessons. Including everyone in oral interaction in the classroom was a priority for all the informants and they mentioned examples of different long-term and short-term strategies to motivate their students. Two teachers asked their students to audio-record themselves when reading texts or talking, to make it easier for them to feel confident expressing themselves and using the target language.

> Q11: We do a lot of audio-recordings of different kinds that they are to send in to me – because that takes away a lot of anxiety for some. They don't dare display their full range when it comes to pronunciation and fluency when they stand up in front of the class. /…/ This [audio recording themselves] has given many students the opportunity to really act out and DARE do what they do. (Hanna)

This seems to be seen as a step on the way to becoming confident enough to share ideas in an interaction in the classroom.

Communication is a priority for all informants. Oral communicative competence is however taught in different ways. To some teachers it seems to imply any kind of exercise involving talking.

> Q12: Well, it can be… You can do these different *one-minute-topics*, where I write something on the board and they get one minute to talk about it. We have done different things, like presenting yourself or talking about your friend. Sometimes they get to act out little plays and then of course sometimes it's reading a text and answering questions. (Robin)

Other informants focus more on the exchange of ideas.

> Q13: We-ell… I often try to have a debate. I place them in groups, I throw out words, I provoke them, because I want to get them going. And then they get to discuss, first in groups and then we summarize it into a thing… (Andrea)

A majority of the teachers described a mixture and a variation of different types of activities to engage students in oral exercises and interaction.

Correctness was mentioned by all informants. To some of the teachers accuracy is, more or less, a focus area, to others its importance is secondary to communication.

> Q14: Of course, you cannot… Some things you cannot get around. You have to drill the irregular verbs into your head, there are things you have to do first, so… so that you have a small tool box that you can use and feel safe with. (Dominique)

> Q15: Then of course when we correct writing, well, then we look very much at sentence structure and you look at the conjugation of verbs and all those parts, you know, all those things, and then it is easily done that you might bring that into the oral part, where it might not have the same weight. (Robin)

To those who find it important that the language used is as correct as possible, there is a tendency to use Swedish to simplify students' understanding of the instruction. One informant stated also using translation of textbook texts as a routine exercise.

The content of the interaction is an issue when the informants describe their teaching. Here is where true communication happens, according to some teachers.

> Q16: I feel you grow a lot from discussions around books /…/. Both as a person and as… Therefore it is fun to be able to bring that [book

> discussions] into the English lessons. /…/ The focus is then not that much on 'now I am speaking English', you know. (Linda)

> Q17: Knowledge of the world is what builds language! /…/ And that's why it's so important to have good materials for teaching. That the materials aren't some sort of "now we're learning English-…", apart from everything else. /…/ So they get a world-…well, an understanding. Language is used out there in the world, isn't it!? (Andrea)

However, some of the participants worried about assessing the opinions of the students.

> Q18: Well, actually I think that is what's most interesting: WHAT you say. Yes, but…/*hesitation*/…/Wait a minute… /…/This could be dangerous, because then I would actually be assessing their… their thoughts! (Jennifer)

To others the immediate interests of the students are in focus for oral activities.

> Q19: You try to connect to different issues that are talked about, of course, you try to find both topical issues or, I mean, /…/ Now with the 6th graders it was Justin Bieber, oh boy! Well, then you go into that a bit and talk about that. (Robin)

> Q20: ...questions about serious things like environmental catastrophes and things like that […] is something they have no opinions about when they are 15-16 years old. (Jennifer)

The analysis of the informants' reflections on oral proficiency and their descriptions of how they develop oral proficiency in the classroom expressed within the two sections above, show that they all work hard to make all students feel confident enough to take part in oral activities and improve their proficiency. However, the informants differ in their orientation to oral proficiency. Based on their statements, roughly three separate focus areas become apparent. One of the teachers reflected on the various approaches like this:

> Q21: Of course you have thought about it and… /*pause*/ also which focus you have in the assessments. It may also be about where you are from, when you got your [teacher] training and a bit, a bit about how you are as a teacher too, what it is you stress. I find it can vary quite a bit, if you are very communicative or if you are more geared towards language accuracy and so on. /…/ All aspects are there [in the documents] and you can have different priorities if you listen to… (Kim)

"All aspects are there" in the steering documents without any clearly defined hierarchy, according to the informant. It then depends on the preferences of the individual teacher, "if you are more geared towards" one orientation or the other.

To summarize: Among the informants of this study there appears to be a difference in focus on content, language or interaction. Some of the teachers in the study stated that the content of the interaction is what motivates students, prompts them to develop their linguistic skills, as well as helps demonstrate their proficiency, therefore the *content*, what is talked about, is at the heart of oral proficiency in year 9.

Other teachers focus on the linguistic tools and state that being able to use the language as appropriately as possible will give the students confidence to use the target language and eventually become proficient communicators. They then see *language* itself as the main aspect of oral proficiency at this stage.

All informants talked about the importance of communication and the majority has a focus on trying to create *interaction* and negotiations of meaning in their classrooms, both to motivate their students and to develop the students' oral proficiency.

6.2 Teachers' Understanding of Policy Documents

Policy documents frame teachers' pedagogical activities. In this study the national curriculum, *Lgr11*, the syllabus for English in the curriculum, the national test and the guidelines for the national tests are referred to as the policy documents defining oral proficiency in English and its assessment.

6.2.1 Informants on Policy Documents

Teachers' perceptions are influenced by the policy documents regulating the school system they work in, by their professional training, as well as personal experiences and beliefs. The documents and how they are interpreted regulate how teachers assess and grade student achievements. To explore the influence of the policy documents on the informants' views on oral proficiency, they were asked to reflect on and compare the former curriculum (*Lpo94*) and the new curriculum (*Lgr11*) and their respective syllabi for English.

Lgr11 was relatively new to the participants when this study was carried out. However, all informants claimed that they had been well informed about

the changes and that they had worked at their individual schools and, in some instances, on a more regional basis with the implementation of the new documents.

> Q22: When *Lgr11* came… We have had a lot of, well, in-service-training and study days and things like that. Yes, everyone has had that, and then we met in groups in the municipality. (Jennifer)

> Q23: It would have been difficult if you had not had the chance to discuss it and kind of dig into it and reflect on it, I think that would have been really… (Linda)

They had also studied the knowledge requirements. One of the informants said that study days and team meetings often were about assessment and grading. Some of the informants said they felt a little *Lgr11*-fatigue and claimed their students felt the same way.

> Q24: Actually, we have had loads of in-service-training. Since this new grading system was introduced, we have… so I feel that that has been run through over and over and been turned inside-out. That has not been done with previous changes during all these years, been given that much time. The students are a bit: NO, we don't want to hear more about it! We know how that circle is to be filled out! (Dominique)

Others were still struggling to convert or "translate" the old goals into the new goals and standards. No one expressed any worries about this, but stated that it might take some time and that they were on their way to master the new documents. Several of the teachers had been involved in compiling new local matrixes, for internal use or to inform students and parents about the core content and the knowledge requirements for English and/or other subjects they teach. All of the informants mentioned matrixes for use on a local or individual level, helping in planning their teaching and assessing.

When asked to describe and compare the previous and the new syllabus for English, most informants spontaneously mentioned changes in the performance standards (now called knowledge requirements). Six informants found no major changes between the two syllabi, whereas three informants suggested that the criteria in *Lgr11* are clearer and easier to communicate to students.

> Q25: Clearer than last time. Easier for the students to understand what's required. I don't remember all the details by heart, but; clear and easy to discuss with the students, in contrast to the former one. I find it easier to talk to the students about how they are doing. (Hanna)

> Q26: Well, overall, I think it's more informative, the new one. (Kari)

The informants were also asked if they found any of the skills described in the syllabus more important than others. Six informants said all skills mentioned in the syllabus are equally important. One informant reflected:

> Q27: …writing and speaking, that they are kind of in the same… that everything is in the same sentence: THAT I find very intriguing! They really make a point of them being equally important… (Eva)

Two other informants stated that writing tends to become the most important skill in school, as in writing there are concrete papers and tests to use as proof. One informant claimed that oral proficiency has become more important, because it is now equal to other skills. However, two out of twelve informants maintained that oral proficiency takes up a minor part of teaching and learning according to the syllabus.

Two informants stated that functional, communicative competence ranks the highest of all skills in the syllabus. A couple of the informants pointed out that the documents can be interpreted in different ways:

> Q28: One has to be able to make oneself understood, always. But no special order [between skills]. No, that's not how I see it. But of course you can read it like the devil reads the bible. (Jennifer)

> Q29: /…/well, it is a bit depending on how you read it… (Kari)

> Q30: /…/There is always a certain amount of subjectivity in it, because of these words of appraisal that are there [in the knowledge requirements]. Well, what is the difference between **relatively** well and **very well**? You can't say, can you!? But the national tests and the example essays are a good guide. But then there is also that forum, you know, to sit down with other teachers and discuss. /…/ No, but I think, you can say or think whatever you want about that, it is always going to be subjective, because there is no other way of doing it. So the only thing you can work on then is to try and reach a general consensus, with everything that implies. /…/Both in your school, in your municipality and on a national level. (Robin)

To summarize: The informants claimed they were well acquainted with the new documents. They did not express any concerns or negative attitudes

towards the syllabus for English in general, but had certain remarks on some of the new criteria, mostly relating to the level of the grading criteria, which will be addressed below (see Section 6.4). The impact of the documents on the perceptions of the teachers seems considerable, judging from how they all use the wording from these documents in their statements. However, the interpretations of the documents, as expressed in reflections on them as well as in reflections on oral proficiency in general, differ partly according to the individual orientations of the teachers. Those teachers who have an interactional orientation to oral proficiency tend to find the new syllabus easier to interpret than the previous one, whereas other informants find little difference between the two. Teachers of all three orientations see their own orientation to oral skills in the documents, as shown in Table 10 below.

Table 10. Oral Proficiency in *Lgr11* according to informants

Orientation	*Lgr11* compared to *Lpo94*	Oral proficiency in *Lgr11* according to informants
Content	No significant difference	To state an opinion, convey a message, thoughts, ideas, feelings and experiences in interaction
Language	No significant difference	"Good" language in interaction, i.e. showing basic linguistic knowledge
Interaction	Clearer objectives and performance standards	Functional communication in oral interaction, less emphasis on formality and correctness

Teachers with a content orientation find no significant differences in the description of oral proficiency between the old curriculum and *Lgr1*. They see a focus on conveying a message and stating a point of view in interaction in the documents. Teachers with a language orientation experience no substantial differences in the definition of oral proficiency in old and new curricula either. They find a focus on basic language skills as a first step to develop good oral proficiency in the documents. Teachers with an interactional orientation claim that the objectives and performance standards are clearer in *Lgr11* and that the functional use of the language is emphasized.

6.2.2 Content, Language and Interaction in Policy Documents

In this section, *Lgr11,* the syllabus for English, as well as the rating criteria from the guidelines (the intended curriculum) will be examined in relation to the three orientations expressed by the informants (the perceived curriculum).

Lgr11 is divided into three main chapters: 1. Fundamental Values and Tasks of the school; 2. Overall Goals and Guidelines; and 3. Syllabi. The initial paragraph of chapter 2 on goals and guidelines states that

> The overall goals set out the norms and values, as well as the knowledge that all pupils should have acquired by the time they leave the compulsory school. The goals specify the orientation of work in the school. (Skolverket, 2011c, p. 14, *official translation*)

In other words, chapter 2 of *Lgr11* is a key to the interpretation of the rest of the document. The chapter is divided into several parts and in three of them, Norms and Values, Knowledge, and Assessment and Grades, guidelines for teaching applicable to this study can be found.

In Norms and Values it is stated that:

> The school should actively and consciously influence and stimulate pupils into embracing the common values of our society, and their expression in practical daily action. (Skolverket, 2011c, p. 14, *offical translation*)

Five goals for the whole school are described under this heading, the first being

> … that each pupil:
>
> can consciously determine and express ethical standpoints based on knowledge of human rights and basic democratic values, as well as personal experiences,… (Skolverket, 2011c, p. 14, *official translation*)

This is followed by five guidelines for the teacher, the first two stating that

> **Teachers should:**
>
> • clarify and discuss with the pupils the basic values of Swedish society and their consequences in terms of individual actions,
>
> • openly communicate and discuss different values, views and problems,… (Skolverket, 2011c, p. 14, bold in original, *official translation*)

It seems clear that every teacher has an obligation to address democratic values and ethical topics, as well as to discuss different views and problems in their instruction.

The next part of chapter 2 is Knowledge. The goals are many, but the second in order of all knowledge goals in the national curriculum is:

> … that each pupil:

> can communicate in English, in both the spoken and written language, …
> (Skolverket, 2011c, p. 15, *official translation*)

This indicates the importance of communicative English in the curriculum.

As can be seen from the above, in chapter 2 of the curriculum it is stressed that students are to be given the opportunity, irrespective of school subject, to develop their ability to openly express values, ideas and personal experiences as well as to communicate and discuss these topics across the curriculum.

The syllabus for English in chapter 3 states the aims and the long-term goals for the subject, then goes on to define the core content and finishes with the knowledge requirements (see Appendix A).

In the long-term goals (or 'abilities to develop') in the first part of the syllabus there is a focus on oral (as well as written) communication and interaction: the students are to develop their ability to express themselves, to communicate in speech, to use strategies to understand and be understood and to adapt their language to the situation and the speaking partner/s. The long-term goals are to give guidance and direction for the subject and are also cited in the guidelines for the national test. Together with the core content of the syllabus, the long term-goals are the basis for the knowledge requirements.[23]

The core content of the syllabus is a new part, as previous syllabi (after 1969) did not specify the content of the subjects, only goals to be reached. The core content describes subject knowledge essential to each academic discipline (see Appendix A; Skolverket, 2011c, pp. 34-35). In the core content section, three listed items address specific language phenomena, such as pronunciation, grammatical structure and sentence structure (Skolverket, 2011c, p. 35).

The knowledge requirements state, in general terms, what learning outcomes are required in order to achieve the different grades (see Appendix A). There is no mention of any specific content or clear linguistic level specified, but the functions of understanding and making oneself understood are stressed.

[23] Commentary on Performance Standards for English, http://www.skolverket.se/publikationer?id=2825

The supplementary assessment factors, found in the guidelines for the national test, are divided into two sections: *content* and *language and expressiveness*, offering guidance in understanding the intended focus of oral proficiency.

There is then support to be found in the documents for the different orientations to oral proficiency held by the informants, as reported above. Some of the statements from the informants also seem to contain references to previous curricula and syllabi. The references to *feeling confident enough to* or 'daring to' as well as to *want to* speak echo the goal for English in *Lgr80* (see Section 2.2.4). Without referring to previous curricula, one of the informants remembers:

> Q31: What did it always use to say before?... about English… let's see…to want to, to dare to, to be able to – that chant, you know. And you mustn't forget, there has to be willingness there, a wish to actually express things /…/ and feelings. That is just as important… (Kim)

However, as *Lgr80* marked the shift to a strong emphasis on student centeredness, it is also possible to view the comments, positive and negative, on *wanting to* and being *confident enough to* speak as voicing a student centered attitude, or arguing against too much of such an approach in favor of a more knowledge centered attitude (see comment from informant Ingrid in Section 6.1.1, Q6).

To summarize: All three identified orientations to oral communication found in the informants' statements, *content*, *language accuracy* and *interaction*, can be viewed as supported by different parts of the policy documents. The parts are not clearly hierarchically ordered, thus leaving room for individual interpretations on how to prioritize between them in teaching and assessing oral proficiency.

6.3 The Oral Part of the National Test for English

In this section, statements about the oral part of the national test will be explored.

The informants all expressed that they were familiar with the national tests for English, had organized and rated them several times and found them to be good tests. The benchmarks were appreciated, though not all of the informants regularly listened to the CD with samples of student interactions

or studied the comments to these samples. Four of the informants stated that they thought the selection of topics and/or the construction of one of the parts of the 2013 oral test was more difficult than previous years.

> Q32: The year 9-test was difficult this year. Their oral was more difficult than it has been previously. The task, they had difficulties starting, many had./…/ I don't know if it was the format, it was… They had difficulties understanding certain words in it too. (Hanna)

> Q33: I thought all of part B [of the oral test] was difficult. Very many of them did not understand what they were supposed to do. /../ And then they got stuck. They don't think independently and freely and they don't talk, they were too caught up in trying to do the task. I tried to get them going, tried to get them to motivate and explain, but then I became too much part of the test. So I thought the B-part, it was not very good, actually. It was too complicated. (Kim)

All four informants who found the 2013 test more difficult claimed that this was exceptional, as the tests normally are received very well by the students, who tend to enjoy the interaction in small groups.

> Q34: I think it is more like it is difficult to make them stop sometimes. 20 minutes, that is nothing! (Andrea)

> Q35: And then I give them lots of time there. Well, I have a time there when it [the test] is officially over, but say that they go on talking, then I don't interrupt [them] /…/ They stay on and keep discussing during their lunch break if they want to. (Dominique)

A great concern shared by all informants is how to group students so they feel confident enough to demonstrate their full proficiency during the test. The informants are well aware of the potential interlocutor effects on the interaction. Issues brought up by informants are: levels of proficiency between students in a test group, how well the students know each other, and how extrovert or introvert a student is. Several informants point out that an extrovert student is needed in each group, to secure the conversational flow. Some informants try to mix different levels of language proficiency to create a more dynamic atmosphere for the conversations, but stress that the variation in proficiency levels cannot be too large.

> Q36: …you suspect that that person could get a bit further, then possibly you would put that person with someone you KNOW is very good. Who might win that person over to the other side. But that's… It's a bit the level too, that I look at… and then it might be that you think that "now that person needs to get a bit of a push in the right direction". /…/ The differences can't be too big either, then it may get too inhibiting for the one who is a bit… weaker. The stronger person usually manages well anyway, more or less. (Eva)

> Q37: It is mostly the grouping that you look at a lot, to make it as good as possible for everyone. Partly regarding security, so that they can perform as well as..., /…/so they don't feel afraid of talking because you are with someone you don't know that well. /…/But also that you think about, if there is someone there who needs someone who, who needs a bit more… who might need someone who starts it off to get going himself. (Hanna)

The concerns about different levels of proficiency are also brought up by informants in the group discussions (see Section 6.4.4).

The role of the teacher as an instructor and a coach, as well as an examiner, and how this may influence the interaction is mentioned and questioned by some informants (see also the informant Kim, Q33, above, this section). In order to encourage the students to speak as much as possible, the teachers try to ask questions or comment when a conversation is dying during the actual test. They might also choose what cards/topics to present to certain groups of students. The teachers are, however, afraid of being too active and at times find it difficult to balance their own involvement.

> Q38: And you… honestly, you try as long as possible, but when the silence becomes unbearable, then, then… then it goes like this: "Have you thought about this?" Or: "If you think like this…" or… Well, yes… So well… Ok, you do interfere a bit more there then, maybe, yes, to get the conversation going, kind of, in… Yes, absolutely. /…/ And then you are also recording and you don't want to be recorded yourself, so you try to keep quiet as long as possible! (Kari)

> Q39: But this year I myself had to be much more involved to push and get them going in the right direction, sort of. And that is also what's so difficult – What am I assessing? (Andrea)

This is a concern discussed in the group discussions as well (see Section 6.4.4.).

All twelve informants in the study normally graded the oral subtest of the students they were teaching, i.e. their own students. In one case only, two teachers were conducting and grading the tests together. Three informants

"helped out" or rated a small random sample of efforts by students they did not teach. Eleven out of twelve informants were the only raters of the oral tests of their own students.

