

‘Unpredictable and full of risks?’

*An evaluation of the exam assessment in English
in the R’94 vocational courses*

Anne Ely Thorenfeldt



Masteroppgave (60 studiepoeng) i engelskdidaktikk

Institutt for lærerutdanning og skoleutvikling

UNIVERSITETET I OSLO

Høsten 2005

1. Introduction	5
1.1 Terminology	6
1.2 Sources and working method	7
1.3 Short outline of the thesis	8
2. Theoretical platform	10
2.1 State of the art	10
2.2 CoE: Threshold level	12
2.3 The Communicative Competence Concept	13
2.4 Assessment and testing on a communicative basis	15
2.5 Reform'94	18
2.5.1 Aims and targets	18
2.5.2 Assessment	20
2.6 Exam test VF2210 – the construct	22
3. Theoretical reflections and previous research	24
3.1 Teaching English in the vocational classes.....	24
3.2 Teaching and testing vocational vocabulary	28
3.3 Presenting a related study	31
4. The testing of reading comprehension	34
4.1 Theoretical reflections on assessing reading	34
4.2 Estimating readability	37
4.2.1 The LIX index	37
4.2.2 The FOG index	38
4.3 Analysing and characterizing texts and comprehension tasks	39
4.3.1 Reading comprehension based on a single text	40
4.3.2 Reading comprehension based on a leaflet	47

4.4 Discussion of the findings	49
5. The testing of writing	53
5.1 Theoretical reflections on assessing writing	53
5.2 VF2210 - Prompts and writing instructions	56
5.3 Organized preparation	59
6. Summing up the written part of the exam test	60
6.1 VF2210 – brief descriptions and analyses of the tests.....	60
6.2. Targets tested	68
6.3 The testing of vocationally related English.....	69
7. The oral part of the exam test	71
7.1 Theoretical reflections on assessing oral competence	71
7.2 Examples based on interviews	73
7.3 Summing up the oral part	77
8. Discussion	79
8.1 Analyses and discussion of selected some findings	79
8.2 Some consequences of R’94 and regional exam design	84
9. Current and future trends	89
9.1 CoE: “Framework”	89
9.2 The ‘Knowledge Promotion’ Reform 2006	92
9.2.1 ‘Høringsutkastet’ – the discussion document.....	92
9.2.2 ‘Knowledge Promotion’ – the new syllabus	93
9.3 An interview with Olav Talberg.....	97
9.4 An interview with Ragnhild Stai Amundsen	100
10. Conclusion	102
11. Bibliography	106

Preface

This thesis is the result of many years of teaching English in the vocational courses of upper secondary education, and an urge for more knowledge about current thinking on the teaching of foreign languages. When I started this study, I was surprised to discover how little research had been done into the vocational areas of study, and how little interest the school authorities had given to these students, especially when it comes to the common general subjects. After 15 years of teaching, my heart still lies with these students, and now more than ever, I would like to dig deeper into the fascinating, though challenging world of teaching languages in the vocational courses.

I am deeply grateful to my supporting family, in particular my 13 year old daughter Jani, and to my colleges and superiors at Thor Heyerdahl vgs for helping me out and allowing me the time and opportunity of completing this study besides working full time as a teacher. A special thanks to Olav Talberg for providing me with important data, and to Ragnhild Stai Amundsen and the teachers who let me use their time and draw on their experience.

Finally, I am much indebted to my supervisor Aud Marit Simensen for indispensable professional advice, valuable encouragement and for believing in me from the start!

Live your dreams!

Larvik, November 2005

Anne Ely Thorenfeldt

1 Introduction

In the autumn of 1994 all young people in Norway were given a statutory right to upper secondary education leading to either qualification for higher education programmes or vocational qualifications. The Norwegian Ministry of Education, Research and Church Affairs introduced the new 'Core curriculum for Primary, Secondary and Adult Education in Norway'. This 44 page document formed the basis of a shift of paradigm in Norwegian teaching, or at least it was supposed to do so, introducing large-scale reforms in primary, secondary and higher education simultaneously.

On this binding foundation separate curricula and subject syllabuses were developed at the different levels of education. The focus of the present thesis will be on the curriculum for upper secondary education, the so-called 'Reform'94', and its implications for formal summative assessment: the exam tests of English as a foreign language in the vocational courses.

R'94 brought major changes to the vocational branches of the Norwegian Upper Secondary School. English was made compulsory for all, and the number of hours taught was doubled. Furthermore, it was decided that the curriculum for the more general subjects, like English and mathematics, would be the same for the vocational courses as for the general academic courses. The county authorities were made responsible for the new local exams.

Research question

In this study, I will look into the consequences of the current practice of making a common regional test in each county, meant to cover all schools and all areas of study. According to my preliminary enquiry, as many as 12 out of 19 Norwegian counties have chosen to do so, and my fear is that this represents a challenge to fairness as well as to the validity of the test and the reliability of the test results. Being a teacher of English in the vocational area of building and construction myself, I found it challenging to examine these points of worry in more detail. With the following research question I hope to highlight what I feel is the main problem:

To what extent has test making on a regional level become a threat to fair assessment in the VF2210 exam?

More detailed research questions will be as follows:

- Will a comparison between exam tests from different counties reveal significant differences in level of difficulty?
- Is the exam assessment in line with the teaching and/or is it in line with the R'94 intentions?
- What kind of test is the VF2210 exam test meant to be, and what is reality like? Is it an achievement test, a general proficiency test or both?
- Is the idea of a local exam lost along the way?
- To what extent are all four basic skills seen as equally important in the assessment?
- The curriculum contains two equally important targets of content. Do the exam tests reflect this in a reasonable way?
- Has the implementation of the R' 94 English curriculum had any significant consequences for the vocational students in particular?
- Are oral exams still 'examinations proper', or have performance-based tasks been granted a place in the oral assessment?
- To what extent is VF2210 in line with the R'94 curriculum and guidelines?

I will start by examining curricula, guidelines, theory on the subject and current practice in teaching to get an idea of what the assessment should be like, and then examine a sample of authentic exam sets and compare the current assessment to the theoretical platform.

1.1 Terminology

A few comments have to be made on the terminology used. As a general rule, I use *task* for the Norwegian 'oppgave' and *prompt* (Weigle 2002) for the Norwegian 'stiloppgave'.

Validation task is used about the kind of tasks asking the student to evaluate and compare the content of given statements to the text, in the form of tick off for 'true/false/doesn't say' or 'right/wrong'.

By¹ *proficiency testing* I understand 'testing that aims at giving information about a student's general level of proficiency' as opposed to *achievement testing* which aim is 'to give information about the achievement of a student in relation to a specific course'. When achievement testing is used at the end of a course, it is called *summative assessment*. Another term used is *criterion-referenced testing* as opposed to norm-referenced. Criterion-referenced

¹ The following definitions are based on Simensen 1998:Chap.10

testing means ‘measuring an examinee’s performance against certain criteria’, that is ‘determining to what degree the students have reached the aims and targets of the course’. By *integrative testing* is meant ‘measuring more than one language skill and/or several language components at the same time’.

The *reliability* of a test refers to its consistency as a test instrument. *Objective* tests, tests which are scored mechanically, are said to have a higher degree of reliability than subjective tests, which require a judgement on behalf of the examiner. The *validity* of a test is a question of to what extent the test measures what it is intended to measure. The term *construct* refers to a characteristic or a definition of the language skill(s) we want to assess. Content may or may not be part of the construct. ‘*Washback*’ effect from tests to teaching means that students as well as teachers tend to study examinations given in preceding years in order to find out what aspects of the school subject are likely to be measured.

Reform’94 will normally be referred to as *R’94*, meaning the R’94 curriculum for English as a common general subject. *R’94 English* refers to the English version. By *L1* I mean mother tongue use of a language, and *L2* will in this connection refer to English as a foreign language. Generally, *L2* is often used to refer to ‘English as a second language’ (Simensen 1998:11). Other abbreviations I will use throughout the thesis are: CCC = Communicative Competence Concept, CLT = Communicative language teaching, CoE = Council of Europe, EFL = English as a foreign language, ESP = English for specific purposes, MWU = multi word unit.