Nearly all of the informants took notes on each student during the tests and several brought matrixes or checklists to the test. Six of the informants stated that they always audio-recorded the tests, others said they sometimes record them. Not all of the informants who audio-record the interactions listened to the recordings afterwards, whereas two informants did this on a regular basis, making a holistic assessment during the test and then listening closer for specific details when listening to the recordings and passing a final grade on the interactions. The other ten informants stated that they too make a holistic assessment at first and then try to pinpoint certain specifics to finalize the grade before the test is over. All informants reported that they have colleagues with whom they discuss test results that they find challenging or problematic to rate. Eleven out of twelve informants stated that they routinely collaborate in the assessment and rating of the written parts of the national tests for English, but not in the assessment of the oral subtest.

Most informants were satisfied with the organization of the oral test at their respective schools, despite a setup where they are organized during regular lessons. The teacher then leaves the classroom with a pair or smaller group of students to carry out and rate the test in a separate room, while the rest of students are left to work on their own. One informant said that because of the traditional setup, where it has been up to the individual teachers to organize, carry out and rate the tests of their own students during regular class-time, the oral test has been seen as less important than the parts of the tests taken on special days assigned by Skolverket, usually organized by the school administration and involving other teachers, etc.

Some informants mentioned that the tests take too much time away from teaching. There were also complaints from several teachers about not having enough time for assessing and discussing the tests with colleagues.

The informants appreciate the tests but are not uncritical. They discuss details that they find less good and most of them have views on rating criteria for the tests in general. Several of the informants commented that the criteria for E were set too low.

> Q40: When you start looking, you think: Oh, is this passed? Well, I am not surprised anymore, but the first times I felt: Oops, I have probably been too severe [when rating]. (Eva)

> Q41: I looked at those from Skolverket that they had passed as a low E-level, and you felt: Cheeze, is this the E-level?! Well, kind of like that. Because I don't know if I would have rated… I think I will have to be a bit self-critical and say that I probably wouldn't have passed it, not in the role I have here. (Robin)

To summarize: All informants had been involved in arranging and rating national tests several times. They organize the tests according to recommendations in the guidelines and find them very good. However, eleven out of twelve teachers were the only raters of their own students' efforts and few discussed their rating of oral proficiency on a regular basis with their colleagues, even though they discuss the ratings of the other parts of the national test regularly. As only six of the informants audio-record the tests, a second rating or opinion is made impossible in many cases.

6.4 Assessment of Oral Proficiency

In the previous sections the data from the interviews have revealed how the informants describe oral proficiency, what grading criteria they find capture this proficiency and how they describe their instruction aimed at developing their students' oral proficiency. An analysis of the results from these three areas indicates that they seem to have different orientations to oral proficiency, focusing mainly on *content*, *language* or *interaction*. The informants have further expressed that they are familiar with the policy documents, which is also shown in their adherence to the actual wording in the documents in their own statements. They have also stated that they are well acquainted with the national tests, approve of them and have organized them regularly. In the following section, their statements on assessment of oral proficiency, as well as their actual rating of oral proficiency in the tests, will be studied.

6.4.1 Informants on Assessment and Rating of Oral Proficiency

The interviews revealed differences in the perception of the assessment of oral proficiency among the informants. One of them stated that s/he gives weaker students more leeway in the oral tests in order not to discourage them, i.e. silence them. S/he said that s/he is very generous when awarding grades for oral proficiency. Two further informants stated that at their school they

pass students who speak very little, in a way they would never pass students who write very few and short texts. One of them reflected:

> Q42: Well, the oral part more than other parts of teaching, it's as if you think of it being so much more a part of your personality that when assessing it becomes difficult. When you coax a student into writing that's not a problem, but if you try to make a student talk [who doesn't want to], that is a problem. And if you assess a student and say: You don't reach the goals for writing, then there is a sort of distance between the person and what he writes, but with speaking it's as if they cling together [personality and performance] and then, then at times maybe you go a little easier there… (Kim)

The wish to not pass judgment on personality nor discourage a student, seem to be possible influences for generosity in grading a student's oral performance. One informant, on the other hand, seemingly finds the demands for the highest grade too high for most students to reach as s/he stated s/he had never awarded any student an A.

When comparing the new curriculum with the previous, several informants commented on the higher demands for the highest grade.

> Q43: Yes, that was one of the first things I noted, that it feels as if it is required of the student… That a more competent student is taken for granted, one who can discuss several subject areas and…/…/ I think it feels as if they have to be able to discuss, well, about living conditions, really, and it says ethical dilemmas and such./…/ I think it is more of that. (Kim)

As seen above (see Section 6.3) yet others find the requirements for the lowest grade E as exemplified in the national tests to be too low.

Several participants stated that they, with more experience, feel more confident in their grading and assessment.

> Q44: Without a doubt! I value [things] in a different way. /…/ I was really rather single-minded, totally square-headed. It had to do with not having enough experience and relying on what was written. (Ingrid)

> Q45: Well, I am probably more careful now, thinking of all the different parts, maybe. I have become better at trying to, for myself, setting up matrixes and kind of… eehh… /*pause*/ more mark and think a bit about the different parts, not simply seeing it as speaking, you know? I think I did that more before. It was kind of like a chunk. Now I see it more as several parts. (Kari)

The teachers in the study all stated that it is not easy to assess oral proficiency but, with experience, their confidence has grown. Discussions with colleagues are seen as important and they would like more discussions, in their own school, but also with teachers from other schools. However, all informants stated that they have colleagues to ask for advice, if in doubt about the grading of a particular student performance.

> Q46: You can always come in and… you can really show: "I am very uncertain here. How should I be thinking?" Or something like that. That's not strange, you know. (Linda)

Several of the informants claimed that the most difficult part is verbalizing the grounds for their assessments and their grading, making them transparent, especially to students and parents.

> Q47: I can't say that it is very well grounded in a rubric of some kind, but it is based on 20 years of experience. (Jennifer)

> Q48: Well, I know, I know approximately what it is supposed to look like. Even the new [syllabus] now… I felt… well, it hasn't been that difficult to integrate, maybe a bit of the differences in levels, then. /…/BUT /…/: How am I going to get this across to the students? (Eva)

The informants agreed that it is fairly easy to decide if a student shows enough proficiency to be passed with the lowest grade, E, but that it is more difficult to assess the higher grades. The margins and differences between the levels are shady and difficult to pinpoint, according to the informants.

Three informants talked about more or less constantly carrying out on-going assessment of the oral proficiency of their students, as so much in the lessons is done orally. Two informants jot down notes after virtually every lesson on as many students as possible and strive to give feedback, even if only very briefly, during class, in the corridor or even in the lunch room as often as they can. Individual, personal feedback on oral efforts is considered of vital importance by a majority of the informants and they try to give this feedback in face-to-face meetings regularly.

The value of arranging frequent occasions for the students to show their ability to interact orally in English was mentioned by several informants, but it was also pointed out by one informant, that if students do not speak English in the classroom during lessons, it will be difficult to arrange enough opportunities for those students to properly demonstrate their oral

communicative proficiency. Normal authentic day-to-day interaction in English is the basis for fair grading and assessment, according to several informants. Often interactive situations similar to the test situation in the national test are arranged during the school year to prepare students for this kind of peer-to-peer talk. A majority of the informants state that the national test is merely a confirmation of the levels of proficiency of their students, which they have already estimated.

In summary: Oral proficiency is a challenging skill to assess and grade, according to the informants. Some claim it is difficult to make the assessment criteria clear to students. Others confess they themselves find it difficult to see the distinction between levels of grading criteria. Many of the informants take special care to arrange situations for students to show their proficiency and a majority is anxious to feed back and feed forward, to encourage their students to develop their oral proficiency. A majority of the informants see the oral subtest as a confirmation of their own on-going classroom assessment of their students' oral proficiency.

6.4.2 Features of Oral Proficiency Noted

To investigate how they go about grading student efforts and how they talk about their grading with colleagues, the informants were invited to a grading and group discussion activity. The analysis of arguments put forward by the participants in the group discussions reveals what aspects of oral proficiency were primarily observed and subsequently influenced the ratings. When their comments on the student performances were coded according to the supplementary assessment factors in the guidelines for the national test, it became clear that the informants focus less on features that can be ascribed to the 'Content'-factors (about a fourth of the comments) and more on features that can be categorized as 'Language and expressiveness'-factors (see Section 5.4.3, Table 6 and Appendix B) in their discussions. This is illustrated in Figure 8 below.

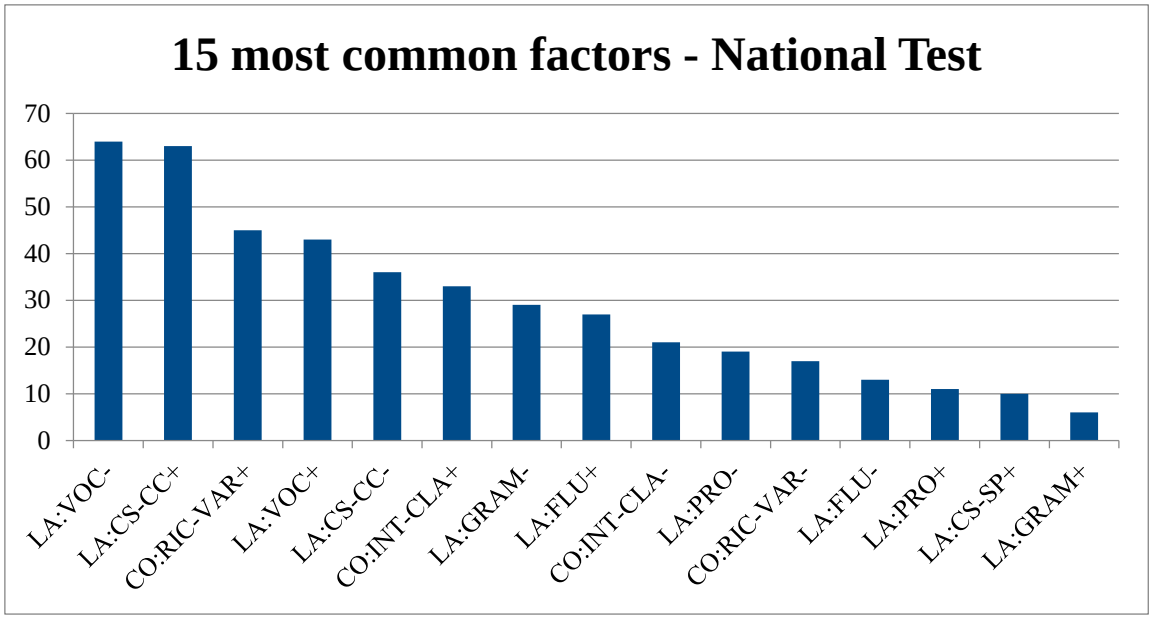


Figure 8. Top 15 assessment factors classified using national test guidelines

Of the 15 most frequently mentioned factors, as shown in the figure above, 11 belong to the 'Language and expressiveness'-category of codes. Among these,

- slightly more than half (161) of the language-related occurrences were negative (indicated by a – in Figure 8 above), suggesting a linguistic weakness,
- as opposed to the (150) language-related occurrences, which were positive (indicated by a + in the figure above).

The most common observation was a *deficiency in vocabulary*. The second most common remarks were positive comments related to one of the communicative strategies categories, i.e. being *good at developing and carrying on a conversation*. The third most noticed feature was *richness and variation of content* and the fourth a *rich vocabulary*. The fifth most noticed feature was a *weakness in developing and carrying on a conversation*. Comments on interaction (LA:CS-CC+ and LA:CS-CC-) make over a third of the total comments in the 'Language and expressiveness'-category as shown in Figure 8 above.

Since interaction was stressed by a majority of the informants during the interviews and since interaction was also frequently mentioned in the discussions, an attempt to capture what particular aspects of interaction the teachers observed was made using a coding key inspired by May (2011), (see Section 5.4.3). The aspects most commonly noted are shown in the figure below.

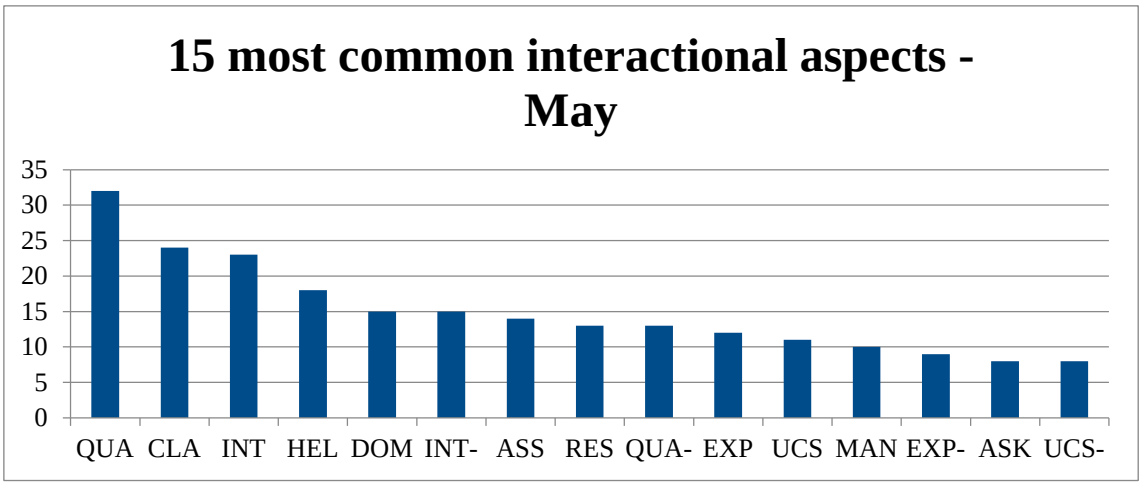


Figure 9. Top 15 interactional aspects noted - matrix inspired by May (2011)

In Figure 9, the richness of different aspects connected with interactional efficiency is clearly visible in the 15 most frequently commented interactional features (see Coding Key Section 5.4.3).

- QUA indicating general good *quality of interaction* is the most frequently used and least specified feature.
- CLA indicating *asks for clarification, confirmation or clarifies own ideas, concept checks* is the second most commonly mentioned feature.
- INT is the third, meaning that what was said is *intelligible*.
- HEL is the fourth, indicating that the speaker *helps and supports* his/her partner in the interaction.
- DOM is the fifth, indicating that one party *dominates* the discussion, usually in a negative way.

The 15 interactional features shown in Figure 9 are not diversified in the supplementary assessment factors used for the rating of the oral subtest, but are summarized mainly in one category, 'Language and expressiveness: communicative strategies' (see Appendix B).

The informants discussed the different linguistic and interactional features of the student interactions in their groups. In the end they compared their holistic final grades, and discussed aspects and assessment factors as the grounds for the grades they had awarded the different performances.

To summarize: The remarks made about the student examples during the group discussions were coded according to two different coding keys to investigate what features were most salient to the informants. The result from the coding, in accordance with the supplementary assessment factors, showed

a majority of remarks on language (cf. Borger, 2014). A closer scrutiny, however, revealed that two of the most frequently mentioned features were interactional features: the use of communicative strategies, and the lack thereof. When using the coding scheme inspired by May (2011), the results showed a rich variety of interactional features noticed by the informants. The second coding scheme makes it obvious that the informants were aware of interactional efficiency dimensions that they had not explicitly derived from the assessment factors (cf. Figure 8 and Figure 9). A number of features in the two keys were rarely or not at all mentioned. They have not been included in this investigation.

6.4.3 Informants Grading Oral Proficiency

The actual grading of the student examples was not observed or recorded. After the individual assessment and grading, the informants in their groups exchanged views on how they had graded the student examples and what criteria or assessment factors the performances fulfilled. (This exchange was recorded and transcribed.) Although the participants were informed that the purpose was not for them to reach an agreement on grades awarded, but merely to exchange thoughts and reflections on their individual ratings, both groups were eager to come to an agreement on how to grade the individual student efforts. The examples of peer-to-peer oral interaction that the informants listened to were graded differently by the two groups. The analysis revealed a clear difference between the two groups with regard to which criteria from the national test guidelines they mostly referred to. The features most commonly referred to were fitted into a grid and a variation became apparent, as shown in Table 11 (for student example 1) and Table 12 (for student example 2) below.

Table 11. Assessment Factors – Student Example 1

Most frequently mentioned features of oral proficiency in assessment factors in National Test guidelines, student example 1. *Positive and negative observations are respectively indicated with (+) and (-).*		
	Group I, ex. 1	**Group II, ex. 1**
Features most frequently commented on coded acc. to NT matrix	Content (+): Richness and variation – Examples and perspectives (CO:RIC-VAR+)	Language and expressiveness (+): Communicative strategies – Develop and carry on conversation (LA:CS-CC+)
	Language and expressiveness (+): Communicative strategies – Develop and carry on conversation (LA:CS-CC+)	Language and expressiveness (-): Richness, variation, clarity and assertiveness – Vocabulary, phraseology and idioms (LA:VOC-)
	Language and expressiveness (-): Richness, variation, clarity and assertiveness – Vocabulary, phraseology and idioms (LA:VOC-)	Content (+): Intelligibility and clarity (CO:INT-CLA+)
	Language and expressiveness (+): Richness, variation, clarity and assertiveness – Vocabulary, phraseology and idioms (LA:VOC+)	Content (+): Richness and variation – Examples and perspectives (CO:RIC-VAR+)
	Language and expressiveness (+): Fluency and ease (LA:FLU+)	Language and expressiveness (-): Communicative strategies – Develop and carry on conversation (LA:CS-CC-)

Both groups commented on the richness of content and the use of communicative strategies in student example 1, as can be seen in Table 11 above. The deficiency in vocabulary was also noted by both groups. Additionally, group I commented on positive aspects of vocabulary and on fluency, whereas group II noted the communicative strategies more than other features, both in a positive and negative sense. More comments were made on interactional than on linguistic features in group II, as the informants elaborated on the intelligibility of the communication, which they considered adequate even if the language used was sometimes rudimentary. The informants in this group also emphasized the helpfulness of the more proficient interlocutor, who kept the interaction going, by returning to comment on this several times in the discussion.

The varied foci in assessment in these two groups ended in a difference in actual grading. Group I awarded an E and a C for the two speakers

respectively, whereas group II rated the two students E/D and A. The focus on interaction and communicative strategies rendered one of the students a higher grade from group II.

In the discussions on the second example of student interaction, there were similar differences in emphasis among the informants, as in the previous example. Group I again had a stronger linguistic attention than group II, as shown in Table 12 below.

Table 12. Assessment Factors – Student Example 2.

Frequency of Features Salient to Groups Rating Example 2 According to Assessment Factors from Guidelines for National Tests. *Positive and negative observations are respectively indicated with (+) and (-).*		
	Group I, ex. 2	**Group II, ex. 2**
Features most frequently commented on coded acc. to NT matrix	Language and expressiveness (-): Richness, variation, clarity and assertiveness – Vocabulary, phraseology and idioms (LA:VOC-)	Language and expressiveness (-): Communicative strategies – Develop and carry on conversation (LA:CS-CC-)
	Language and expressiveness (-): Richness, variation, clarity and assertiveness – Grammatical structure (LA:GRAM-)	Language and expressiveness (+): Communicative strategies – Develop and carry on conversation (LA:CS-CC+)
	Content (-): Richness and variation – Examples and perspectives (CO:RIC-VAR-)	Content (-): Intelligibility and clarity (CO:INT-CLA-)
	Language and expressiveness (+): Richness, variation, clarity and assertiveness – Vocabulary, phraseology and idioms (LA:VOC+)	Language and expressiveness (+): Richness, variation, clarity and assertiveness – Vocabulary, phraseology and idioms (LA:VOC+)
	Content (+): Richness and variation – Examples and perspectives (CO:RIC-VAR+)	Language and expressiveness (-): Richness, variation, clarity and assertiveness – Vocabulary, phraseology and idioms (LA:VOC-)

From the five most frequently mentioned features shown in Table 12 above, it is clear that the discussions were mostly centered round deficiencies in the performances by the students in example 2 (indicated with (-) in Table 12). A positive aspect mentioned in both groups was vocabulary. Group I predominantly discussed the language used, with a focus on grammatical mistakes and the lack of variation, whereas for group II the two most

frequently commented factors were the 'Language and expressiveness' factors concerned with communicative strategies. Group II also in this example concentrated on the interaction and commented on one of the interlocutors dominating the other and preventing him/her from demonstrating his/her proficiency.Group I further commented on slurred or unclear speech due to a non-Swedish accent in this example. They referred grammatical problems (LA:GRAM-) of the students to their non-Swedish background. The informants in group II also noted occasional difficulties in understanding what the students in example 2 were saying, but did not refer the unclear speech to any specific assessment factor, but found it had a negative impact on several factors. They therefore noted that the content was un-intelligible and un-clear (CO:INT-CLA-), for instance. No comments on a non-Swedish accent were made.