1.2 Sources and working method

I have chosen a flexible design with a qualitative approach for my research. To some extent I have categorized and quantified the data, but my method has not been systematic enough to be characterized as quantitative research. The study is an evaluation of existing material, supplemented by interviews. My intention was to find out what the exam test in the subject ‘VF2210 Engelsk’ is supposed to be like according to the existing curriculum and guidelines, and compare this to reality. I also wanted to document the consequences of the current system of test design for the students in question. VF2210 is a combined exam. I will mainly deal with the written part, and only briefly look into the oral part. My main source is a

representative probability sample, randomly selected from the population of regionally designed exam sets for the spring term exams, carried out in the different Norwegian counties the past few years. In addition, I have based my study on the given curriculum in English for vocational courses in R'94, the platform of the test makers as shown in the guidelines and regulations made by the Norwegian Board of Education, theory on the topic, research within the area and interviews with teachers.

To get hold of the needed material, I sent a letter of enquiry to all 19 counties, asking for samples of regionally made exam tests carried out in the period 2002 – 2005². Nine counties answered. Half a year later I made a new approach, this time with a better result – all the remaining counties except for one responded. Unfortunately, I have still not been able to get hold of this last sample.

Ending up with a total of more than 50 exam sets, I had to decide upon some criteria for selection, and I chose to limit my population of tests to the two last effected spring term exams from each county. Due to practicality, I had to leave out the seven counties in which they make local exams proper. In those counties a single exam set is developed for each particular group of students by teachers who know their specific syllabus, and the problem I have chosen to focus on, the consequences of regional tests, consequently do not exist. I could of course have randomly chosen some samples in order to compare the two types, but that would raise new questions, beyond the scope for this thesis.

This left me with a sample of 21 exam sets, all made for the same purpose by different test makers. In addition, I have collected information among five randomly, though accidentally sampled teachers about the oral part of the exam, and made interviews with two noteworthy persons involved with development of teaching and testing foreign languages.

1.3 Short outline of the thesis

The thesis is divided into ten chapters. In the introduction I have defined the aim of the study, commented on the terminology used and described my sources and working methods. The second chapter constitutes a theoretical platform for the study, including a presentation of Reform'94 and the state of the art: the communicative trend in EFL language teaching.

² The reason why I could not choose one specific year is that this kind of exams is not carried out everywhere every year.

Chapter 3 looks into related research and discusses this in relation to my own study. In chapter 4, the testing of reading comprehension is being examined and discussed, and the different sample tests compared with respect to readability, text difficulty and test tasks. Chapter 5 is a study of the testing of writing, including a brief outline of the use of preparation sessions and pre-writing material. Chapter 6 is a comment on the written part of the exam as a whole. Here I briefly describe and compare the different VF2210 exam sets, and sum up the findings. In addition, the chapter provides a closer look into the vocational aspect and points out the need for achievement testing with respect to this target. Chapter 7 makes available some examples of how the compulsory oral part of the exam is being carried out through interviews with five randomly chosen teachers.

The main discussion of the findings will take place in chapter 8, supplied by closer examinations of some chosen exam sets. Current and future trends are demonstrated in chapter 9, like the syllabuses in progress and the new common reference documents from CoE, and I will cite from interviews with two experts, both in different ways ‘decision-makers’ within the area, in order to take us beyond the focus on R’94 assessment and into the field of teaching and testing foreign languages in the future. Finally, in the last chapter I will present the conclusions from the study, and draw attention to areas in need of further research.

2 Theoretical platform

2.1 State of the art³

Since the 1980s the communicative and meaning-oriented approach (Simensen 1998: Ch10) to second- and foreign language teaching and testing has been influential, and with the R'94 this trend fully reached Norway. The previous work done by the Council of Europe has clearly been the platform of the Norwegian Ministry of Education as to developing the R'94. This includes the 1975 'Threshold level', the 1980s specifications of objectives and the communicative competence concept and finally the 1990 'Threshold Level' (to be further dealt with below). Traces of these ideas and concepts are easily found both in the 'Core curriculum', the specific curricula for the different subjects and in the assessment guidelines.

However, the implementation of R'94 changed the scene completely for the subject of the present thesis: English as a foreign language in upper secondary vocational education. New ideas of teaching methods, project-work, inter-disciplinary topics, self-directed learning, criterion-based assessment, performance-based tests, testing language in context, discussing evaluation-criteria with the students and so on confused and scared quite a few language teachers around the country. New textbooks trying to grasp the gist of the new concept, but aiming far too high, were just another part of the picture – and there were no clear descriptors to back up the criterion-based assessment, even if serious attempts were made to meet that need...