For student example 2, the factors mentioned and emphasized differed to some extent, showing different approaches in the two groups. In their rating, however, the groups reached the same grades, E and D, despite the variation in what aspects they commented on.

To summarize: All informants in the two groups dwelled to some extent on interactional efficiency. Although the supplementary assessment factors and knowledge requirements include only one or two obvious interactional language features, group II put more emphasis on interaction than group I. Group II also noted a rich variety of interactional features not explicitly mentioned in the policy documents. The aforementioned difference in orientation to oral proficiency was again discernible and here resulted in a difference in rating and, in one case, even a difference in grading. In group I there was a representation of all three orientations among the participants, whereas in group II no participant had a language orientation, seemingly influencing the group's collective perception of the assessment factors and of the student performances.

6.4.4 Informants on Rating

In addition to features relating to oral proficiency, the group discussions also covered the supplementary assessment factors and knowledge requirements per se, which some participants considered to be rather unclear or confusing.

Q49:

- /…/Are we looking at these assessment factors that are listed or are we looking at the knowledge requirements? Because if I look at the assessment factors I would almost rate her a little harsher, but if I look at the knowledge requirements, well, then it is "relatively varied", "relatively coherent", kind of, it's just… I mean if I look at the knowledge requirements I think she's on a C, clearly C on everything. She doesn't dip anywhere.
- But then we agree…
- Yes, we agree that it's a C.
- Right.
- That feels really good!
- But if you look at…
- Otherwise we would have to pity our students!
- /*ironically*/"But I really think she should have an A…"
- If you look at the assessment factors that's where she is lacking. It's in richness, variation and clarity.
- Yes, absolutely!
- But not in assertiveness.
- No.
- No, I don't think so either.
- Nor do I. (Group I, student example 1)

The participants are uncertain about how to use the supplementary assessment factors and the knowledge requirements, and about how they are related to one another, but seem to be able to find ways to assess the student effort in spite of this.

The raters further compare between student efforts, instead of comparing them with the assessment factors and performance standards.

Q50:

- Then of course you get influenced just as we just said, you compare with the other [speaker]…
- Mmm
- And clearly her language is much more advanced than what he is, but…(Group A, student example 1)

Q51:

- You want to compare between them. It's only human! (Group I, student example 2)

The matching of proficiency levels and personal characteristics of test takers are of great concern to most of the informants.

Q52:

- We are probably in full agreement that the boy is weaker than the girl?
- Yes, absolutely!
- And the difference is very big.
- Mmm
- /…/ I am a bit confused here, what were they [who matched the students] thinking?
- Yes, I wouldn't have done that, because both of them lose, I think, right?
- Mmm
- The stronger student loses because she gets no opposition
- Yes, and no response…
- Yes
- …that can make her move on. And the weaker [student] probably feels…
- With her fluency!
- He feels bad. /…/Not only linguistically but also… it's also… How can I express this? Intellectually. (Group II, example 1)

The perceived mismatch leads to speculation about what the student could possibly have achieved had s/he been matched with another partner.

Q53:

- But the question is, well that's how I felt, if she had been speaking with someone else, maybe she could have shown a little more [of her proficiency], but I don't know… (Group I, example 1)

Q54:

- /…/even if her vocabulary could have been richer…
- And it surely had, had she been stimulated by someone else. That's the way it is. (Group II, example 1)

Both groups react strongly to the teacher interfering in the peer-to-peer action.

Q55:

- Well, what bothers me here is that she was very active, the teacher.
- Yes, exactly!
- Because in the beginning they… well they had… they talked to one another, but then she came in and the more questions she asks the quieter they become towards one another. That was really a pity! (Group I, example 2)

Q56:

- She interfered so much, the teacher here, that there was no flow [in the conversation]. (Group II, example 2)

In interviews prior to, as well as after the group discussions, the interference of a teacher/instructor in the oral part of the test was commented on by most informants. They all found it a problem when testing weaker or quieter students and here commented negatively on this type of interference again (see Section 6.3).

To summarize: Several factors, other than the obvious interactional and linguistic features, are noticed by the informants. Uncertainty about how to use knowledge requirements and supplementary assessment factors add to the difficulty of rating the oral test. Taking aspects like the perceived impact of a speaking partner and the teacher into consideration further complicates the matter. However, after addressing the issues and discussing them in the groups, the informants arrive at rating decisions shared by the group members.

6.4.5 Informants Reflecting on Assessing Oral Proficiency

The informants' reflections after the group discussions were all very positive. They claimed to feel more confident in their assessment and grading after having discussed their ratings in the groups, and found it rewarding to have exchanged ideas and opinions with colleagues from other schools. Several informants expressed that they find this kind of activity an excellent kind of professional development.

No questions were asked by the researcher on reasons for the variation in rating during group discussion, but most of the informants spontaneously reflected on possible reasons for the differences. One informant brought up a debate that had been going on at her school and that she felt was partly echoed in the group discussion:

> Q57: Of course you have to think of the grammar, that it's there, and that the correctness is there, but that is not really the focus here… in the material you get [together with the national tests] it's much more about communicating and conveying a clear message and such and that can be done even if you're not very good at… at the formalities. /…/ The dangerous thing for me as a teacher of Swedish is that I cannot look too much at, well, certain parts: "Listen, you haven't backed all your arguments here now!" Well, you know… (Eva)

She further explained that the debaters were teachers of English and a third language, and teachers of English and Swedish. This informant found that teachers of foreign languages were more prone to focus on linguistic accuracy and teachers of Swedish tended to focus more on interaction and content. Another informant commented that s/he felt they (the informants) were generally in agreement during the group discussions, but when there were differences of opinion, s/he felt they were mainly due to age and experience. Teachers with more experience have higher demands on linguistic variation to pass a higher grade, according to this teacher.

One informant stated that the group discussions really showed how complex the matter of assessing oral interaction is. To listen to other teachers' reasoning on how and what they rate was said to be interesting and edifying.

> Q58: And there [in the guidelines for assessment of the oral part of the national test] they rather stress the interaction bit, /…/not that much emphasis on specific choices of words or sometimes possible grammar mistakes or formal mistakes or what you call them. Rather, there is much more in this about how you take part in a conversation, that you are an active [speaking] partner, ask questions, follow-up and those kinds of things. And there, there you could see that we [the participants in the group discussions] sometimes looked for different things, where some perhaps think more about: "How do students pronounce the words? Do they make mistakes on certain verbs?", or whatever. The difficult thing is maybe that you sometimes put on your writing glasses, even though it is the oral part. (Robin)

To summarize: The informants stress that they find discussions with colleagues a sound and rewarding kind of professional development in assessment. They, further, voluntarily commented on the differences they experienced in the views on assessment in the group discussions as stemming from differences in approaches to language, depending on their subject specialties and/or the length of their experiences.

6.5 Frame Factors

To further investigate what may influence the informants' assessment of oral proficiency, factors established as traditionally framing educational processes were explored, i.e. *time*, *resources* and *student body*. The informants were found to include all of these in their statements.

Time is a factor mentioned in all interviews and group discussions. Time is a crucial aspect in the teaching process and in assessment practice. For instance, in this study half of the informants state that they lack time to record and listen a second time to the student interactions of the oral subtest. In the following quote, a teacher mentions several areas where s/he feels more time is needed for the oral test:

> Q59: …And that we get better conditions to carry out the tests. And better conditions for rating discussions. And to listen… [a second time to the recordings]. (Chris)

All but two maintain they do not have enough time for rating discussions with colleagues in connection with the national tests.

> Q60: I just talked to a colleague and we said that we would try to get time for it [discussion on rating of national tests], but it doesn't… there is no… We won't get the time. Unfortunately. We did get time [organized by the school administration] last year, we sat down with the 6th graders' [test] a couple of us, but it's… it's really, really tough. So… we had… we really want to!! But… (Chris)

Some of the informants also point out that discussions on assessment in general are needed to calibrate the assessment among teachers at a school, as well as between schools, and that there is rarely time for this type of discussion, which in turn affects the rating of student achievement. One of the informants finds that discussions on the interpretation of the policy documents and the knowledge requirements are needed in order to come to some sort of consensus (see quotation by Robin, Q30, Section 6.2.1).

Material *resources* are mostly mentioned in positive terms, as the teachers talked about the availability of technical devices and digital equipment making audio-recording much easier than previously. There were, however, also limiting aspects, when resources were connected to time and substitute teachers.

> Q61: /…/I know that there, when they have oral tests, then they are always two teachers listening. /*pause*/ I believe that is good, but here we don't have those resources. (Andrea)

The *student body* is another factor explicitly brought up by the informants. Most of them comment on the general proficiency level of their students and how the student body of their school tends to influence their interpretation of the different levels of proficiency.

> Q62: Well, I think we sort of landed in that we, sadly enough, are guided by the kind of students you have. In a way you end up in an assessment mode up or down depending on the student body you have. /…/ and you cannot be sure that is the normal mode. /…/If I have a couple of classes where no students are very strong, then my level will drop, and vice versa rise. Which I imagine happens between schools and also within schools if we don't have the possibility to meet and talk and meet one another's student. (Andrea)

> Q63:
>
> - What I mean and what I was sort of into was the assessment, if it is correct. Had it been different if you had had classes with noticeably weaker… ? Because then it [the assessment] might have… /…/
>
> - Well, as we said before, you are very influenced by your own students, that's what's interesting.
>
> - Yes, we are and that's very important to keep in mind, so that you don't get carried away. (Group I, student example 1)

Another factor could be discerned in what the teachers chose to comment on. A number of the informants stated that the traditions and the culture of the *community of teachers* they are part of at their schools influence them. A local culture among the teachers can contribute to establishing standards for that particular teacher community.

> Q64: That is everyone [all English teachers] here has loads of experience. And we have set our own standards here, what we think and so on. And it's always like that at every work place, isn't it? /…/Even if there are performance standards and you have… well, well, you create your own rating scales anyway, at every… I think. Here we are quite meticulous, I think. That is a culture you internalize. Ehh… I think. (Dominique)

Q65:

- Because schools, I feel schools develop a rating culture…
- Exactly!
- A school develops an assessment culture and then the teachers fall in.
- Yes. You have a…
- You get dragged into it and you have…
- Hmm… The strong ones… That worries me now… (Group I, student example 1)

A community of teachers can thus form a strong local school culture that influences teaching approach and assessment style for all teachers of that teacher community, according to the informants. This is discussed by the informants both as a supportive practice and as a limiting practice.

To summarize: It seems that lack of *time* leads to a more uncertain grading practice and less collaboration on rating oral achievement. *Resources* are mainly talked about in terms of digital and technological equipment, supporting the execution of the oral subtest, since recording equipment is readily available in all schools. However, the teachers seem to feel that they do not have the opportunity to make full use of these resources in connection with the oral tests, because they lack the time to do so. Awareness of differences in *student body* creates a further uncertainty and anxiousness in grading, resulting in a wish for more collaboration within schools as well as between schools. The informants also recognize the potential limitation, as well as support of, a strong *community of colleagues* at a local work place, which further underlines their wish to collaborate and discuss student assessment inside as well as outside their own schools. *Time* available, local *resources,* their own *student body* and the *community of teachers* they are part of all seem to be factors perceived as framing the assessment processes for the informants in this study.

6.6 Summary

The analysis of the statements made by the informants reveals that three different orientations to oral proficiency can be distinguished. One orientation is focused on the *content* of the interaction between speaking partners, one emphasizes *language accuracy* to help and support the interaction, and one is primarily focused on the *interaction* as such, stressing motivation and techniques to keep a conversation going. Informants who emphasize *content*

view the ability to exchange and negotiate meanings and ideas intelligibly with interlocutors as a primary characteristic of oral proficiency. This is seen as meaningful and motivational interaction. Teachers with a *language* focus see knowledge about the structure of language and the ability to use the basics of this knowledge as a precondition for clear and efficient communication, which in turn is believed to make the students confident in their oral interaction. To the third group of informants, the overall objective is communication and social interaction. These teachers encourage the use of the language at hand in the most efficient way in social activities and conversation. The three perspectives, *content*, *language* and *interaction*, are to be found in the documents and criteria. As there is no obvious hierarchy between them, either apparent or declared in the information, different individual interpretations are possible.

The policy documents seem to have a considerable impact on the perceptions of the informants. Implementation activities had just been carried out previous to the interviews and all informants stated they had worked more with the new curriculum than with the previous one. This is further indicated by the manner in which the teachers in this study use the same wording as that of the documents (see Section 6.1.1). National tests are here seen as policy documents. The general attitude to the national test, and the oral subtest in particular, is positive among the teachers in the study and they follow the guidelines and instructions regarding the organization and grading of the tests to the best of their understanding.

When it comes to grading the national tests, all informants stated they have colleagues with whom they can discuss, if they have difficulties rating a student effort and want a second opinion or advice. However, only half of the informants audio record the oral tests on a regular basis. Not recording the tests makes it difficult to examine students' oral achievements in rating discussions with colleagues. Eleven out of twelve participants were the sole raters of the performandes of their own students. Not all informants had studied the benchmarks in the guidelines for the test before testing their students' oral proficiency. They further stated that they do not do this regularly.

In the assessment activity and group discussions, the grading differed. None of the informants disagreed with the rating criteria, but there were still different interpretations of how to apply the supplementary assessment factors and the knowledge requirements to the student performances. The differences were found to roughly correspond with the individual orientations

of the group members. In other words, group I had a focus on language and content, whereas group II had a stronger focus on communication and interaction.

The teachers in this study reported they felt confident rating their students' performances, but stated they had not had any assessment training. Their confidence in rating student achievement has grown as they have become more experienced as teachers and assessors. All of the participants talked about collaboration with other teachers and colleagues when in doubt about assessment and grading. They mentioned discussions and rater meetings among English teachers and wish that there would be time for more collaboration. Some of them worry about not meeting teachers from other schools and other school districts.

Factors mentioned by informants as framing their practice are *time*, *resources*, *composition of learner groups* and *community of teachers*. Lack of time is mentioned as limiting the audio-recording of oral tests, listening a second time, and also as a limiting factor with respect to rating meetings and time to reflect on student performance. Most of the teachers in the study would prefer to have time for more discussions with colleagues, both at their own schools but also with teachers from other schools. A majority of the informants talked about the proficiency level of their students and worry about setting their own performance standards according to their own group of students. Some of them openly refer to a strong community among teachers at their local school, acknowledging that this local culture affects their own perception of teaching and assessing.

Chapter 7: Discussion and Conclusion

The following chapter reviews the main findings and presents some didactical implications of the results. It then makes suggestions for further research and, finally, provides some concluding reflections.

7.1 Main Findings

The general focus on education and student results has intensified discussions on grades and grading practices in Swedish schools. Grades are seen not only as a record of the proficiency and knowledge of the individual student, but are also a means to measure quality in education. Efforts to reverse negative trends in education and to strengthen pedagogical development have resulted in a new national curriculum, *Lgr11*, a new grading scale and more national tests. These changes are aimed at strengthening the focus on subject matter knowledge (Prop./Government Bill 2008/09:87, *Skollag*/The Education Act SFS 2010:800) and further securing fair and equal grades.

There is a stress on English as an important school subject in the national curriculum. Communicative competence is in focus and oral proficiency, together with written proficiency, is emphasized. Oral communication is further seen as the most important English language skill to learn, according to both teachers and students (Erickson, 2010).

The competence and ability of Swedish teachers to award fair grades have been questioned (Skolinspektionen, 2013). Although the re-rating of national tests for English showed the least deviation between original ratings and re-ratings of the different subjects tested, Skolinspektionen deemed the differences to be unacceptably large (Ibid.). Oral proficiency was not included in the investigation by Skolinspektionen. This study attempts to explore how twelve skilled teachers of English perceive oral proficiency and how they grade the oral part in the obligatory tests. The aim is to make teachers' implicit assumptions and perceptions more explicit and accessible.

7.1.1 Informants' Perception of Oral Proficiency

The informants in this study perceive oral proficiency in various ways. Their different understandings of the phenomenon can be grouped into three different orientations, stressing *content*, *language* (linguistic features) or communicative *interaction* respectively, a pattern similar to that found in other studies (e.g. Apelgren, 2013).

The teachers with a *content*-orientation to oral proficiency focus on the negotiation of views and ideas, echoing language learning theories stating that more complex language will result from discussing more complex and engaging topics, and ideas stating that tasks demanding more elaborate output will generate language learning (Swain, 2000; Swain & Lapkin, 1995). They seem to believe, that by pushing the learners to process language more deeply, in interaction with partners, linguistic knowledge will be constructed by the learners, just as maintained by some researchers (Swain, 2000, p. 97).

Informants holding a *language*-orientation appear to rely on traditions, as discussed in other studies (Apelgren, 2001). Basic linguistic tools are seen as the starting point for oral communication, but also as a means to instill confidence, making the students feel confident enough to take part in oral interaction in the classroom.

The teachers showing an *interactional* orientation seem to be inspired by ideas on how comprehensible input and the negotiation of meaning in oral exchange leads to linguistic development (Long, 1985). These informants appear to believe, in accordance with Long, that modifying your speech to be understood by an interlocutor and working to understand information from a speaking partner, i.e. cooperating to reach mutual understanding, is a stage in language learning (Long, 1983). The teachers with an interactional orientation in this study also often stress the use of familiar topics to enhance interaction (Skehan & Foster, 1997).

Depending on personal beliefs and experiences, teachers in this study tend to favor either form or function as the prerequisite for the development of oral proficiency. Their respective perceptions of oral proficiency seem to be influenced by their individual pedagogical philosophies, resulting in their different orientations.

Regardless of orientation, a majority of the teachers find their students to be very good at speaking English, which tallies with national and international reports (Erickson, 2010; Skolverket, 2012b). Still, several of the informants are

concerned that not all students are confident enough to engage in conversations and discussions in English during lessons. Virtually all of the informants elaborate on learners' *willingness* to speak English, mostly in connection with classroom interaction. *Willingness*, or *wanting to* speak English was, however, also mentioned in association with assessment, which can be seen as an echo of old curricula (*Lgr80*, see Section 2.2.4) and a confusion in the transition from previous to new objectives and rating criteria (see Appendix L), but also as a sign of teachers being concerned about the social consequences of rating, of teachers rewarding, motivating and encouraging their students (Brookhart, 1994; Oscarson & Apelgren, 2011).

7.1.2 Informants' Interpretation of Policy Documents

Unlike recent reports (Skolverket, 2013), the statements in the interviews indicate that the informants of this study feel well acquainted with, and knowledgeable about, current policy documents. They state that they have participated in local efforts in their schools – some even in collaborative efforts between schools - to adapt their practice to new standards and objectives. This suggests that measures have been taken on a local as well as regional level to implement the new documents. This is contrary to what was reported about the previous shift in curriculum and grading system in 1994 (i.e. Riksrevisonen, 2004; Selghed, 2004; Tholin, 2006). The informants indicate that they feel reasonably confident using the documents, implying that they are working on incorporating them into their pedagogical practice and that they seem to find the new goals and objectives adequate.