R'94 lines up six targets for English; four targets focusing on the basic language skills (Brown 1996) and two targets specifying content. These targets will be presented in section 2.5.1. The basic skills as defined in R'94 are the receptive skills *reading* and *listening*, and the production skills *writing* and *speaking*. The targets underline the difference in mode by specifying 'understanding' as the main competence in connection with target 1 (listening) and 2 (reading), and 'use of' in connection with target 3 (speaking) and 4 (writing). But to what extent has this fact influenced the assessment?

The last two targets specify the content: target 5 - The English-speaking world (focusing on the USA in module 1 and the UK in module 2), and target 6 - English in relation to the

³ This section is based on a paper written by me in the autumn term 2004 (Thorenfeldt 2004)

specific areas of study. Shouldn't these targets then, according to R'94 (R'94:3.2), be subject to achievement testing in the final exams? The focal point in interpreting criterion-referenced assessment scores is, according to Brown (1996), always on how much of the material each student has learned in absolute terms. Criterion-based tests are usually produced to measure achievement according to well-defined and fairly specific objectives, often specific to a particular course (Brown 1996:2).

State-of-the-art thinking on assessment has taken on a whole new approach compared to the previous 'discrete-point' trend, focusing not on the mistakes made, but on the actual achievement in relation to the criteria, allowing all candidates to perform to the best of their ability. Primary focus is on the students' ability to perform communicative tasks effectively, according to Read (1997). The learner is accorded a more significant role in the teaching/learning process, and slogans like 'self-directed learning' and 'student participation' call for deeper involvement from both sides of the desk, and a need for comprehensible objectives. To what extent is this reflected in the assessment?

As a result of the influence from the integrative-communicative movement (Brown 1996), based on the view that the whole is more than the sum of its discrete parts, language proficiency came to equal communicative competence, and the weight was put on procedural rather than declarative knowledge. R'94 calls for a 'holistic assessment of the students' competence as described in the aims and targets', but a holistic approach, defined by Weigle (2000:65) as 'single scores based on an overall impression', opens for subjective weighting between the different skills and content areas involved, and represents a challenge to reliability and fairness.

The concept of integrative-communicative, criterion-referenced, holistic assessment implied a new type of tests. One problem with integrative test-types (Simensen 1997:Ch10) is that reliability tends to be low. The situation for the test in question, the VF2210, is that it is being rated not only subjectively, but also quickly; in most cases in the period between the written and the oral part of the combined exam. I will return to this in section 7.3. Brown (1996:Ch1) emphasizes the need for well-defined and fairly specific objectives, and the importance of letting the students know in advance what kind of tasks and content to expect. Current practise on test making seems not to be quite in line with this view.

2.2 Council of Europe: “The Threshold Level” 1975 - 1990

The research of the Committee for Cultural Cooperation of the Council of Europe represents pioneer work in syllabus design. The committee’s work resulted in ‘threshold levels’ in language learning objectives and specifications for various European languages, and built a foundation for syllabus design and EFL curricula all over Europe. The Threshold Level Project consists of the 1975 Threshold Level document, drawn up by Dr. J A van Ek on behalf of an expert group convened by the CoE, Scope from 1986 specifying objectives for second language learning, and the revised and extended Threshold Level 1990. The project provided models on which national and local syllabuses may be based.

Threshold Level 1975 was primarily concerned with the functional and semantic sides of language learning, specifying what learners need *to do* with language, and in relation to what notions he or she needs to do it (van Ek 1975:4-7). Among other things, the need to know culturally acceptable and appropriate ways of interacting verbally is highly stressed.

The Threshold Level 1990 is, as mentioned above, a revision and an extension of the 1975 document on the basis of practical experience with the first model and new theoretical insight gained since 1975. This insight is related first and foremost to aspects of language in context. The most influential conceptual framework for the teaching of communicative competence is the one developed by Canal and Swain in 1980. Here communication is interpreted as a form of social interaction (discussed in Simensen 1998:105). Four components of knowledge and skills were included: grammatical competence, sociolinguistic competence, discourse competence and strategic competence. ‘Knowledge’ was defined as what one knows, consciously and unconsciously, and ‘skills’ as how well one can perform this knowledge in actual communication. These four competences were included in Threshold level 1990, complemented by another two: socio-cultural competence and social competence.

While the 1975 document specifies the ‘hard core’ only of communicative ability, the 1990 document is representative of the complexity of the communicative competence concept specifying a total of six competences: linguistic, sociolinguistic, discourse, strategic, socio-cultural and social competence(van Ek and Trim 1990:iv).

1. Linguistic competence (concerns mastery of 'grammatical' knowledge + knowledge of vocabulary, pronunciation, orthography as well as sentence structure and morphology)
2. Sociolinguistic competence (concerns mastery of which language forms as well as meanings are appropriate in different contexts)
3. Discourse competence (concerns the ability to receive as well as produce connected texts)
4. Strategic competence (concerns mastery of verbal and non-verbal strategies to compensate for breakdowns in communication and gaps in the language user's knowledge of the other competences)
5. Sociocultural competence (concerns the ability to understand and accept the feelings of others and accept people with different social and cultural background)
6. Social competence (which applies to both the will and skill to interact with others)

In principle, the model is limited to decisions about what to teach. Objectives were described in operational terms ('can do'-statements), and the focus were on the minimal requirements for effective communication only (Van Ek and Trim 1990:118). They were arrived at on the basis of systematic analysis of the needs of the learners as prospective speakers of the target language, such as *situations* in which the learner will use the target language, *topics* which have to be handled at a threshold level and *language activities* in which the learner will take part (Simensen 1998:106). The Threshold Level also includes new theoretical concerns in other fields. There is, for example, emphasis on notions like insightful learning, consciousness in learning, 'learning to learn', learner autonomy, self-directed learning, individual variation and learning strategies, all recognizable in R'94.

The Council of Europe's understanding of communicative competence as defined in the Threshold Level project is clearly the basis of the Curriculum of English in the Norwegian secondary school. This is evident in chapter three of the Curriculum Guidelines where it is stated that the basis for the final assessment of the learners is the degree to which they can be said to have achieved communicative competence (R'94:3.4). The notion of 'communicative competence' in connection with assessment is explicitly defined in the list of criteria on the previous page. Assessment within the CCC will be further discussed in section 2.4.

2.3 The Communicative Competence Concept and a socio-cultural perspective

However, CoE was not the one and only promoter of a communicative approach to language teaching and learning. As early as in 1971, American anthropologist and sociologist Dell Hymes brought forth a new concept compared to the existing *linguistic* competence concept;

the *communicative* competence concept, defined as knowing “when to speak, when not, and as to what to talk about with whom, when, where, in what manner” (Hymes,1972:277). As we can see, Hymes concept includes language knowledge as well as ability for use. Aud Marit Simensen (Simensen, 1998:78) underlines that this is a definition of the native speaker’s comprehensive competence, a competence which enables him or her to produce and understand utterances in relation to specific contexts, social and situational, and specific communicative purposes. Thus, rather than to aim at full communicative competence for most Norwegian L2 students, a more realistic objective is to aim at the highest possible degree of communicative competence (R’94:3.4).

Hymes’ communicative competence concept sparked off a series of studies, and initiated a new interest in pragmatic knowledge for appropriate communicative behaviour in L2 as well as in L1. His ideas gave birth to a communicative trend in language teaching in general, further developed by among others Canale & Swain in 1980 and the CoE in the ‘Threshold level’, and, as mentioned above, easily recognizable in the R’94 syllabuses. In R’94 it is explicitly stated that the goal of English teaching is that ‘the students should reach a high level of communicative competence, and this competence should be developed through working with the four basic language skills’ (R’94 English:Chap.2) .

The prevailing theory on L2 education argues for a socio-cultural approach to education. Social interaction plays an important role in the learning process (Warshauer and Kern, discussed in Ulriksen 2002). The Russian psychologist Lev Vygotsky insisted that language learning is interactive and social. Vygotsky described language as having two functions; a psychological, i.e. language used to make sense of experiences, and a cultural; language used for communication, the latter being the focal point in foreign language teaching and learning. Vygotsky’s most widely known concept is the “zone of proximal development”, referring to the difference between what the learner is capable of when acting alone and what he or she is capable of when acting under a more experienced other. Vygotsky shared many of Piaget’s⁴ views about children and learning, but placed more emphasis on the social aspect of learning. According to Elizabeth Murphy (Murphy 2000, discussed in Ulriksen 2002), the constructivists came up with a new conception of language learning, seeing learning as a result of the student’s active participation in tasks that they find relevant and engaging.