The teachers with a *content*-orientation, emphasizing the expression of opinions and experiences, find support for their focus in parts of the curriculum and the syllabus mandating discussions and exchange of thoughts on topical issues, ethical dilemmas, etc. (see Section 6.2.2 and Appendix A). There is also backing for this orientation in the supplementary assessment factors, which devote one section to aspects of content.

Informants holding a *language*-orientation appear to rely on parts of the core content of the syllabus describing linguistic features to be treated in the instruction of English (see Appendix A, Core content), as well as the supplementary assessment factors, where a variety of language features are highlighted in the Language and expressiveness-section. This focus on language accuracy is also similar to what was stated in earlier syllabi, where

accuracy was seen as an essential tool and prerequisite for interaction (see Section 2.2.4).

The teachers showing an *interactional* orientation seem to rely on the descriptions of the exchange of views and opinions in the curriculum and syllabus (i.e. the same as the teachers with a content orientation), but also on the importance of language as a tool for learning in social interaction (Skolverket, 2011c, p. 32; Vygotskij & Kozulin, 1986) as stated in Content of Communication in the syllabus (Appendix A). There are also supplementary assessment factors addressing communication and communicative strategies, to support this orientation.

The informants construct various understandings of the documents, according to their individual pedagogical beliefs, experiences and orientations, but stay within the limits of what is mandated in those policy documents.

7.1.3 Informants on the National Tests

The teachers in the study report that they are well acquainted with, and positive to, the national tests. They follow instructions and guidelines and, having organized them regularly, they state that they feel confident about how to manage and grade the tests. Some informants found that the 2013 oral subtest was more difficult than previous tests, which might be due to the fact that not all of the informants had studied the benchmarks or listened to the CD with examples in connection with the 2013 test. This is noteworthy, as the 2013 test was partly altered, and the grading criteria, as well as the grades, were new.

That testing and assessing oral proficiency is not an easy task is acknowledged by the informants. They are aware of aspects mentioned in other studies, such as the influence among interlocutors in general in the peer-to-peer test format (Brown, 2003; Chalhoub-Deville, 2003; Davis, 2009), and of the influence of certain personal characteristics of the speakers in the tests (Nakatsuhara, 2011; Ockey, 2009). Several informants mention concerns about pairing students with different levels of language ability, as well as about how well acquainted the students are with one another in test situations. These are concerns also mentioned in previous research (Davis, 2009; Gan, 2010; Ikeda, 1998; Iwashita, Brown, McNamara, & O'Hagan, 2008). Most informants stress the importance of interaction between the speakers in the peer-to-peer conversations. They are conscious of the risk of interlocutors

speaking in parallel to one another performing "solos" instead of interacting, and in their comments reflect findings from previous studies (Galaczi, 2008). Furthermore, the teachers in this study are clearly mindful of their own influence on the student performances, as discussed by e.g. Sandlund and Sundqvist (2011).

In other words, these skilled and experienced English teachers are well aware of the complexity of testing oral proficiency. However, the lack of time makes it difficult for them to prepare by properly reviewing benchmarks and assessment factors before the tests, potentially creating a gap between the prescribed rating criteria and the criteria actually used.

7.1.4 Informants on Grading Oral Proficiency

Although the teachers in this study are experienced as teachers and raters, they express having difficulties in making the grounds for their grading explicit and transparent. This is consistent with research that has shown that experienced raters consult the full rating rubrics less often (Joe et al., 2011; Orr, 2002) and rely on implicit criteria more often (Bejar, 2012; Joe et al., 2011). As already mentioned, not regularly reviewing benchmarks and assessment factors allows for larger degrees of subjective judgments, potentially jeopardizing fair grades.

When grading the student examples in the group activities of the study, the informants largely interpreted the rating criteria according to their respective orientations with regard to oral proficiency. In one instance, the groups awarded different grades; in the other, they awarded the same grade, but on different grounds, which reflects findings from previous research (Brown, 2000; Orr, 2002).

Being lenient or harsh as a rater is a common reason for variation in rating, according to several studies (Borg, 2003). This is commented on in the individual interviews by some of the teachers (see Section 6.4.1), showing that they are aware of this phenomenon. Being experienced is another factor that influences rating, according to the informants. Experience gives confidence in grading, they claim. The positive impact of experience has previously been reported in other research involving Swedish teachers (e.g. Forsberg & Wermke, 2012).

The general proficiency level of a conversation in an oral test tends to influence the raters to focus on either content or language. A lower level of proficiency generates more comments on language, whereas a higher level of

proficiency stimulates more comments on content, according to reports (Ang-Aw & Goh, 2011; Brown et al., 2005; McNamara, 1996). This tendency may have contributed to the focus in the group discussions of this study as well, as student example 2, by both groups considered to show a rather low proficiency, resulted in fewer comments on content than example 1. However, the orientations of the group members seem to have had a stronger effect on what features were commented on in general, as group I in both examples reflected more on linguistic than on other features (6 out of the 10 most frequently noted categories of features). In group II, out of the 10 most frequently referred to categories of features, 4 were instead interactional features. The focus of the two groups seem to reflect the orientations to oral proficiency of the group members; group I representing all three orientations and group II lacking members with a language orientation.

A certain insecurity about the new criteria appears to lead to comparisons between student efforts, instead of comparing with rating criteria, which has been described previously (Bejar, 2012; Orr, 2002). The fact that teachers are able to verbally use the concepts and notions of new policy documents does not necessarily mean these concepts have been internalized and that the teachers are able to fully apply them.

In the rating activity of the study, the informants discussed their assessment and grading in ways that aligned with findings in previous research, demonstrating differences among skilled subject teachers making qualitative judgments. The subjective element in qualitative judgment of student attainment is apparent in the way the grades and the comments of the teachers reflect their various orientations to oral proficiency. The anticipated difficulty in assessing interactional features due to lack of detailed interactional assessment factors was not detectible in the discussions.

7.1.5 Informants on Framing Factors

In the interviews, the informants make distinct references to perceived differences in interpretations of policy documents and assessment factors among teachers, (c.f. Q21, Q28, Q29, Q30 and Q57). All of them clearly identify their own school as a certain "type" of school with a specific "type" of students, indicating their awareness of the growing differences between schools (Östh et al., 2013). They all refer to their respective *student bodies* and comment on the risk that they might interpret the performance standards

according to the results among their own group of students, instead of assessing according to the official performance standards. They therefore want *time* to meet teachers from other schools for deliberations on policy documents, assessment and grading, as well as more time for preparation for, and discussions on, national tests at their own schools. The lack of time and routines for discussions and meetings on these issues among teachers in Sweden has been shown in other studies (Wedin, 2007). Several of the informants also state that the *community of teachers* at their school has a strong influence on how the test is organized and how different grade levels are interpreted and broken down in matrixes and rubrics. They view themselves as parts of these communities, being influenced as well as influencing them. Differences in student body, community of colleagues and time made available are thus factors that the informants interpret as framing their rating practice (C. Gustafsson, 1999) when it comes to the oral part of the national test. These factors contribute to possible differences in assessment of student achievement, according to the informants. Age and subject specialty were also brought up, but were not investigated here.

7.2 Answers to Research Questions

The objective of this study has been to investigate how teachers perceive, test and assess oral proficiency. Through interviews and group discussions after a rating activity, the perceptions and rating practice of a number of teachers were explored. The statements of the informants were transcribed and analyzed and answers to the three research questions were sought. Below tentative answers to the questions are summarized.

RQ1: How do the informants define oral proficiency?

The informants differ in how they describe oral proficiency in ways that reflect three identifiable orientations with a focus on either, content, language or interaction. These orientations to, or pedagogical beliefs about, oral proficiency are strongly held and influence these teachers' interpretations of policy documents as well as their practice (see Section 4.1). The three orientations are in accordance with the intended curriculum, do not exclude one another and can be seen as expressions of professional and pedagogical content knowledge or pedagogical constructions (Gudmundsdottir, 1991; Gudmundsdottir & Shulman, 1987; Hashweh, 2005; Shulman, 1986).

RQ2: How do the informants describe the organization of the oral part of the national test of English in year 9 at their respective schools?

The national test, with its oral part, is seen as a natural part of practice; it is well known and accepted. The informants state that they have well-functioning routines for the procedures of the test. They acknowledge the complications of the test format, which leads to pedagogical and didactical considerations when organizing the test. Great efforts are put into grouping students in optimal ways and organizing for every learner to show his/her proficiency.

The rating of student efforts is carried out holistically, as prescribed in the instructions. Eleven out of twelve informants are the sole raters of their own students' oral performances. Half of the participants state that they audio-record the tests, according to recommendations in the guidelines for the test, thus making a second rating possible. However, only two teachers listen a second time to the interactions and only one teacher routinely has the support of a second rater in the grading of the test. Not all informants listen to the benchmark CD every year, which means that not all of the informants regularly review grade levels before rating the student efforts.

RQ3: What influences how the informants rate the oral proficiency tests?

The study shows that the *orientations* of the informants influence how they interpret policy documents, and consequently how they rate students' oral proficiency in the national test. Individual pedagogical beliefs seem to lead to the different orientations among the teachers in the study.

Their own *teacher experience* is reported by informants themselves as an important source for knowledge about, and confidence in, assessment, which tallies with other research (Forsberg & Wermke, 2012; Munby & Russell, 1993). *Discussions* among colleagues is another source of perceived support and professional development in assessment practice, which has been found in other research also (Forsberg & Wermke, 2012).

The informants further identify certain local circumstances, here referred to as *frame factors*, as having an influence on their assessment practice. The 'school culture', shaped by the colleagues as well as the student body of the schools, is perceived as such a framing factor. Another factor emphasized by all informants as impacting their pedagogical decisions, are time constraints,

limiting rating discussions with colleagues, and time granted for proper preparation and reflection. This is seen as a threat to fairness in grading.

Didactical Implications

The results of this study have several didactical implications. Firstly, to enhance conformity and equity in grading the oral part of the national test of English further, teachers need to be given the proper conditions to prepare, carry out and assess the tests. This implies more time to prepare the test and assess the results.

Secondly, teachers need to discuss the test at their schools, but they also need to be allowed time for meetings between schools and within school districts or regions, to discuss and calibrate their interpretations of knowledge requirements and assessment factors, as well as their grading of student performances on the test.

Thirdly, to support a common understanding and interpretation of policy documents, time for continuous assessment discussions on all levels, (not only for the national test), is vital. Principals, school administrations and municipalities need to encourage such discussions.

Fourthly, to address the perceived lack of previous assessment instruction and training, the annual guidelines for the national test need to be studied carefully and treated as continuous professional development in assessment by schools and teachers. The national test is time consuming and needs to deliver not only support for fair grading, but also development of expert knowledge and competence in assessment and grading to schools and individual teachers.

Teachers' assessment and rating practices need to be strengthened instead of questioned. Qualitative assessments by skilled professionals are needed. Complex competencies cannot and should not be reduced to what is easily measurable and comparable. Oral proficiency as a multi-faceted skill cannot be captured with statistical exactness in multiple-choice tests or simple matrixes; experienced and professional language teachers are needed to make the judgments. A modern school system needs to make the necessary investments to develop its ability to achieve reasonable exactness through qualitative methods of measuring performances, and then adapt its use of the results to the level of exactness that can be achieved with these methods.

7.3 Further Research

In compulsory school, Swedish teachers grade the attainment of their own students. A large number of policy documents support them in their assessment practice. In addition to the national curriculum, with the subject syllabi and the national tests for several subjects, there are diagnostic materials for many subjects. There is also advisory material as well as additional regulations in documents of different kinds available to all teachers (and the general public) on the home page of Skolverket. In December 2014, the following documents concerning aspects of English and assessment of English, among others, were available:

Table 13. Further policy documents

Materials for English	Materials for Grading and Assessment
Materials for Group Discussions – English[24]	The Grading Scale and the Grades B and D[25]
Commentary on the Syllabus for English (Skolverket, 2011a)	Aspects of Assessment[26]
Commentary on Performance Standards for English[27]	Collaborative Rating[28]
	Knowledge Assessment in School – Practice, Notions, Problems and Opportunities (Skolverket, 2011b)

A closer reading of these texts would provide a broader picture of the intended curriculum for oral proficiency in English. An investigation into how well known and how well used these documents are by teachers would be appropriate in connection with this, and could also provide guidance for Skolverket on future directions.

The present study also highlights further problem areas with regard to assessment of oral proficiency. A study investigating a representative sample of authentic, recorded and graded student performances from the oral subtest, collected from schools all around the country, would be valuable to more securely establish what factors lead to deviations, versus what helps establish a consensus among raters.

[24] http://www.skolverket.se/publikationer?id=2531
[25] http://www.skolverket.se/publikationer?id=2953
[26] http://www.skolverket.se/publikationer?id=3259
[27] http://www.skolverket.se/publikationer?id=2825
[28] http://www.skolverket.se/publikationer?id=3172

The possible impact of a 'school culture', (i.e. specific student body and community of colleagues), on the grading of student achievement in English, as reported by teachers in this study, also needs to be investigated.

7.4 Concluding Reflections

The statements made by the informants in the study show that they consider fair and uniform grades to be of vital importance. As fair grades underpin the fundamental democratic principles of equal opportunities, policy documents explicitly mandate procedures to ensure that grading criteria are consistently and reliably employed. All informants in this study emphasize their commitment to meet these expectations. They also have the professional competences and qualifications required to do so.

Quality assurance in the production of national tests in general has not been questioned. Skolverket has found the tests to be valuable instruments for achieving equivalence in grading (Skolverket, 2007). Other studies have shown them to be reasonably valid in measuring students' subject knowledge (Skolverket, 2007; Åberg-Bengtsson & Erickson, 2006). Yet, the results of this study confirm that the nature of oral proficiency remains difficult to capture with instrumental exactness. Potential deviations between different teachers' ratings of student performances seem difficult to avoid, even if competent professionals work hard to do it 'by the book'. More time for deliberations and discussions within schools as well as between schools are needed to minimize differences.

As the grading system is based on the underlying assumption that teachers understand the rating criteria in a similar way, these discussions and social moderation practices (Jönsson & Thornberg, 2014) are important, to develop fair grading and to give teachers confidence and strengthen their competence in grading their students efforts. These discussions are also necessary to fulfill the first objective of the national tests, "The national tests' primary objectives are to support fair and equal assessment..."[29] The findings in this study indicate that there is room for improvement, both within individual schools and on the level of municipalities or regions.

An assessment practice that feeds back to both teachers and students and feeds forward for further improvement is, today, a natural and integral part of

[29] http://www.skolverket.se/bedomning/nationella-prov, retrieved 2015-01-05, *my translation*.

a school system. However, when assessments are used outside their primary context of the teaching practice in schools, it is important to be aware of what the results of testing really suggest about students' achievements. School results and grades have to be interpreted in their context, to be understood as measures of educational efficiency and attainment.

> There is also a growing interest in assessment outcomes as performance indicators, as well as national and political pressure for highly reliable "objective" assessment in order to be able to rank order and grade the performance of individuals and schools (and even countries). At the same time, /…/, developments in cognition and learning are telling us to assess more broadly, in context, and in depth. (Gipps, 1999, p. 384)

Skolinspektionen suggested, after the investigation into teachers' grading of national tests, that performance testing should be excluded from the tests (Skolinspektionen, 2013). A recent report also claimed that the second purpose of the national tests, "to enable an analysis of to what extent knowledge requirements have been met within individual schools and school districts as well as on the national level"[30] is not fully met with the current tests (J.-E. Gustafsson, Cliffordson, & Erickson, 2014). Qualitative measures of complex proficiencies are simply not easily captured in clearly comparable results. It is a question of skilled professionals with pedagogical content knowledge assessing complex tasks, in which students demonstrate their level of mastery of multi-faceted proficiencies. Surely, the consequence of the above reports cannot be to exclude the measuring of complex proficiencies, thus diminishing their importance in education (Mickwitz, 2011)? Instead, further development of ways to measure and ways to reach a common understanding of policy documents should be the way forward. Also the worries of teachers, that their local practice and interpretation of the policy documents do not coincide with that of other schools, need to be taken seriously. This is particularly important in times of growing segregation in society and schools, when the risk is high that specific school cultures frame how time available is used and how policy documents are interpreted.

[30] http://www.skolverket.se/bedomning/nationella-prov, retrieved 2015-01-05, *my translation*.

Svensk sammanfattning

Betyg är idag, förutom att vara ett kvitto för den enskilde eleven på lärande och uppnådd kunskap, ett sätt att utvärdera hur lärare och skolor, kommuner och nationer lyckas i sina utbildningsuppdrag. Kunskapsresultaten har kommit att spela en allt viktigare roll i den allmänna debatten om skolan. För Sveriges del har denna debatt intensifierats ytterligare av de sjunkande resultaten i internationella jämförelser som PISA[31], TIMSS[32] och PIRLS[33] samt av diskussioner om den ökande segregationen i samhället och skolan (Östh et al., 2013). Bedömningars och betygs likvärdighet har ifrågasatts och undersökts (se t ex Skolinspektionen, 2013).

Följaktligen har en rad förändringar under senare tid genomförts inom skolans område. Bland annat har en ny läroplan, *Lgr11*, med nya kursplaner i alla ämnen och nya betyg introducerats för att stärka kunskapsuppdraget och tydliggöra målen för utbildningen (Prop. 2008/09:87). Engelska är fortsatt ett framträdande ämne i läroplanen för grundskolan och en kommunikativ språksyn framhålls tydligt. Den muntliga kommunikativa produktionen och interaktionen är, jämte den skriftliga kommunikativa förmågan och interaktionen, framlyft i styrdokumenten. Muntlig språkfärdighet är även det område både elever och lärare finner viktigast att ha tillägnat sig efter avslutad skolgång (Erickson, 2010). Hur denna färdighet testas, bedöms och betygsätt är därför av intresse.

Hur lärare definierar muntlig språkfärdighet och hur de uppfattar att färdigheten beskrivs och definieras i de nya styrdokumenten samt vad som påverkar deras bedömningspraktik när det gäller muntlig språkfärdighet i engelska, är ämnet för denna licentiatuppsats. Forskningsfrågorna är:

- Hur uppfattar lärare muntlig språkfärdighet?
- Hur organiserar, bedömer och betygsätter lärare den muntliga delen av det nationella provet i engelska i årskurs 9?
- Vad påverkar lärares bedömning av muntlig språkfärdighet?

---

[31] Programme for International Student Assessment, http://www.oecd.org/pisa/
[32] Trends in International Mathematics and Science Study, http://timssandpirls.bc.edu/#
[33] Progress in International Reading Literacy Study, http://timssandpirls.bc.edu/pirls2011/index.html

I undersökningen har tolv lärare intervjuats. Sex av de intervjuade lärarna deltog även i sambedömningsmöten med efterföljande gruppdiskussion. Deltagarna i sambedömningen blev därefter intervjuade ytterligare en gång. Intervjuer och gemensamma bedömningssamtal har transkriberats och analyserats. Som verktyg har konkordansanalys (med hjälp av Antconc[34]), delar av Spradleys etnografiska intervjuanalys (Spradley, 1979, 1980) och ramfaktorer (C. Gustafsson, 1999; Lundgren, 1972) använts.

Lärarna i denna undersökning tycker att de fått god information om *Lgr11*och de nya betygen, vilket skiljer sig från resultaten i andra undersökningar (Skolverket, 2013). Den vidare analysen av utsagorna i studiens intervjuer och gruppdiskussioner visar på tre olika inriktningar till muntlig språkfärdighet. En grupp informanter fokuserar på det *innehåll* som behandlas i den muntliga interaktionen, en grupp är mer inriktad på det *språk* som används och en tredje grupp har fokus på själva *interaktionen* elever emellan. Dessa olika uppfattningar om tyngdpunkten i muntlig språkfärdighet verkar baserade på starka personliga övertygelser, en pedagogisk filosofi, och påverkar hur dessa lärare tolkar styrdokument, kunskapskrav och bedömningsfaktorer. En personlig pedagogisk filosofi är samtidigt en förutsättning för att lärare ska kunna omvandla ett ämnesinnehåll till ett fungerande pedagogiskt innehåll (Hashweh, 2005; Shulman, 1986).

Lärare som prioriterar *innehållet* i den muntliga interaktionen fäster avseende vid elevernas förmåga att förmedla en åsikt, att formulera idéer och argumentera för olika ståndpunkter i diskussioner och debatter med andra. Detta liknar språkinlärningsteorier som menar att mer krävande uppgifter, där eleverna måste anstränga sig mer för att bli förstådda, genererar språkinlärning, som i t.ex. the Output Hypothesis (Swain, 2000; Swain & Lapkin, 1995). Stöd för att innehållet i diskussioner och åsiktsutbyten är av vikt finns både i läroplanens kapitel 2 (Skolverket, 2011c, sid. 12-13), i kursplanens skrivningar (se appendix A) och i bedömningsfaktorerna till det muntliga provet, där olika delar av kommunikationens innehåll fokuseras (se appendix B).

Lärare som fokuserar på *språket* talar om att utrusta eleverna med en "verktygslåda" av grundläggande uttryck, idiom och fraser, såväl som grammatik att använda i muntlig interaktion. Det handlar om att ge alla elever

[34] http://www.antlab.sci.waseda.ac.jp/software.html

förutsättningar att göra sig förstådda, men också om att ge eleverna en säkerhet när de interagerar, så att de vågar tala och delta i olika typer av samtal. Den här inställningen liknar mer en traditionell inställning till språkinlärning; baskunskaperna först, så att inläraren vet vilka verktyg som finns till hands för att uttrycka ett budskap. I det centrala innehållet i kursplanen finns en del språkliga punkter som grund för denna åsikt (Skolverket, 2011c, sid 32-33), såväl som i bedömningsfaktorerna, där olika aspekter av språk och uttrycksförmåga lyfts fram.

Lärarna som anser att *interaktionen* är viktigast uppmuntrar eleverna att tala med varandra, att hålla ett samtal igång på ett så naturligt och korrekt sätt som möjligt. De försöker motivera alla elever genom att aktivt erbjuda samtalsämnen som är aktuella för eleverna och som engagerar. Själva utbytet av information och åsikter, den sociala interaktionen, står i centrum. Detta påminner om teorier som t.ex. the Interaction Hypothesis (Long, 1983, 1985), där den muntliga kommunikationen i sig, där samtalspartners försöker göra sig förstådda, respektive förstå varandra, ses som en viktig del i språkinlärningen. Också här finns stöd i läroplanen, som t.ex. i skrivningen om språket som ett verktyg för lärande i kommunikation med andra (Skolverket, 2011c, s. 30). En jämförelse med styrdokumenten ger alltså vid handen att dessa är så formulerade, att stöd för samtliga tre inriktningar kan hittas i dokumenten.

Det nationella provet ses av lärarna i studien som en naturlig del av undervisningen. De genomför dem årligen och känner väl till hur de fungerar. De lägger vidare ner stor möda på att arrangera proven så, att eleverna har optimala möjligheter att visa sina muntliga språkfärdigheter. Medvetenheten om den muntliga språkfärdighetens komplexitet och svårigheten att fånga denna färdighet i ett prov är stor. Trots detta spelar bara hälften av lärarna in proven och bara ett fåtal lyssnar regelbundet en andra gång på samtalen innan betygen sätts. Elva av de tolv lärarna är ensamma om att betygsätta sina egna elevers kunskaper i den muntliga delen av provet. Flera av informanterna angav i samtalen att de inte regelbundet, dvs. varje år, lyssnar igenom exempel-CD:n som medföljer proven. Några läste kommentarerna till bedömningsexemplen, men inte alla. Inte heller våren 2013, då betygskriterierna var nya och proven för första gången var baserade på *Lgr11,* hade samtliga deltagare orienterat sig i bedömningsexemplen. De flesta deltagarna angav att de ansåg sig ha för lite tid till förberedelse och bedömning av proven.

I den sambedömningsaktivitet som ingick i studien betygsatte de två grupperna informanter elevexemplen olika. I det ena exemplet fick elevernas prestationer olika betyg, i det andra gavs prestationerna samma betyg i båda grupperna, men med olika motivering. I motiveringen till bedömningarna kunde informanternas tre olika inriktningarna till muntlig språkfärdighet spåras.

Lärarna i studien tog själva upp faktorer som de ansåg kan påverka bedömningen. De oroar sig för att de sätter sina betygsnivåer efter det elevunderlag deras skola har och att de påverkar varandra inom sina respektive lärarkollegier. De ansåg även att bristen på tid för förberedelse, genomförande och framförallt gemensamma diskussioner runt bedömning och betygsättning av proven påverkar deras möjligheter att försäkra sig om att de sätter rättvisa betyg. De önskar mer tid för proven och för gemensamma bedömnings-diskussioner, något som behöver uppmärksammas med tanke på den ökande skolsegregationen (Östh et al., 2013).

För att elevers kunskaper och färdigheter ska bli likvärdigt bedömda krävs att lärare delar synen på hur kunskaper, språkfärdighet och kunskapskrav bör tolkas. När det gäller en komplex färdighet som muntlig språkfärdighet i ett främmande språk, krävs kvalitativa bedömningar av erfarna och välutbildade lärare. Kvalitativa pedagogiska bedömningar är emellertid till en del subjektiva. Likvärdigheten i sådana bedömningar kräver därför mycket tid för gemensamma tolkningsdiskussioner, så att skillnaderna i bedömningarna minimeras och så att så stor samstämmighet som möjligt uppnås.

Slutsatser av denna studie är att, för att öka samstämmigheten i bedömningen och betygsättningen av den muntliga språkfärdigheten i engelska i årskurs 9, behöver lärare för det första, årligen ges tid till förberedelse, till genomförande och tid för bedömning av proven. För det andra behöver lärare diskutera sina bedömningar med varandra på den lokala skola, men även med lärare från andra skolor. Förutsättningar för sådana gemensamma diskussioner behöver skapas av ledare på skolnivå såväl som på kommunal nivå. För det tredje bör gemensamma diskussioner rörande tolkningen av styrdokumenten och kunskapskraven i allmänhet, utanför de nationella proven, anordnas, för att ytterligare öka samsynen lärare emellan. För det fjärde behöver lärarinformationen som tillhandahålls med de nationella proven ses som en årlig fortbildning i syfte att stärka lärares bedömarkompetens och därför ges det utrymme den kräver.

Lärare behöver fortsatt stärka sin bedömarkompetens och få stöd i den kvalitativa bedömningen av mångfacetterade färdigheter genom kollegiala diskussioner. Förutsättningarna för att göra professionella pedagogiska bedömningar behöver ständigt förbättras. Risken finns annars att det lätt mätbara görs till det viktigaste och att mer kvalitativa värderingar avlägsnas från de nationella proven till förmån för mer instrumentella bedömningar (se Skolinspektionen, 2013). Detta kan i sin tur leda till att moment som kräver kvalitativa bedömningar får mindre utrymme i undervisningen (Mickwitz, 2011). En utveckling av fler mätinstrument för muntlig språkfärdighet och vägar att nå samsyn när det gäller styrdokument är en bättre väg framåt. Lärares oro över att de inte delar en gemensam tolkning av styrdokumenten och betygskriterierna måste tas på allvar. Detta är särskilt viktigt i tider av ökad segregation i samhälle och skola. Risken är annars stor att en lokal skolkultur präglar tolkningen av dokumenten och hur den tid som finns till förfogande används, vilket i sin tur ytterligare ökar segregationen.

Bibliography

Ang-Aw, H. T., & Goh, C. C. M. (2011). Understanding discrepancies in rater judgement on national-level oral examination tasks. *RELC Journal, 42*(1), 31-51.

Apelgren, B. M. (2001). *Foreign Language Teachers' Voices. Personal Theories and Experiences of Change in Teaching English as a Foreign Language in Sweden.* (PhD), Gothenburg University, Göteborg. (Göteborg Studies in Educational Sciences 155)

Apelgren, B. M. (2013). What's Grammar Got to Do with it? In G. Bergh, Bowen, R. & Mobärg, M. (Ed.), *Language, Football and All That Jazz. A Festschrift for Sölve Ohlander.* Göteborg.

Archer, M. S. (1995). *Realist social theory : the morphogenetic approach.* Cambridge: Cambridge University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford Univ. Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice : developing language assessments and justifying their use in the real world.* Oxford: Oxford University Press.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research, 23*(4), 459-484.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133-180.

Bejar, I. I. (2012). Rater cognition: Implications for validity. *Educational Measurement: Issues and Practice, 31*(3), 2-9.

Black, P., Harrison, C., Lee, C., & Marshall, B. (2003). *Assessment for learning: Putting it into practice.* Buckingham: Open University Press.

Black, P., & Wiliam, D. (1998). *Inside the black box : raising standards through classroom assessment.* London: GL Assessment.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*(02), 81-109.

Borger, L. (2014). *Looking Beyond Scores. A Study of Rater Orientations and Ratings of Speaking.* (Licentiate Thesis), Gothenburg University, Gothenburg

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15.

Broadfoot, P. (1996). *Education, assessment, and society: A sociological analysis.* Buckingham: Open University Press.

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35-36.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142.

Brookhart, S. M. (1994). Teachers' Grading: Practice and Theory. *Applied Measurement in Education, 7*(4), 279.

Brookhart, S. M., & Freeman, D. J. (1992). Characteristics of entering teacher candidates. *Review of educational research, 62*(1), 37-60.

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing, 26*(3), 341-366. doi: 10.1177/0265532209104666

Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing, 12*(1), 1-15.

Brown, A. (2000). An investigation of the rating process in the IELTS oral interview. *IELTS research reports, 3*, 49-84.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20*(1), 1-25.

Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks. *Research Report - Educational Testing Service, Princeton RR, 5*.

Carlgren, I., & Lindblad, S. (1991). On teachers' practical reasoning and professional knowledge: Considering conceptions of context in teachers' thinking. *Teaching and Teacher Education, 7*(5–6), 507-516. doi: 10.1016/0742-051X(91)90045-Q

Carlone, H. B. (2004). The cultural production of science in reform-based physics: Girls' access, participation, and resistance. *Journal of Research in Science Teaching, 41*(4), 392-414.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187-1218.

Chalhoub-Deville, M. (2003). Second language interaction: current perspectives and future trends. *Language Testing, 20*(4), 369-383. doi: 10.1191/0265532203lt264oa

Chesebro, J. W., & Borisoff, D. J. (2007). What makes qualitative research qualitative? *Qualitative Research Reports in Communication, 8*(1), 3-14.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: M.I.T. Press.

Clandinin, D. J., & Connelly, F. M. (1992). Teacher as curriculum maker. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 363-401). New York: Macmillan.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. Milton Park, Abingdon, Oxon, [England]: Routledge.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEFR). Council of Europe. Retrieved from http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp.

The Common European Framework of Reference: The globalisation of language education policy. (2012). (M. Byram & L. Parmenter Eds. Vol. 23). Bristol: Multilingual matters.

Cuban, L. (1990). Reforming Again, Again, and Again. *Educational Researcher, 19*(1), 3-13. doi: 10.2307/1176529

Cuban, L. (2013). Why so many structural changes in schools and so little reform in teaching practice? *Journal of Educational Administration, 51*(2), 109-125.

Dahllöf, U. (1967). *Skoldifferentiering och undervisningsförlopp : komparativa mål- och processanalyser av skolsystem 1*. Stockholm: Almqvist & Wiksell.

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing, 26*(3), 367-396. doi: 10.1177/0265532209104667

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*(3), 305-334.

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of Speaking Proficiency. *Studies in Second Language Acquisition, 34*(1), 5-34. doi: 10.1017/S0272263111000489

Denscombe, M. (2007). *The good research guide for small-scale social research projects*. Maidenhead: Open University Press.

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing, 11*(2), 125-144.

Ducasse, A. M. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing, 26*(3), 423-443.

Ek, J. A. v. (1975). *The Threshold Level*. Strasbourg: Council of Europe.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford Univ. Press.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition, 24*(02), 223-236.

Erickson, G. (1991). Muntliga prov i engelska i årskurs 8. *LMS Lingua*(3), 97-102.

Erickson, G. (2009). *Nationella prov i engelska - en studie av bedömarsamstämmighet*. Stockholm: Skolverket.

Erickson, G. (2010). A New Look at Teaching and Testing: English as Subject and Vehicle. In T. Kao & Y. Lin (Eds.), *Good Practice in Language Testing and Assessment - A matter of Responsibility and Respect* (pp. 237-258). Taipei, Taiwan: Bookman Books Ltd.

Erickson, G. (2012). *National assessment of foreign languages in Sweden*. Göteborg: Gothenburg University.

Erickson, G., & Börjesson, L. (2001). Bedömning av språkfärdighet i nationella prov och bedömningsmaterial. In R. Ferm & P. Malmberg (Eds.), *Språkboken. En antologi om språkundervisning och språkinlärning*. Stockholm: Skolverket.

Feldt, J. (2010). *Öppna jämförelser. Grundskola 2010*. Stockholm: Sveriges kommuner och landsting.

Ferm, R., & Malmberg, P. (Eds.). (2001). *Språkboken : en antologi om språkundervisning och språkinlärning*. Stockholm: Skolverket.

Forsberg, E., & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers' continuing professional development of assessment knowledge. *Professional Development in Education, 38*(5), 741-758.

Fulcher, G. (1996). Testing tasks: issues in task design and the group oral. *Language Testing, 13*(1), 23-51.

Galaczi, E. D. (2008). Peer-Peer Interaction in a Speaking Test: The Case of the "First Certificate in English" Examination. *Language Assessment Quarterly, 5*(2), 89-119.

Gan, Z. (2010). Interaction in group oral assessment: A case study of higher- and lower-scoring students. *Language Testing, 27*(4), 585-602. doi: 10.1177/0265532210364049

Gan, Z., Davison, C., & Hamp-Lyons, L. (2009). Topic Negotiation in Peer Group Oral Assessment Situations: A Conversation Analytic Approach. *Applied Linguistics, 30*(3), 315-334. doi: 10.1093/applin/amn035

Giddens, A. (1979). *Central problems in social theory : action, structure and contradiction in social analysis*. Berkeley: Univ. of California Press.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of research in education, 24 (1999)* 355-392.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*(2), 479-507.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, Calif: Sage.

Gudmundsdottir, S. (1991). Ways of seeing are ways of knowing. The pedagogical content knowledge of an expert English teacher. *J. Curriculum Studies, 23*(5), 409-421.

Gudmundsdottir, S., & Shulman, L. (1987). Pedagogical content knowledge in social studies. *Scandinavian Journal of Educationl Research, 31*(2), 59-70.

Gustafsson, C. (1999). Ramfaktorer och pedagogiskt utvecklingsarbete. *Pedagogisk forskning i Sverige, 4*(1), 43.

Gustafsson, J.-E., Cliffordson, C., & Erickson, G. (2014). *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter*: SNS Utbildningskommission.

Gustafsson, J.-E., & Erickson, G. (2013). To trust or not to trust?—teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability, 25*(1), 69-87.

Hall, K., & Harding, A. (2002). Level descriptions and teacher assessment in England: towards a community of assessment practice. *Educational Research, 44*(1), 1-16.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. London: Arnold.

Hammersley, M. (1984). *The ethnography of schooling : methodological issues*. Driffield: Nafferton.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 1051-1078). Amsterdam: North-Holland.

Hargreaves, A. (1994). *Changing teachers, changing times: Teachers' work and culture in the postmodern age*. London: Cassell.

Harik, P., Clauser, B. E., Grabovsky, I., Nungester, R. J., Swanson, D., & Nandakumar, R. (2009). An examination of rater drift within a generalizability theory framework. *Journal of Educational Measurement, 46*(1), 43-58.

Hashweh, M. Z. (2005). Teacher pedagogical constructions: a reconfiguration of pedagogical content knowledge. *Teachers and Teaching, 11*(3), 273-292.

Hedge, T. (2001). *Teaching and learning in the language classroom* (Vol. 106): Oxford University Press Oxford,, UK.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly, 40*(1), 109-131.

Hyltegren, G. (2014). *Vaghet och vanmakt - 20 år med kunskapskrav i den svenska skolan*. (PhD), Gothenburg University, Göteborg.

Hymes, D. (1962). The ethnography of speaking. *Anthropology and human behavior, 13*(53), 11-74.

Hymes, D. (1972). On Communicative Competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics : selected readings*. Harmondsworth: Penguin.

Ikeda, K. (1998). The paired learner interview: A preliminary investigation applying Vygotskian insights. *Language Culture and Curriculum, 11*(1), 71-96.

Iwashita, N. (1999). The validity of the paired interview format in oral performance assessment. *Melbourne Papers in Language Testing, 8*(1), 51-66.

Iwashita, N. (2010). *Features of oral proficiency in task performance by EFL and JFL learners*. Paper presented at the Selected Proceedings of the 2008 Second Language Research Forum.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24-49.

Joe, J. N., Harmes, J. C., & Hickerson, C. A. (2011). Using verbal reports to explore rater perceptual processes in scoring: A mixed methods application to oral communication assessment. *Assessment in Education: Principles, Policy & Practice, 18*(3), 239-258.

Jönsson, A., & Thornberg, P. (2014). Sambedömning: effektiv kompetensutveckling eller redskap för likvärdigbedömning i skolan? *Pedagogisk forskning i Sverige, 19*, 386-402.

Kramsch, C. (1986). From Language Proficiency to Interactional Competence. *Modern Language Journal, 70*(4), 366.

Kramsch, C. (2006). From Communicative Competence to Symbolic Competence. *Modern Language Journal, 90*(2), 249-252. doi: 10.1111/j.1540-4781.2006.00395_3.x

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Oxford University Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Krashen, S. D. (1992). Formal grammar instruction. Another educator comments. *TESOL Quarterly, 26*(2), 409-411.

Krashen, S. D. (1993). *The power of reading: Insights from the research*. Englewood, Connecticut: Libraries Unlimited.

Kvale, S. (1995). The social construction of validity. *Qualitative inquiry, 1*(1), 19-40.

Kvale, S. (1997). *Den kvalitativa forskningsintervjun*. Lund: Studentlitteratur.

Lee, J. S., Nargund-Joshi, V., & Dennis, B. (2010). Progressing through the Haze in Science and Mathematics Education Research: Contemporary

Use of Spradley's Qualitative Inquiry in Two Case Studies. *International Journal of Qualitative Methods, 10*(1), 42-57.

Leung, C. (2005). Convivial communication: recontextualizing communicative competence. *International Journal of Applied Linguistics, 15*(2), 119-144.

Leung, C., & Teasdale, T. (1997). What do teachers mean by speaking and listening: a contextualised study of assessment in the English National Curriculum. In A. Huhta, V. Kohonen, L. Kurki-Sounio, & S. Luoma (Eds.), *New contexts, goals and alternatives in language assessment.* (pp. 291-324). Jyväskylä: University of Jyväskylä.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, Calif.: Sage.

Lincoln, Y. S., & Guba, E. G. (2000). The only generalization is: There is no generalization. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case study method: Key Issues, Key Texts* (pp. 27-44). London: Sage.