⁴ Jean Piaget, influential Swiss psychologist 1896-1980

Williams and Burden list ten propositions (quoted in Ulriksen 2002) that may serve as a guide for teaching and learning languages from a socio-constructivist perspective in the 21st century, among them the need for putting “the learner at the centre of the learning process”.

2.4 Assessment and testing on a communicative basis

The focus in the communicative and meaning-oriented trend in L2 testing is on measuring language in context (Simensen 1998:270). Integrative test types are usually favoured, and an ideal is to make the students perform on basis of several types of input, listening and/or reading. Appropriate responses will be as important as correct ones. An argument in favour of integrative test types is their high degree of validity, but a problem has been that the reliability of many of the favoured test types in this trend tends to be too low, especially when testing oral proficiency, and that the examinees’ responses have to be subjectively rated.

As opposed to discrete-point testing, the key word of integrative testing is *integration*. If communication is ‘...a functional, context-bound and culturally-specific use of language involving an integrated view of receptive and productive skills...’ (Candlin in Oller 1979 quoted in Pettersen 2001:15), then communicative language testing should mirror this. Since a test must provide a satisfactory sample of the syllabus in order to be valid in relation to this specific syllabus, and communicative language teaching is based upon normal language use, communicative tests should also reflect normal language use (Pettersen 2001: 15-16).

Pettersen argues that even if most test makers seem to have consented to integrative test types, we have not yet succeeded in designing tests that measures communicative abilities, perhaps partly as a result of the difficulties with defining what these abilities really are.

Since the implementation of integrative test types in the Norwegian exam assessment, teachers and examiners have expressed their confusion as to what they really test, and have had doubts about the validity as well as the reliability of these test types. As early as in 1979, T. Abrahamsen wrote an article in the magazine ‘Språk and språkundervisning’ (discussed in Pettersen 2001:78), in which he claimed that *essay* was not a suitable test of students’ communicative ability. His point was that vocational students were able to write short impeccable essays with a very limited vocabulary, and as he consider the lack of words an

essential weakness in students' communicative ability, he called for language tests aiming to encourage students to expand their vocabulary.

However, his critique has had little impact on Norwegian exam design, and the essay is still by far the most stable and trusted element in the exam tests. R'94 was meant to bring a change from the tradition of testing the students' reproduction ability to testing their ability to make use of their knowledge. Both Talberg (below in section 9.3) and Gulbrandsen (1997:25) express their criticism against the reactionary attitudes among the teachers, and argue in favour of more 'open' tasks giving the students a possibility to choose both topic and perspective themselves, with the intention of letting them include topical knowledge of their own specific syllabus and area of interest. It is also the clear intention of the National Exam Secretariat, stated in the guidelines for assessment (SUE/vg – 95-005) and repeated in a report from 1997 (Talberg 1997) that the 'examinees should be given the possibility of showing their knowledge and skills in 'real life' situations. Attempts has been made to define what a 'real life situation' is with respect to exam assignments (i.e. in SUE/vg -99-022), but the issue is still open to discussion. I will return to this in section 8.

When assessing in a school context, there are certain principles that ought to be considered (Simensen 1998:chap.10). The formats used for testing should be familiar to the students from the daily work in the classroom, the test should aim at measuring what the students know, and not what they do not know, and the criteria for evaluation and notably the rating scales for rating their performance should be discussed with the student. However, if the teaching is not in line with the intentions of the curriculum, and the exam tests are made by others than the class teacher, problems are likely to occur.