Lindblad, S., Linde, G., & Naeslund, L. (1999). Ramfaktorteori och praktiskt förnuft. *Pedagogisk forskning i Sverige, 4*(1), 93.

Lindblad, T. (1991). Allt är relativt: centralt utarbetade prov i svenska, matematik, engelska, franska, tyska, fysik och kemi: beskrivning, analys, förslag *Rapporter från institutionen för pedagogik* (Vol. 1991:03). Göteborg: Göteborgs universitet, Institutionen för pedagogik.

Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126-141.

Long, M. (1985). Input and second language acquisition theory. *Input in second language acquisition*, 377-393.

Lundahl, C. (2006). *Viljan att veta vad andra vet: kunskapsbedömning i tidigmodern, modern och senmodern skola.* (PhD), Uppsala universitet, Stockholm. (2006:8)

Lundahl, C. (2009). *Varför nationella prov?: framväxt, dilemman, möjligheter.* Lund: Studentlitteratur.

Lundahl, C. (2011). *Bedömning för lärande.* Stockholm: Norstedt.

Lundgren, U. P. (1972). *Frame factors and the teaching process : a contribution to curriculum theory and theory on teaching.* Stockholm: Almqvist & Wiksell.

Lundgren, U. P. (1979). *School curricula : content and structure and their effects on educational and occupational careers.* Stockholm: Dep. of educational research, Stockholm school of education

Lundgren, U. P., & Nihlfors, E. (2005). *Det gemensamma-en studie av några kunskapselement i den för alla lärarutbildningar gemensamma delen. Högskoleverkets rapportserie 2005:17* R.

Luoma, S. (2004). *Assessing speaking.* Cambridge: Cambridge University Press.

Magnusson, A. (1998). *Lärarkunskapens uttryck.* (PhD), Linköping university, Linköping. (Linköping studies in Education and Psychology No 58)

Marklund, S. (1987). *Skolsverige 1950-1975: D. 5, Läroplaner*. Stockholm: Liber/Utbildningsförl.

May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing, 26*(3), 397-421. doi: 10.1177/0265532209104668

May, L. (2010). Developing speaking assessment tasks to reflect the 'social turn' in language testing. *University of Sydney Papers in TESOL, 5*(5), 1-30.

May, L. (2011). Interactional Competence in a Paired Speaking Test: Features Salient to Raters. *Language Assessment Quarterly, 8*(2), 127-145. doi: 10.1080/15434303.2011.565845

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34-43.

McNamara, T. F. (1996). *Measuring second language performance*. New York: Longman.

Messick, S. (1993). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13-104). Phoenix, Arizona: Oryx Press.

Mickwitz, L. (2011). *Rätt betyg för vem?: Betygsättning som institutionaliserad praktik*. (Licentiate Thesis), Stockholm University Stockholm. (Studier i språkdidaktik 3)

Mishler, E. G. (1991). *Research interviewing*. Cambridge, Mass.: Harvard University Press.

Munby, H., & Russell, T. (1993). The Authority of Experience in Learning To Teach: Messages from a Physics Methods Class. *Journal of Teacher Education*(March 1994 45 (2)), 86-95. doi: 10.1177/0022487194045002002

Munthe, E. (2003). Teachers' workplace and professional certainty. *Teaching and Teacher Education, 19*(8), 801-813.

Naeslund, L. (2004). *Prövostenar i praktiken: grundskolans nationella provsystem i ljuset av användares synpunkter*. Stockholm: Skolverket.

Nakatsuhara, F. (2008). Inter-interviewer variation in oral interview tests. *ELT Journal: English Language Teachers Journal, 62*(3), 266-275. doi: 10.1093/elt/ccm044

Nakatsuhara, F. (2011). Effects of test-taker characteristics and the number of participants in group oral tests. *Language Testing, 28*(4), 483-508. doi: 10.1177/0265532211398110

North, B. (1995). The development of a common framework scale of descriptors of language proficiency based on a theory of measurement. *System, 23*(4), 445-465.

North, B. (2014). *The CEFR in Practice* (Vol. 4). Cambridge: Cambridge University Press.

Ockey, G. J. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test scores. *Language Testing, 26*(2), 161-186. doi: 10.1177/0265532208101005

OECD. (2009). *Creating effective teaching and learning environments : first results from TALIS*. Paris: OECD.

OECD. (2010). *PISA 2009 results. Vol. 5, Learning trends : changes in student performance since 2000*. Paris: OECD.

Orr, M. (2002). The FCE speaking test: Using rater reports to help interpret test scores. *System, 30*(2), 143-154.

Oscarson, M., & Apelgren, B. M. (2005). *Nationella utvärderingen av grundskolan 2003 (NU-03) : engelska*. Stockholm: Skolverket.

Oscarson, M., & Apelgren, B. M. (2011). Mapping language teachers' conceptions of student assessment procedures in relation to grading: A two-stage empirical inquiry. *System, 39*(1), 2-16. doi: 10.1016/j.system.2011.01.014

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of educational research, 62*(3), 307-332.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, Calif.: Sage.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6*(02), 186-214.

Pienemann, M., Belay, L., Brown, B. B., Hinckley, P., Levi, A., & Weinstein, T. (1995). *Second Language Acquisition: A First Introduction*. Macarthur: NLLIA/LARC, University of Western Sydney Macarthur.

Pollitt, A., & Murray, N. L. (1996). What raters really pay attention to. *Studies in language testing, 3*, 74-91.

Pope, M., & Denicolo, P. (2000). *Transformative Education: Personal Construct Approaches Ot Practice and Research*. Hoboken, New Jersey: Wiley-Blackwell.

Prop. 2008/09:87. *Tydligare mål och kunskapskrav: Nya läroplaner för skolan*. Retrieved 2015-03-11 from www.regeringen.se/sb/d/10003/a/117269.

Puhan, G. (2008). Detecting and correcting scale drift in test equating: An illustration from a large scale testing program. *Applied Measurement in Education, 22*(1), 79-103.

Ragin, C. C., & Becker, H. S. (1992). *What is a case?: exploring the foundations of social inquiry*. Cambridge: Cambridge University Press.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*(3), 249-258.

Richards, J. C. (2001). *Curriculum Development in Language Teaching [Elektronisk resurs]*. Cambridge: Cambridge University Press.

Riksrevisonen. (2004). *Betyg med lika värde?: en granskning av statens insatser [Elektronisk resurs]* (Vol. 2004:11). Stockholm: Riksrevisionen.

Rothstein, J. (2008). Teacher quality in educational production: Tracking, decay, and student achievement. *Working Paper #25*. Princeton, New Jersey: Education Research Section, Princeton University and National Bureau of Economic Research.

Ruohotie-Lyhty, M. (2013). Struggling for a professional identity: Two newly qualified language teachers' identity narratives during the first years at work. *Teaching and Teacher Education, 30*, 120-129.

Sandlund, E., & Sundqvist, P. (2011). Managing task-related trouble in L2 oral proficiency tests: Contrasting interaction data and rater assessment. *Novitas-ROYAL, 5*(1), 91-120.

Saussure, F. d. (1970). *Kurs i allmän lingvistik*. Staffanstorp: Cavefors.

Savignon, S. J. (1983). *Communicative Competence: Theory and Classroom Practice. Texts and Contexts in Second Language Learning*. Reading, Mass.: Addison. Wesley.

Scheurich, J. J. (1995). A postmodernist critique of research interviewing. *International Journal of Qualitative Studies in Education, 8*(3), 239-252. doi: 10.1080/0951839950080303

Selghed, B. (2004). *Ännu icke godkänt : lärares sätt att erfara betygssystemet och dess tillämpning i yrkesutövningen*. (PhD), Malmö högskola, Malmö. (Malmö studies in educational sciences, 1651-4513 ; 15)

SFS 2010:800. *Skollag*/The Education Act. Stockholm: Utbildningsdepartementet.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 4-14.

Silverman, D. (2006). *Interpreting qualitative data : methods for analyzing talk, text and interaction*. London: Sage.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185-211. doi: 10.1177/136216889700100302

Skinner, B. F. (1957). *Verbal behavior*. London: Methuen & Co.

Skolinspektionen. (2010). *Tillsyn och kvalitetsgranskning : sammanfattning av Skolinspektionens erfarenheter och resultat*. Stockholm: Skolinspektionen.

Skolinspektionen. (2011). *Engelska i grundskolans årskurser 6-9 [Elektronisk resurs]*. Stockholm: Skolinspektionen.

Skolinspektionen. (2013). Olikheterna är för stora. Omrättning av nationella prov i grundskolan och gymnasieskolan, 2013. *Redovisning av regeringsuppdrag Dnr U2011/6544/GV*.

Skolverket. (2000a). *Grundskolan : kursplaner och betygskriterier*. Stockholm: Statens skolverk.

Skolverket. (2000b). *Grundskolan : kursplaner och betygskriterier. Kommentarer*. Stockholm: Statens skolverk.

Skolverket. (2001). *Språkboken: en antologi om språkundervisning och språkinlärning / [Elektronisk resurs]*. Stockholm: Myndigheten för skolutveckling.

Skolverket. (2004). *Engelska i åtta europeiska länder : en undersökning av ungdomars kunskaper och uppfattningar*. Stockholm: Skolverket.

Skolverket. (2005). *Att mäta skolors relativa effektivitet : en modellanalys baserad på resurser och resultat*. Stockholm: Skolverket.

Skolverket. (2007). *Provbetyg-slutbetyg-likvärdig bedömning? : en statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolan 1998-2006* (Vol. 300). Stockholm: Skolverket.

Skolverket. (2011a). *Kommentarmaterial till kursplanen i engelska*. Stockholm: Skolverket.

Skolverket. (2011b). *Kunskapsbedömning i skolan: praxis, begrepp, problem och möjligheter*. Stockholm: Skolverket.

Skolverket. (2011c). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011*. Stockholm: Skolverket.

Skolverket. (2012a). *Bedömning av språklig kompetens : en studie av samstämmigheten mellan Internationella språkstudien 2011 och svenska styrdokument*. Stockholm: Skolverket.

Skolverket. (2012b). *Internationella språkstudien 2011 : elevernas kunskaper i engelska och spanska*. Stockholm: Skolverket.

Skolverket. (2013). *TALIS 2013 : en studie av undervisnings- och lärmiljöer i årskurs 7-9 Internationella studier* (Vol. Rapport 408:2014). Stockholm: Skolverket.

Skolöverstyrelsen. (1962). *Läroplan för grundskolan*. Stockholm: Kungl. Skolöverstyr.

Skolöverstyrelsen. (1969a). *Läroplan för grundskolan : Lgr 69*. Stockholm: Utbildningsförl.

Skolöverstyrelsen. (1969b). *Läroplan för grundskolan. 2, Supplement : kompletterande anvisningar och kommentarer. Engelska*. Stockholm: Utbildningsförl.

Skolöverstyrelsen. (1980). *1980 års läroplan för grundskolan : inledning : mål och riktlinjer*. Stockholm: LiberFörlag/Allmänna förl.

Skolöverstyrelsen. (1990). *Läroplaner. 1990:1, Undervisning i främmande språk : kommentarmaterial*. Stockholm: LiberUtbildningsförl.

SOU 2010:96. *Riktiga betyg är bättre än höga betyg: förslag till omprövning av betyg : betänkande av betygsrgrprövningsutredningen*. Stockholm: Fritze.

Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt.

Spradley, J. P. (1980). *Participant observation*. New York: Holt, Rinehart and Winston.

Sundberg, D., & Wahlström, N. (2012). Standards-based curricula in a denationalised conception of education: the case of Sweden. *European*

Educational Research Journal, 11(3), 342-356. doi: 10.2304/eerj.2012.11.3.342

Sundqvist, P. (2009). *Extramural English matters: out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary* (PhD), Karlstad University, Karlstad. (2009:55)

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371-391.

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL, 24*(03), 302-321. doi: 10.1017/S095834401200016X

Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing, 17*(2), 163-184.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling a typology with examples. *Journal of Mixed Methods Research, 1*(1), 77-100. doi: 10.1177/2345678906292430

Tholin, J. (2006). *Att kunna klara sig i ökänd natur : en studie av betyg och betygskriterier - historiska betingelser och implementering av ett nytt system.* (PhD), Högskolan i Borås, Borås. Retrieved from http://hdl.handle.net/2320/1532 (Skrifter från Högskolan i Borås, 0280-381X ; 1)

Tornberg, U. (2005). *Språkdidaktik*. Malmö: Gleerup.

Utbildningsdepartementet. (1994). *Läroplaner för det obligatoriska skolväsendet och de frivilliga skolformerna : Lpo 94 : Lpf 94*. Stockholm: Utbildningsdepartementet.

Van den Akker, J., Kuiper, W., & Hameyer, U. (Eds.). (2003). *Curriculum perspectives: An introduction*. Dordrecht: Kluwer Academic Publishers.

Wedin, A.-S. (2007). *Lärares arbete och kunskapsbildning: utmaningar och inviter i den vardagliga praktiken.* (PhD), Linköping University, Linköping. (Linköping Studies in Pedagogic Practices 2)

Velling Pedersen, D. (2013). Ämnesprovet i engelska. I K. Hector-Stahre (Red.). In K. Hector-Stahre (Ed.), *Ämnesproven 2012 i grundskolans årskurs 9 och specialskolans årskurs 10*. Stockholm: Skolverket.

White, L. (1987). Against comprehensible input: the input hypothesis and the development of second-language competence. *Applied Linguistics, 8*(2), 95-110. doi: 10.1093/applin/8.2.95

Wiliam, D. (2007). Content then process: Teacher learning communities in the service of formative assessment. In D. B. Reeves (Ed.), *Ahead of the*

curve: The power of assessment to transform teaching and learning (pp. 183-204). Bloomington, Indiana: Solution Tree.

Vygotskij, L., & Cole, M. (1978). *Mind in society : the development of higher psychological processes*. Cambridge, Mass.: Harvard U.P.

Vygotskij, L., & Kozulin, A. (1986). *Thought and language*. Cambridge, Mass.: MIT Press.

Vygotskij, L. S., & Öberg Lindsten, K. (1995). *Fantasi och kreativitet i barndomen*. Göteborg: Daidalos.

Yin, R. K. (2009). *Case study research : design and methods*. London: Sage.

Young, B., & Tardif, C. (1992). Interviewing: Two sides of the story. *Qualitative Studies in Education, 5*(2), 135-145.

Åberg-Bengtsson, L., & Erickson, G. (2006). Dimensions of national test performance: A two-level approach. *Educational Research & Evaluation, 12*(5), 469-488. doi: 10.1080/13803610600808115

Östh, J., Andersson, E., & Malmberg, B. (2013). School choice and increasing performance difference: A counterfactual approach. *Urban Studies, 50*(2), 407-425.

Appendix A: Syllabus for Enlgish

Extract from National Curriculum Lgr11

3.2 ENGLISH

Language is the primary tool human beings use for thinking, communicating and learning. Having a knowledge of several languages can provide new perspectives on the surrounding world, enhanced opportunities to create contacts and greater understanding of different ways of living. The English language surrounds us in our daily lives and is used in such diverse areas as politics, education and economics. Knowledge of English thus increases the individual's opportunities to participate in different social and cultural contexts, as well as in international studies and working life.

Aim

Teaching of English should aim at helping the pupils to develop knowledge of the English language and of the areas and contexts where English is used, and also pupils' confidence in their ability to use the language in different situations and for different purposes.

Through teaching, pupils should be given the opportunity to develop all-round communicative skills. These skills involve understanding spoken and written English, being able to formulate one's thinking and interact with others in the spoken and written language, and the ability to adapt use of language to different situations, purposes and recipients. Communication skills also cover confidence in using the language and the ability to use different strategies to support communication and solve problems when language skills by themselves are not sufficient.

In order to deal with spoken language and texts, pupils should be given the opportunity to develop their skills in relating content to their own experiences, living conditions and interests. Teaching should also provide pupils with opportunities to develop knowledge about and an understanding of different living conditions, as well as social and cultural phenomena in the areas and contexts where English is used.

Teaching should help pupils to develop their skills in searching for, evaluating, choosing and assimilating the content of spoken language and texts from different sources. They should also be equipped to be able to use different tools for learning, understanding, being creative and communicating. Teaching should encourage pupils to develop an interest in languages and culture, and convey the benefits of language skills and knowledge.

Teaching in English should essentially give pupils the opportunities to develop their ability to:

- understand and interpret the content of spoken English and in different types of texts,
- express themselves and communicate in speech and writing,
- use language strategies to understand and make themselves understood,
- adapt language for different purposes, recipients and contexts, and
- reflect over living conditions, social and cultural phenomena in different contexts and parts of the world where English is used.

Core content

In years 1–3

Content of communication

- Subject areas that are familiar to the pupils.
- Interests, people and places.
- Daily life and ways of living in different contexts and areas where English is used.

Listening and reading – reception

- Clearly spoken English and texts from various media.
- Simple instructions and descriptions.
- Different types of simple conversations and dialogues.
- Films and dramatised narratives for children.
- Songs, rhymes, poems and sagas.
- Words and phrases in their local surroundings, such as those used on signs and other simple texts.

Speaking, writing and discussing – production and interaction

- Simple presentations.
- Simple descriptions and messages.
- Songs, rhymes and dramatisations.

In years 4–6

Content of communication

- Subject areas that are familiar to the pupils.
- Daily situations, interests, people, places, events and activities.
- Views, feelings and experiences.
- Daily life, ways of living and social relations in different contexts and areas where English is used.

Listening and reading – reception

- Clearly spoken English and texts from various media.
- Oral and written instructions and descriptions.
- Different types of conversations, dialogues and interviews.
- Films and dramatised narratives for children and youth.
- Songs, sagas and poems.
- Strategies to understand key words and context in spoken language and texts, for example, by adapting listening and reading to the form and content of communications.

- Different ways of searching for and choosing texts and spoken English from the Internet and other media.
- Language phenomena such as pronunciation, intonation, grammatical structures, spelling and also fixed language expressions in the language pupils encounter.
- How words and fixed language expressions, such as politeness phrases and forms of address, are used in texts and spoken language in different situations.
- How different expressions are used to initiate and complete different types of communications and conversations.

Speaking, writing and discussing – production and interaction

- Presentations, instructions, messages, narratives and descriptions in connected speech and writing.
- Language strategies to understand and make oneself understood when language skills are lacking, such as through reformulations.
- Language strategies to participate in and contribute to discussions, such as questions, and phrases and expressions to confirm understanding.
- Language phenomena to clarify and enrich communication such as pronunciation and intonation, spelling and punctuation, polite phrases, and other fixed language expressions and grammatical structures.

In years 7–9

Content of communication

- Current and subject areas familiar to the pupils.
- Interests, daily situations, activities, sequences of events, relations and ethical questions.
- Views, experiences, feelings and future plans.
- Living conditions, traditions, social relations and cultural phenomena in various contexts and areas where English is used.

Listening and reading – reception

- Spoken English and texts from various media.
- Spoken English with some regional and social variants.
- Oral and written instructions and descriptions.
- Different types of conversations, dialogues, interviews and oral communications.
- Literature and other fiction in spoken, dramatised and filmed forms.
- Songs and poems.
- Oral and written information, as well as discussions and argumentation for different purposes, such as news, reports and newspaper articles.

- Strategies to understand details and context in spoken language and texts, such as adapting listening and reading to the type of communication, contents and purpose.
- Different ways of searching for, choosing and assessing texts and spoken language in English from the Internet and other media.
- Language phenomena such as pronunciation, intonation, grammatical structures, sentence structure, words with different registers, as well as fixed language expressions pupils will encounter in the language.
- How texts and spoken language can be varied for different purposes and contexts.
- How connecting words and other expressions are used to create structure and linguistically coherent entities.

Speaking, writing and discussing – production and interaction

- Different ways of working on personal communications to vary, clarify, specify and adapt them for different purposes.
- Oral and written narratives, descriptions and instructions.
- Conversations, discussions and argumentation.
- Language strategies to understand and be understood when language skills are lacking, such as reformulations, questions and explanations.
- Language strategies to contribute to and actively participate in conversations by taking the initiative in interaction, giving confirmation, putting follow-up questions, taking the initiative to raise new issues and also concluding conversations.
- Language phenomena to clarify, vary and enrich communication such as pronunciation, intonation and fixed language expressions, grammatical structures and sentence structures.

Knowledge requirements

Knowledge requirements for grade E at the end of year 6

Pupils can understand **the most essential content** in clearly spoken, simple English at a relaxed pace in simple texts about daily and familiar topics. Pupils show their understanding by reporting content **in a simple form** with comments on content and also with **acceptable** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and texts, pupils can choose and apply **a strategy** for listening and reading. Pupils can choose texts and spoken language of a simple nature and from different media and **with some relevance** use the selected material in their own production and interaction.

In oral and written production, pupils can express themselves **simply and understandably in phrases and sentences**. To clarify and vary their communication, pupils can work on and make some **simple** improvements to their communications. In oral and written interaction, pupils can express themselves **simply**

and understandably in words, phrases and sentences. In addition, pupils can choose and use **a strategy** that solves problems and improves their interaction.

Pupils comment **in simple forms** on some phenomena in different contexts and areas where English is used, and can also make simple comparisons with their own experiences and knowledge.

Knowledge requirements for grade D at the end of year 6

Grade D means that the knowledge requirements for grade E and most of C are satisfied.

Knowledge requirements for grade C at the end of year 6

Pupils can understand **the main content and clear details** in simple English, clearly spoken at a relaxed pace, and also in simple texts on daily and familiar topics. Pupils show their understanding by reporting content **in a simple form** with comments on content **and details** and also with **satisfactory** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply **strategies** for listening and reading. Pupils can choose from texts and spoken language of a simple nature and from different media and **in a relevant way** use the selected material in their own production and interaction.

In oral and written production, pupils can express themselves **simply, relatively clearly and to some extent coherently**. To clarify and vary their communication, pupils can work on and make **simple** improvements to their communications. In oral and written interaction, pupils can express themselves **simply and relatively clearly in words, phrases and sentences**. In addition, pupils can choose and apply **some different strategies** to solve problems and improve their interaction.

Pupils comment **in simple forms** on some phenomena in different contexts and areas where English is used, and can also make simple comparisons with their own experiences and knowledge.

Knowledge requirements for grade B at the end of year 6

Grade B means that the knowledge requirements for grade C and most of A are satisfied.

Knowledge requirements for grade A at the end of year 6

Pupils can understand **the whole and important details** in clearly spoken, simple English at a relaxed pace in simple texts on daily and familiar topics. Pupils show their understanding by presenting an **overview** with their comments on content **and details** and also with **good** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply **strategies** for listening and reading. Pupils can choose from texts and spoken

language of a simple nature and from different media and **in a relevant and effective way** use the material chosen in their own production and interaction.

In oral and written production, pupils can express themselves simply, relatively clearly and relatively coherently. To clarify and vary their communication, pupils can work on and make simple improvements to their communications. In oral and written interaction, pupils can express themselves **simply and clearly in words, phrases and sentences, which to some extent are adapted to purpose, recipient and situation**. In addition, pupils can choose and apply **several different strategies** to solve problems and improve their interaction.

Pupils comment **in overall terms** on some phenomena in different contexts and areas where English is used, and can also make simple comparisons with their own experiences and knowledge.

Knowledge requirements for grade E at the end of year 9

Pupils can understand **the main content and basic details** in English spoken at a moderate pace and in basic texts in various genres. Pupils show their understanding by presenting an **overview** with discussion and comments on content and details and also with **acceptable** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and texts, pupils can choose and apply **a strategy** for listening and reading. Pupils can choose texts and spoken language from different media and with some relevance use the selected material in their own production and interaction.

In oral and written production, pupils can express themselves **simply, understandably and relatively coherently**. To clarify and vary their communication, pupils can work on and make **simple** improvements to their communications. In oral and written interaction in different contexts, pupils can express themselves **simply and understandably** and also **to some extent** adapted to purpose, recipient and situation. In addition, pupils can choose and apply **basically functional** strategies which **to some extent** solve problems and improve their interaction.

Pupils discuss **in overall terms** some phenomena in different contexts and areas where English is used, and can also make **simple** comparisons with their own experiences and knowledge.

Knowledge requirements for grade D at the end of year 9

Grade D means that the knowledge requirements for grade E and most of C are satisfied.

Knowledge requirements for grade C at the end of year 9

Pupils can understand **the main content and essential details** in English spoken at a moderate pace and in basic texts in various genres. Pupils show their understanding by presenting a **well grounded** account with discussion on content and details and also with **satisfactory results** act on the basis of the message and

instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply **strategies** for listening and reading. Pupils can choose texts and spoken language from different media and **in a relevant way** use the selected material in their own production and interaction.

In oral and written production, pupils can express themselves in **relatively** varied ways, relatively clearly and relatively coherently. Pupils express themselves also **with some ease and to some extent adapted to purpose, recipient and situation**. To clarify and vary their communication, pupils can work on and make **well grounded** improvements to their own communications. In oral and written interaction in different contexts, pupils can express themselves **clearly and with some ease** and **with some adaptation** to purpose, recipient and situation. In addition, pupils can choose and use **functional** strategies to **solve problems** and improve their interaction.

Pupils discuss in detail some phenomena in different contexts and areas where English is used, and can then also make well developed comparisons with their own experiences and knowledge.

Knowledge requirements for grade B at the end of year 9

Grade B means that the knowledge requirements for grade C and most of A are satisfied.

Knowledge requirements for grade A at the end of year 9

Pupils can understand **both the whole and the details** in English spoken at a moderate pace in ordinary texts in various genres. Pupils show their understanding by giving a **well grounded and balanced** account where they discuss and comment on content and details, and with **good** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply strategies for listening and reading. Pupils can choose texts and spoken language from different media and **in a relevant and effective way** use the material chosen in their own production and interaction.

In oral and written production, pupils can express themselves in **relatively varied ways, clearly and coherently. Pupils express themselves with ease and some adaptation to purpose, recipient and situation**. To clarify and vary their communication, pupils can work on and make **well grounded** improvements to their own communications. In oral and written interaction in different contexts, pupils can express themselves **clearly and with ease**, and also **with some adaptation** to purpose, recipient and situation. In addition, pupils can choose and apply **well functioning** strategies to **solve problems** and improve their interaction **and take it forward in a constructive way**.

Pupils discuss **in detail and in a balanced way** some phenomena from different contexts and areas where English is used, and can also make **well developed and balanced** comparisons with their own experiences and knowledge.

Knowledge requirements at the end of year 6

Grade E	Grade D	Grade C	Grade B	Grade A
Pupils can understand **the most essential content** in clearly spoken, simple English at a relaxed pace in simple texts about daily and familiar topics. Pupils show their understanding by reporting content **in a simple form** with comments on content and also with **acceptable** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and texts, pupils can choose and apply **a strategy** for listening and reading. Pupils can choose texts and spoken language of a simple nature and from different media and **with some relevance use** the selected material in their own production and interaction. In oral and written production, pupils can express themselves **simply and understandably in phrases and sentences**. To clarify and vary their communication, pupils can work on and make some **simple** improvements to their communications. In oral and written interaction, pupils can express themselves **simply and understandably in words, phrases and sentences**. In addition, pupils can choose and use **a strategy** that solves problems and improves their interaction.	Grade D means that the knowledge requirements for grade E and most of C are satisfied.	Pupils can understand **the main content and clear details** in simple English, clearly spoken at a relaxed pace, and also in simple texts on daily and familiar topics. Pupils show their understanding by reporting content **in a simple form** with comments on content **and details** and also with **satisfactory** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply **strategies** for listening and reading. Pupils can choose from texts and spoken language of a simple nature and from different media and **in a relevant way** use the selected material in their own production and interaction. In oral and written production, pupils can express themselves **simply, relatively clearly and to some extent coherently**. To clarify and vary their communication, pupils can work on and make **simple** improvements to their communications. In oral and written interaction, pupils can express themselves **simply and relatively clearly in words, phrases and sentences**. In addition, pupils can choose and apply **some different strategies** to solve problems and improve their interaction.	Grade B means that the knowledge requirements for grade C and most of A are satisfied.	Pupils can understand **the whole and important details** in clearly spoken, simple English at a relaxed pace in simple texts on daily and familiar topics. Pupils show their understanding by presenting an **overview** with their comments on content **and details** and also with **good** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply **strategies** for listening and reading. Pupils can choose from texts and spoken language of a simple nature and from different media and **in a relevant and effective way** use the material chosen in their own production and interaction. In oral and written production, pupils can express themselves simply, relatively clearly and relatively coherently. To clarify and vary their communication, pupils can work on and make simple improvements to their communications. In oral and written interaction, pupils can express themselves **simply and clearly in words, phrases and sentences, which to some extent are adapted to purpose, recipient and situation**. In addition, pupils can choose and apply **several different strategies** to solve problems and improve their interaction.

Knowledge requirements at the end of year 6

Grade E	Grade D	Grade C	Grade B	Grade A
Pupils comment **in simple forms** on some phenomena in different contexts and areas where English is used, and can also make simple comparisons with their own experiences and knowledge.	Grade D means that the knowledge requirements for grade E and most of C are satisfied.	Pupils comment **in simple forms** on some phenomena in different contexts and areas where English is used, and can also make simple comparisons with their own experiences and knowledge.	Grade B means that the knowledge requirements for grade C and most of A are satisfied.	Pupils comment **in overall terms** on some phenomena in different contexts and areas where English is used, and can also make simple comparisons with their own experiences and knowledge.

Knowledge requirements at the end of year 9

Grade E	Grade D	Grade C	Grade B	Grade A
Pupils can understand the main content and basic details in English spoken at a moderate pace and in basic texts in various genres. Pupils show their understanding by presenting an overview with discussion and comments on content and details and also with acceptable results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and texts, pupils can choose and apply a strategy for listening and reading. Pupils can choose texts and spoken language from different media and with some relevance use the selected material in their own production and interaction. In oral and written production, pupils can express themselves simply, understandably and relatively coherently. To clarify and vary their communication, pupils can work on and make simple improvements to their communications. In oral and written interaction in different contexts, pupils can express themselves simply and understandably and also to some extent adapted to purpose, recipient and situation. In addition, pupils can choose and apply **basically functional** strategies which **to some extent** solve problems and improve their interaction.	Grade D means that the knowledge requirements for grade E and most of C are satisfied.	Pupils can understand **the main content and essential details** in English spoken at a moderate pace and in basic texts in various genres. Pupils show their understanding by presenting a **well grounded** account with discussion on content and details and also with **satisfactory results** act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply **strategies** for listening and reading. Pupils can choose texts and spoken language from different media and **in a relevant way** use the selected material in their own production and interaction. In oral and written production, pupils can express themselves in **relatively** varied ways, relatively clearly and relatively coherently. Pupils express themselves also **with some ease and to some extent adapted to purpose, recipient and situation**. To clarify and vary their communication, pupils can work on and make **well grounded** improvements to their own communications. In oral and written interaction in different contexts, pupils can express themselves **clearly and with some ease** and **with some adaptation** to purpose, recipient and situation. In	Grade B means that the knowledge requirements for grade C and most of A are satisfied.	Pupils can understand **both the whole and the details** in English spoken at a moderate pace in ordinary texts in various genres. Pupils show their understanding by giving a **well grounded and balanced** account where they discuss and comment on content and details, and with **good** results act on the basis of the message and instructions in the content. To facilitate their understanding of the content of the spoken language and the texts, pupils can **to some extent** choose and apply strategies for listening and reading. Pupils can choose texts and spoken language from different media and **in a relevant and effective way** use the material chosen in their own production and interaction. In oral and written production, pupils can express themselves in **relatively varied ways, clearly and coherently. Pupils express themselves with ease and some adaptation to purpose, recipient and situation.** To clarify and vary their communication, pupils can work on and make **well grounded** improvements to their own communications. In oral and written interaction in different contexts, pupils can express themselves **clearly and with ease**, and also **with some adaptation** to purpose, recipient and situation. In addition, pupils can choose and apply **well**

Knowledge requirements at the end of year 9

Grade E	Grade D	Grade C	Grade B	Grade A
	Grade D means that the knowledge requirements for grade E and most of C are satisfied.	addition, pupils can choose and use **functional** strategies to **solve problems** and improve their interaction.	Grade B means that the knowledge requirements for grade C and most of A are satisfied.	**functioning** strategies to **solve problems** and improve their interaction **and take it forward in a constructive way**.
Pupils discuss **in overall terms** some phenomena in different contexts and areas where English is used, and can also make **simple** comparisons with their own experiences and knowledge.		Pupils discuss **in detail** some phenomena in different contexts and areas where English is used, and can then also make **well developed** comparisons with their own experiences and knowledge.		Pupils discuss **in detail and in a balanced way** some phenomena from different contexts and areas where English is used, and can also make **well developed and balanced** comparisons with their own experiences and knowledge.

Appendix B: Assessment Guideline for Part A

Extract from Teacher Information for National Test in English, Year 2012/2013

Översikt: Bedömning Delprov A

Ämnets syfte

Detta delprov relaterar framför allt till fyra av de långsiktiga målen i kursplanens syftestext:

Eleverna ska ges förutsättningar att utveckla sin förmåga att

- förstå och tolka innehållet i talad engelska …
- formulera sig och kommunicera i tal …
- använda språkliga strategier för att förstå och göra sig förstådda
- anpassa språket efter olika syften, mottagare och sammanhang

Bedömning av muntlig produktion och interaktion

Bedömningen av muntlig språkfärdighet utgår från att eleven, med utgångspunkt i den givna uppgiften, baserad på kursplanen, vill och kan uttrycka och utveckla ett innehåll, på egen hand och i samspel med andra.

Vidstående bedömningsfaktorer bygger på den kommunikativa och handlingsorienterade språksyn som ligger till grund för de svenska kurs- och ämnesplanerna i engelska och moderna språk. Faktorerna är avsedda att vara ett stöd för analysen vid en helhetsbedömning och ska ses som olika aspekter av kvaliteter i talat språk.

Bedömningsfaktorer

Innehåll

- begriplighet och tydlighet
- fyllighet och variation
  - olika exempel och perspektiv
- sammanhang och struktur
- anpassning till syfte, mottagare och situation

Språk och uttrycksförmåga

- kommunikativa strategier
  - för att utveckla och föra samtal vidare
  - för att lösa språkliga problem genom t.ex. omformuleringar, förklaringar och förtydliganden
- flyt och ledighet
- omfång, variation, tydlighet och säkerhet
  - vokabulär, fraseologi och idiomatik
  - uttal och intonation
  - grammatiska strukturer
- anpassning till syfte, mottagare och situation

Vid betygsättningen av delprov A relateras framför allt till följande delar av kunskapskraven, som särskilt betonar muntlig produktion och interaktion.

Kunskapskrav

Betyget E
I muntliga … framställningar i olika genrer kan eleven formulera sig **enkelt, begripligt och relativt sammanhängande**.

För att förtydliga och variera sin kommunikation kan eleven … göra **enkla** förbättringar av egna framställningar.*

I muntlig … interaktion i olika sammanhang kan eleven uttrycka sig **enkelt och begripligt** samt **i någon mån anpassat** till syfte, mottagare och situation.

Dessutom kan eleven välja och använda sig av **i huvudsak fungerande** strategier som **i viss mån** löser problem i och förbättrar interaktionen.*

Betyget C
I muntliga … framställningar i olika genrer kan eleven formulera sig **relativt varierat, relativt tydligt och relativt sammanhängande. Eleven formulerar sig även med visst flyt och i någon mån anpassat till syfte, mottagare och situation**.

För att förtydliga och variera sin kommunikation kan eleven bearbeta och göra **välgrundade** förbättringar av egna framställningar.*

I muntlig … interaktion i olika sammanhang kan eleven uttrycka sig **tydligt och med visst flyt** samt **med viss anpassning** till syfte, mottagare och situation.

Dessutom kan eleven välja och använda sig av **fungerande** strategier som **löser problem** i och förbättrar interaktionen.*

Betyget A
I muntliga … framställningar i olika genrer kan eleven formulera sig **relativt varierat, tydligt och sammanhängande. Eleven formulerar sig även med flyt och viss anpassning till syfte, mottagare och situation**.

För att förtydliga och variera sin kommunikation kan eleven bearbeta och göra **välgrundade** förbättringar av egna framställningar.*

I muntlig … interaktion i olika sammanhang kan eleven uttrycka sig **tydligt och med flyt** samt **med viss anpassning** till syfte, mottagare och situation.

Dessutom kan eleven välja och använda sig av **väl fungerande** strategier som **löser problem** i och förbättrar interaktionen **och för den framåt på ett konstruktivt sätt**.*

Betyget D
Kunskapskraven för betyget E och till övervägande del för C är uppfyllda.

Betyget B
Kunskapskraven för betyget C och till övervägande del för A är uppfyllda.

* Fokuseras inte specifikt, men delprovet ger möjlighet till bedömning av denna förmåga.

Appendix C: Instructions for the Test & Warm-up

People's Choices

You are going to talk about choices you and other people make – and why.

Be active and speak English all the time. Help each other with questions and comments to keep the conversation going.

Warm-up

Take turns to tell each other about some things you – and perhaps your family – often do.

Which of these things do you …

- really **want to** do?
- really **have to** do?

Part One

(*mind map + yellow cards*)

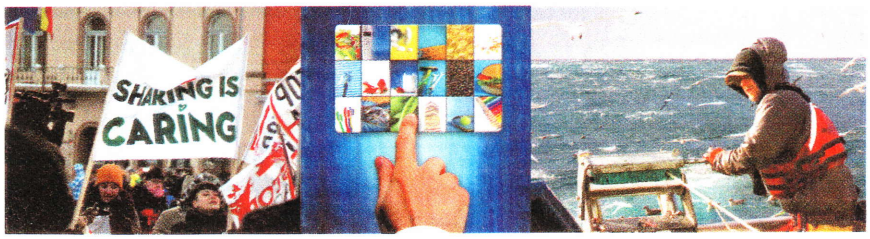
You are going to discuss some things that may influence people and be important when they make their choices.

Take turns to pick a yellow card. Read what it says on your card and put it on the table. The mind map can help you in your discussions.