To comply with the objectives in the curriculum, an R'94 exam test should comprise 'real life' integrated tasks in order to assess the different skills in specified contexts, focusing on communicative functions and on students' performance, and emphasising appropriacy more than correctness. As discussed above, essay writing may or may not be seen as a suitable test in this connection, but it is the preferred type for testing writing competence. The writing instructions and wording are decisive as to whether writing an essay may be seen as a 'real life' task or not. I will return to the testing of writing in section 5. The main discussion in relation to the testing of reading comprehension deals with the problem whether or not it is possible to make authentic test tasks without involving writing skills. Reading comprehension

will be discussed extensively in section 4. The testing of listening skills may be included in the written part as well, as dictation of sentences, or as listening to an oral text while doing validation or multiple choice tasks. In the latter case, the students' writing skills will not interfere with the assessment of their listening skills. Listening may also be part of an oral test, in the form of listening to a recorded oral text or conversation, and subsequently talk about it. This, in addition to speaking skills will be touched upon in section 6.

In R'94, the objectives of the course form the criteria for evaluation⁵, and these objectives include, as shown above, six targets of knowledge and skills and the six components of the communicative competence concept. In addition, the most relevant guidelines⁶ for assessment of the VF2210 are listed below. The combined exam of the vocational courses was awarded little or no attention in these documents.

- R'94 Core Curriculum
- SUE/vg -95-005 Central guidelines for local and central exams
- SUE/vg -99-022 Oral exam in modern languages in upper secondary education
- SUE/vg -99 -027 Guidance in local making of oral exam tests
- 1994 – preliminary assessment guide for written English in upper secondary education
- 1999 – preliminary assessment criteria for English in upper secondary education
- Methodology guide for English

On this platform, can we expect the VF2210 exam test to be an achievement test or a proficiency test? Brown (1996:14) underlines that 'achievement testing usually involves testing to find out how much each person has learned within the program, thus an achievement test must be designed with very specific reference to a particular course. This link with a specific program usually means that the achievement tests will be directly based on course objectives and will therefore be criterion-referenced. Such tests will typically be administered at the end of a course to determine how effectively students have mastered the instructional objectives'.

Hence, the purpose of achievement testing is to determine the degree of learning for advancement or graduation, and the focus is the terminal objectives of the course. Proficiency testing, on the other hand, is norm-referenced and normally administered before entry to a

⁵ *Metodisk rettleiing. Engelsk grunnkurs. 1994*

⁶ My translation of the titles

course, and proficiency tests are likely to be very general. Consequently, it seems final exam assessment in general should be based on achievement testing.

2.5 Reform'94

English is described as a general subject common to all areas of study in the upper secondary school curriculum. In vocational areas of study, all students have a total of 150 periods⁷ - comprising module 1 with 2 periods of teaching per week in the foundation course and module 2 with 2 periods per week in advanced course 1. Students taking vocational areas of study but seeking university entrance qualifications must in addition take module 3 with 37 periods.

2.5.1 Aims and targets

According to the R'94 curriculum for EFL, the overall aim of the education is that the pupils will attain a high degree of communicative competence. Optimal communicative competence in English as a foreign language is defined as the ability to understand authentic English in all kinds of communicative contexts and ability to use correct and idiomatic English in all kinds of situations, but in Norwegian upper secondary education the aim needs to be put lower than optimal competence (R'94:3.4). Language learning consists of acquiring knowledge and mastering skills. Knowing about language and its use, and about social and cultural affairs, all help to enhance the general mastery of a language. The knowledge targets consist of subject matters and working material, and the skills targets contain requirements with respect to both comprehension and production. Ideally, all activities in the classroom should be 'meaningful and realistic'. According to the guidelines of methodology (Metodisk Rettleiing 1994), all activities should be directed towards the superior objective, namely a good communicative competence.

The **aims** of the pupil's study of English (R'94: Ch 2) are⁸:

- to be able to use English which is suitable both in informal and formal situations, and to know how the social context affects the use of the language
- to develop a varied general vocabulary, and a specialized vocabulary appropriate to the pupil's area of study

⁷ A period = 45 minutes

⁸ Quoted from the English version of the R'94 curriculum

- to acquire good clear pronunciation and sufficient familiarity with the rules of English pronunciation to achieve this
- to be able to grasp the meanings and connections of spoken and written English, and express him-/herself so as to bring out intentions and connections clearly
- to acquire sufficient knowledge about the language as a system to be able to understand grammatical explanations and correct errors
- to become acquainted with social and cultural conditions and ways of life in English-speaking societies, and to develop the willingness to the social skills needed to communicate in various situations and to interact with people with other cultural backgrounds
- to be able to make use of such aids as dictionaries, grammars, reference works and such information technology as may be available

English as a school subject also comprises definite aims with regard to attitudes and experience; these do not lend themselves easy to assessment (R'94:Ch.2). Mastery of a foreign language is developed by working on the four skills, targets 1 to 4, focusing on the contents specified in targets 5 and 6. Each target includes more specified sub-targets containing the main points of study. The sub-targets of target 1 to 4 will be found in different chapters below⁹. The sub-target of target 5 and 6 are listed here, as they are relevant to testing of all the different skills.