Part Two

(*green cards*)

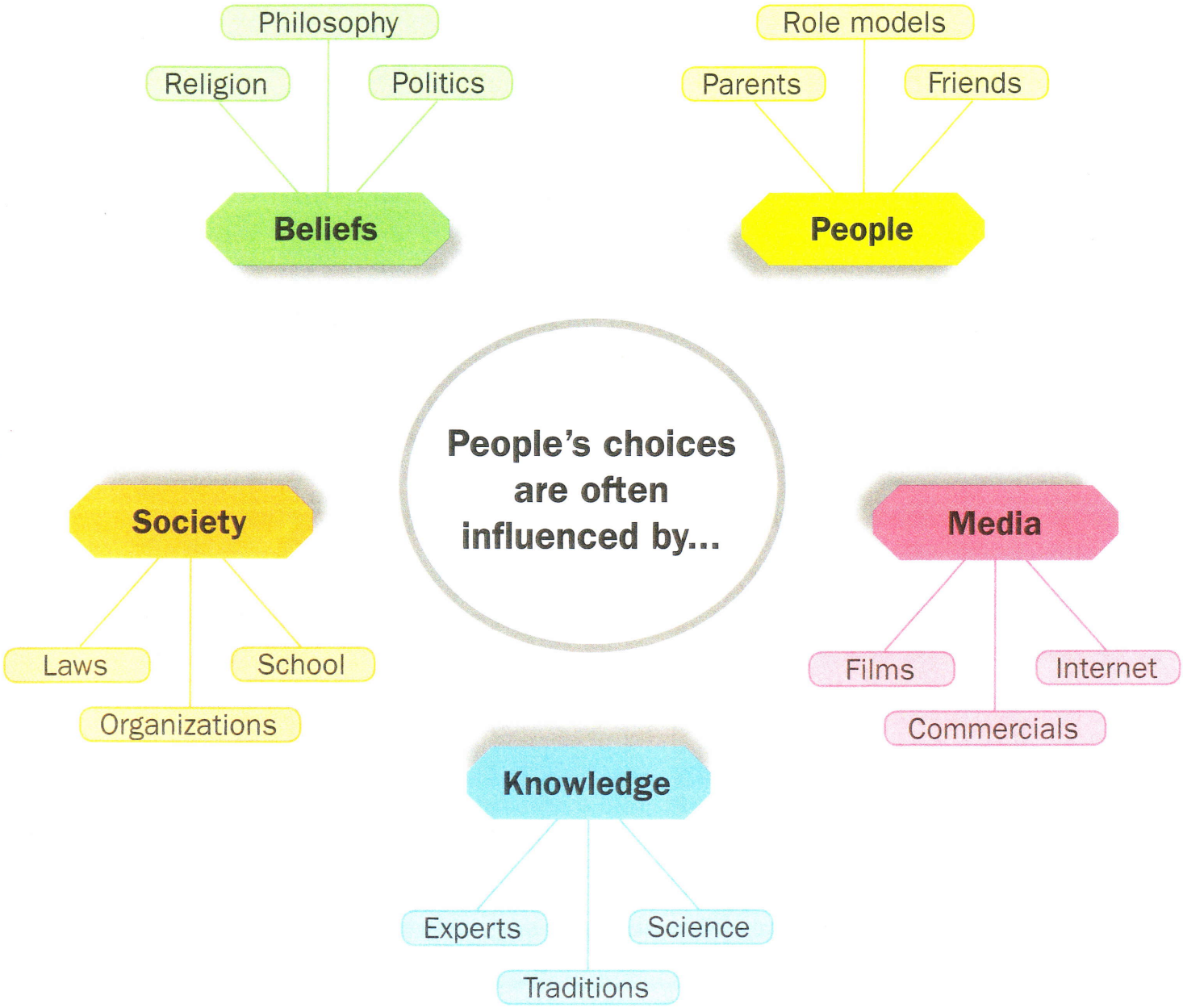
You are going to discuss some statements. Take turns to pick a green card. Read what it says on your card and put it on the table.



All photos: Shutterstock.com

Appendix D: Part 1 of the Test

Part One


Philosophy
Religion
Politics
Beliefs
Role models
Parents
Friends
People
People's choices are often influenced by...
Society
Laws
School
Organizations
Media
Films
Internet
Commercials
Knowledge
Experts
Science
Traditions

Appendix E: Part 1, cards

You need grades in school to motivate students to study

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

Gun laws should be stricter in all countries

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

People listen more to celebrities than to politicians

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

The picture of the USA that you get in TV-series/films is not correct

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

Schools should use more time to teach about alcohol and other drugs

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

I try to think of the environment in everything I do

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

Health experts are right when they say we should eat less carbohydrates

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

Media give you the wrong picture of life in many countries

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

Computer games make kids become more violent

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

You should always do as your parents say

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

People are starting to forget World War II

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

The Internet is a good tool for people who want to change things in the world

Agree? Disagree?
Explain why and give examples.
Discuss with your friend.

What you do on your holiday is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What music you like is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What pets you have is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What traditions you celebrate is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What clothes you choose is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What organizations you join is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What you do online is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What you do in your spare time is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What products you buy is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What food you eat is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

What education/school you choose is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

Where you live is often influenced by...

Explain what you think and give examples.
What about other people?
Discuss with your friend.

Appendix F: Information for Principals

GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Förutsättningar för bedömning av muntlig språkfärdighet

Studiens syfte
I en studie som syftar till en licentiatexamen undersöks frågor kring bedömning av muntlig kompetens, ett tema som är högst aktuella i denna brytpunkt mellan två läroplaner, betygsskalor och utökade nationella prov. Syftet är att undersöka *uppfattningar om bedömningskriterier, bedömningspraxis* och *förutsättningarna för bedömning* bland engelsklärare i åk 9.

Studiens uppläggning och genomförande
Studien genomförs i tre delar;

1) Moment I (sammanlagt 8-10 lärare från olika skolor): Lärarna intervjuas individuellt. Intervjun tar högst 60 min. Frågorna handlar om utbildningsbakgrund, undervisningserfarenhet och bedömar-erfarenhet när det gäller det nationella provet i engelska.

2) Moment II (8-10 lärare): Lärarna samlas i grupp med lärare från olika skolor och diskuterar och bedömer inspelade elevexempel. Denna diskussion och sambedömning spelas in och analyseras senare.

3) Moment III (8-10 lärare): Lärarna intervjuas individuellt som en uppföljning av den första intervjun och sambedömningstillfället. Lärarna besvarar frågor om sambedömningen i grupp, om bedömningen av de egna klassernas nationella prov och den egna bedömningspraktiken.

Det är viktigt att poängtera att allt deltagande sker på frivillig basis och att vem som helst kan avbryta sitt deltagande i studien när som helst utan motivering. Skolor och deltagare anonymiseras och kodas under bearbetningen och analysen av data. Om du har några frågor, är du välkommen att ställa dem via e-post: maria.frisch@gu.se.

Göteborg, februari 2013
Maria Frisch
Forskarskolan FRAM

Appendix G: Information for Teachers

GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Förutsättningar för bedömning av muntlig språkfärdighet

Studiens syfte
I denna studie undersöks frågor kring bedömning av muntlig språkfärdighet i det nationella provet i engelska i åk 9, en aktuell och viktig fråga i skolan idag. Syftet är att undersöka *uppfattningar om bedömningskriterier, bedömningspraxis* och *förutsättningarna för bedömning* bland engelsklärare som undervisar i år 9. I studien undersöks bakgrundsfaktorer som erfarenhet av samt utbildning i bedömning och betygssättning. Vidare undersöks hur lärare i grupp sambedömer elevexempel.

Studiens uppläggning och genomförande
1) Moment I (sammanlagt 12-15 lärare på olika skolor i landet): En individuell intervju med varje lärare. Intervjun kommer att ta ca 1 timme och handla om lärarens utbildning, erfarenheter och bedömningspraxis. Moment I genomförs under våren 2013. Tider för intervju bestäms individuellt med dig som lärare.

2) Moment II (Samma lärare): Sambedömning. Lärare samlas i grupp med lärare från två eller flera skolor och diskuterar inspelade elevexempel. Denna diskussion och sambedömning spelas in och analyseras senare. Moment II genomförs i början av höstterminen 2013.

3) Moment III (5-6 lärare): En individuell intervju, en uppföljningsintervju på ca 1 timme med ett mindre antal lärare. Lärarna besvarar frågor om sambedömningen av elevexempel i grupp, om bedömning av elevers muntliga språkfärdighet i allmänhet samt beskriver sin bedömningspraktik i engelska. Moment III genomförs under hösten 2013. Tider för intervju bestäms individuellt med dig som lärare.

Urval och frivilligt deltagande
Allt deltagande sker på frivillig basis. Skolor och deltagare anonymiseras och kodas under bearbetning och analys av data. Du kommer alltså att förbli anonym, liksom din skola.
Detta betyder att ditt svar inte kommer gå att urskilja, inga namn eller uppgifter som kan identifiera dig eller din skola kommer att finnas med. Intervjumaterial kommer att hanteras med stor försiktighet och sekretess.

Appendix G: Information for Teachers

Möjlighet till reflektion
Hela undersökningen kan ses som ett led i fortbildning kring bedömning och betygssättning. Lärare som deltagit i liknande studier vittnar om att frågorna sätter igång en reflektionsprocess. Intervjun ger dig möjlighet att sätta ord på dina erfarenheter genom att beskriva och förklara olika val du gör i bedömningsarbetet. Diskussionen och sambedömningen av elevexempel ger dig tid och möjlighet att diskutera bedömning av specifika språkliga förmågor med kollegor.

Jag hoppas att du som deltar skall uppleva att medverkan i studien ger dig stöd i ditt viktiga arbete som engelsklärare och i det bedömningsuppdrag som vi lärare i Sverige har. Jag hoppas också att ditt deltagande skall öppna för nya diskussioner mellan dig och dina kollegor.

Om du har några frågor, är du välkommen att ställa dem via e-post: maria.frisch@gu.se.

Göteborg, i mars 2013
Maria Frisch
Forskarskolan FRAM
Institutionen för pedagogik och specialpedagogik
Göteborgs universitet
maria.frisch@gu.se
070 662 83 89

Appendix H

Informants' backgrounds and experience

Description	#
Female	10
Male	2
Age 30+	4
Age 40+	4
Age 50+	4
Bilingual Swedish – English	1
Lived for more than 5 years in English speaking country	2
Bilingual Swedish – other language than English	2
Studied to become a teacher as an adult	4
Years of professional experience as a teacher	5-35
Teaches English only	1
Teaches English and Swedish	5
Teaches English and one other foreign language	3
Teaches English and two other foreign languages	1
Teaches English and Civics	2

Appendix I: Interview guide for the first interview

Intervju nr. 1 - frågor

Introduktion

- Presentation av intervjuaren
- Presentation av studien
- Diskussion/samtal ang. *written consent*

Det muntliga provet

- Vilka är dina erfarenheter av de muntliga proven i ämnesprovet i engelska hittills?
- Vad tycker du om de muntliga proven i engelska?
- Hur upplever du elevernas möjligheter att verkligen visa vad de kan muntligt på de nationella proven?

Provorganisation

- Hur organiserar ni de muntliga proven i de olika ämnena på din skola?
- Vem ansvarar för arrangemanget?
- Hur förbereder du proven?
- Hur använder du exempel-Cd:n som följer med provet?
- Spelar ni in proven? Varför? Varför inte?
- Vem genomför proven?
- Hur sätts paren/grupperna ihop?
- Vilka anpassningar gör du/ni på din skola för elever som ni bedömer behöver det?
- Hur gör du när du bedömer en muntlig prestation?

Kursplan och styrdokument

- Vad för fortbildning och information har du fått om den nya läroplanen, den nya kursplanen och de nya betygen?
- Hur skulle du beskriva den nya kursplanen i engelska?
- Hur uppfattar du målen för muntlig språkfärdighet i kursplanen?
- Hur ser du på de olika färdigheterna i engelska?
- Vilken vikt lägger du på de olika färdigheterna?

Egen undervisning

- Vilka faktorer är viktigast för dig när du undervisar i engelska?
- Vad för sorts muntliga aktiviteter arbetar ni med i klassrummet?
- Bedömningen av den muntliga språkfärdigheten i klassrumssituationen, hur går den till i din undervisning? Vilken feedback får eleverna?
- Vilka arbetsformer är vanligast när ni arbetar med muntlig färdighet?
- Har din engelskundervisning förändrats över tid? Hur?

Appendix I: Interview guide for the first interview

- Har din inställning till engelskundervisning förändrats?
- Har din inställning till muntlig språkfärdighet förändrats?

Muntlig språkfärdighet

- Vad utgör en god muntlig språkfärdighet enligt dig? Vad betyder det att kunna prata engelska?
    - *Hur viktiga tycker du intonation, uttal och betoning är för den kommunikativa förmågan? Hur viktigt tycker du det är att eleven talar grammatiskt riktigt? Vad betyder det att anpassa talet efter situationen för dig? Bedömer du att dina elever känner till idiomatiska uttryck?*

Bedömning

- Hur går du tillväga när du bedömer den enskilda elevens prestation på den muntliga delen av NP engelska?
    - *Gör du en helhetsbedömning först och tittar på enskildheterna sedan, eller bedömer du först enskilda moment, för att sedan göra en sammanvägning? Vilka enskildheter bedömer du/fokuserar du på? (Begriplighet och tydlighet, fyllighet och variation, sammanhang och struktur, anpassning, använda strategier, flyt, säkerhet, uttal och intonation, betoning, grammatisk korrekthet samspelet med de andra i gruppen…)*
- Vad har du fått för utbildning för att bedöma muntlig språkfärdighet?
    - *Vilken sorts stöd och hjälp skulle du vilja ha när det gäller att bedöma elevers muntliga språkfärdighet? Vad för sorts fortbildning skulle du vilja ha för att känna dig mer säker i bedömningssituationen? Har du blivit erbjuden fortbildning, men tackat nej?*
- Diskuterar du bedömningarna med dina elever?
- Diskuterar du dina bedömningar med dina kollegor?
- Hur många elever, ungefär, uppfyller inte kunskapskraven för ett godkänt betyg (E) på den muntliga delen av ämnesprovet i engelska varje år?
    - *Hur hanterar du det? Hur hanterar ni det på skolan?*
- Enligt din åsikt, vet eleverna vad som förväntas av dem under den muntliga delen av provet? Vet de vilka kriterier de behöver uppnå?
- Har din bedömning av elevers kunskaper och färdigheter förändrats över tid? Hur?

Bakgrund

- Vad har du för utbildning?
- Hur länge har du undervisat i engelska?
- Vilka årskurser undervisar du nu?
- Hur länge har du undervisat 9:or?
- Hur länge har du undervisat på den här skolan?
- Vad tycker du om att undervisa i engelska? Har din uppfattning om detta förändrats över tid?

Appendix J: Interview guide for second interview

Intervju nr. 2 - SAMTAL

Introduktion

Ingen intervju utan ett samtal där vi talar med varandra. Jag är mycket intresserad av att höra hur du ser på muntlig språkfärdighet ur olika aspekter.

Samtalspunkter

- Berätta lite om gruppdiskussionen i september!
    - Togs aspekter/bedömningskriterier upp som du reagerade på?
    - Hur rättvisa tycker du att bedömningarna blir?
- Berätta om hur du bedömer muntlig färdighet!
    - Berätta om hur du ser på bedömning av muntlig färdighet i klassrummet, till vardags.
- Känner du dig trygg i bedömningen av muntlig språkfärdighet (jfr med andra färdigheter)?
- Gör ni på liknande sätt, ni som undervisar här?
    - Hur diskuterar du detta med dina kollegor?
- Hur diskuterar du bedömning av muntlig språkfärdighet med eleverna?
- Berätta om hur du ser på den muntliga förmågan i relation till andra språkfärdigheter?

Har du gått någon fortbildning eller upplevt något annat som stött dig i din utveckling av bedömning och betygssättning?

Har du gått någon fortbildning i engelska sedan din examen?

Appendix K: Consent form

Förutsättningar för bedömning av muntlig språkfärdighet

Studiens syfte
I denna studie undersöks frågor kring bedömning av muntlig språkfärdighet i det nationella provet i engelska i åk 9, en aktuell och viktig fråga i skolan idag. Syftet är att undersöka *uppfattningar om bedömningskriterier, bedömningspraxis* och *förutsättningarna för bedömning* bland engelsklärare som undervisar i år 9. I studien undersöks bakgrundsfaktorer som erfarenhet av samt utbildning i bedömning och betygssättning. Vidare undersöks hur lärare i grupp sambedömer elevexempel.

Studiens uppläggning och genomförande
1) Moment I (sammanlagt 12-15 lärare på olika skolor i landet): En individuell intervju med varje lärare. Intervjun kommer att ta ca 1 timme och handla om lärarens utbildning, erfarenheter och bedömningspraxis. Moment I genomförs under feb-apr 2013. Tider för intervju bestäms individuellt med dig som lärare.

2) Moment II (Samma lärare): Sambedömning. Lärare samlas i grupp med lärare från två eller flera skolor och diskuterar inspelade elevexempel. Denna diskussion och sambedömning spelas in och analyseras senare. Moment II genomförs under mars-maj.

3) Moment III (5-6 lärare): En individuell intervju, en uppföljningsintervju på ca 1 timme med ett mindre antal lärare. Lärarna besvarar frågor om sambedömningen av elevexempel i grupp, om bedömning av elevers muntliga språkfärdighet i allmänhet samt beskriver sin bedömningspraktik i engelska. Moment III genomförs under hösten 2013. Tider för intervju bestäms individuellt med dig som lärare.

Urval och frivilligt deltagande
Allt deltagande sker på frivillig basis. Skolor och deltagare anonymiseras och kodas under bearbetning och analys av data. Du kommer alltså att förbli anonym, liksom din skola.
Detta betyder att ditt svar inte kommer gå att urskilja, inga namn eller uppgifter som kan identifiera dig eller din skola kommer att finnas med. Intervjumaterial kommer att hanteras med stor försiktighet och sekretess.

Appendix K: Consent form

Möjlighet till reflektion
Hela undersökningen kan ses som ett led i fortbildning kring bedömning och betygssättning. Lärare som deltagit i liknande studier vittnar om att frågorna sätter igång en reflektionsprocess. Intervjun ger dig möjlighet att sätta ord på dina erfarenheter genom att beskriva och förklara olika val du gör i bedömningsarbetet. Diskussionen och sambedömningen av elevexempel ger dig tid och möjlighet att diskutera bedömning av specifika språkliga förmågor med kollegor.

Jag hoppas att du som deltar skall uppleva att medverkan i studien ger dig stöd i ditt viktiga arbete som engelsklärare och i det bedömningsuppdrag som vi lärare i Sverige har. Jag hoppas också att ditt deltagande skall öppna för nya diskussioner mellan dig och dina kollegor.

Om du har några frågor, är du välkommen att ställa dem via e-post: maria.frisch@gu.se.

Göteborg, 20 mars 2013
Maria Frisch
Forskarskolan FRAM
Göteborgs universitet
maria.frisch@gu.se
070 662 83 89

Deltagande

Jag har tagit del av den skriftliga informationen om studien och har haft möjlighet att ställa kompletterande frågor. Jag har även informerats om att jag kan avbryta mitt deltagande i studien när som helst.
Härmed lämnar jag mitt samtycke till deltagande i studien. Samtycket inkluderar också presentationer/publikationer av studiens resultat samt användande av mina data för eventuella framtida studier.

_____ _____
Datum Namnteckning

_____
Namnförtydligande

Appendix L: Guidelines for assessment from www.skolverket.se, accessible during spring 2013

Part A– *fokus:* fri muntlig interaktion och produktion

Den muntliga förmågan bedöms främst med tonvikt på *viljan och förmågan att delta i samtalet och att förmedla ett innehåll på ett begripligt sätt*. En muntlig prestation som uppvisar många språkliga brister kan vara godkänd om budskapet trots allt går fram och kommunikationen fungerar.

Bedömningsfaktorer

Vid bedömningen kan följande faktorer analyseras:

Vilja och förmåga att samtala och tala

- att interagera – ta initiativ, uppfatta vad andra säger och föra samtalet vidare
- att berätta, beskriva och argumentera
- att anpassa det som sägs till situation, ämne och mottagare

Innehåll

- fyllighet och idérikedom
- behandling av ämnet (fokuserad / fördjupad – kortfattad / ytlig)

Språk

- begriplighet – förmåga att uttrycka ett budskap klart och tydligt
- ledighet, variation och säkerhet – flyt
- strategier för att lösa språkliga problem
- vokabulär och idiomatik (omfång, variation, korrekthet)
- artikulation, uttal och intonation
- grammatik (omfång, variation, korrekthet)

Se också avsnittet ”Bedömningens inriktning” i kursplanen

© Skolverket