Target 1 Comprehension of spoken English¹⁰

The pupils should become skilful at understanding various forms of spoken English, ranging from informal to more formal uses of the language.

Target 2 Comprehension of written English

The pupil should become skilful at understanding various types of English texts, according to their nature and structure and the purpose for which they are read

Target 3 Use of spoken English

The pupil should develop good communicative skills in various situations, and be able to adapt his/her language Use to the particular situation.

Target 4 Use of written English

The pupils should be able to take written notes, to express his/her own attitudes and opinions in writing and To write various kinds of texts.

Target 5 The English-speaking world

The pupil should acquire a certain amount of knowledge of literature in English and of topics relating to social and cultural life in English-speaking countries.

⁹ Target 2 in chapter 4, target 4 in chapter 5, and target 1 and 3 in chapter 7.

¹⁰ The targets are quoted from the English version of the R'94 curriculum

Main points: The pupil should

- 5a
- have an overall picture of the English-speaking world
 - know about the history and geography of the USA
 - know about the social conditions, customs and values in the USA
 - know about education, working life, industry and environment in the USA
 - be able to describe and discuss the contents of at least two short stories or excerpts from a play, either read or seen
- 5b
- know about English as an international language
 - know about the history and geography of the Great Britain
 - know about the social conditions, customs and values in the Great Britain
 - know about education, working life, industry and environment in the Great Britain
 - be able to describe and discuss the contents of a literary work

Target 6 English in relation to the respective areas of study:

The student should be able to understand and use English (Targets 1-4) in contexts that are relevant to his/her area of study.

Main points: The pupil should

- 6a
- be able to understand and use English on subject-related and inter-disciplinary topics as appropriate to the pupil's year and area of study
- 6b
- be able to understand and use English on specialized subject matters as appropriate to the pupil's year and area of study

Note: According to the syllabus the amount of material studied in each module under Targets 5 and 6 must be balanced, so that neither target exceeds 60% or falls under 40% of the total.

2.5.2 Assessment

As already established (section 2.1), the assessment of English in R'94 is criterion-referenced and holistic. The assessment should reflect the individual student's level of knowledge and skills, seen in relation to the criteria and targets found in the curriculum. This seems to call for a combination of achievement and proficiency testing (Brown 1996:11). I have been searching for official guidelines about assessment for the combined exam in particular, but except for appendix 2 in the R'94 curriculum for English as a common general subject, they all relate to the more 'academic' branches of study, and only 'bits and pieces' apply to the assessment of VF2210. The R'94 Curriculum, section 3.2, points out three specific areas of consideration¹¹:

¹¹ My translation

- The assessment should be done in relation to the targets established in the core curriculum and the curriculum of English as a common general subject.
- It is the pupil's holistic competence which is to be assessed, as it is described in the aim of the education (see below)
- The intention of the assessment is to show to what extent the pupil has reached the targets in the curriculum.

In a school context we find two types of assessment, formative assessment during the course and formal summative assessment at the end of a course, and they serve different purposes (Simensen 1998). The purpose of the formal summative assessment is to inform society, future employers and further educational institutions about the pupil's competence in English. There are two kinds of summative assessment: Assessed attainment at the end of the course and the formal result of the exam test. According to appendix 2 of the R'94 curriculum, the students in the vocational courses will be given summative assessment as follows¹²:

- **Assessed attainment:** Pupils will get one single grade after Module 1 the first year, assessing the 2-period foundation course, and one single grade after Module 2 the second year, assessing the 4-period course as a whole = Module 1 + Module 2.
- **Exam assessment:** In addition, pupils may be randomly selected through multistage cluster sampling for a combined written and oral examination. The exam tests are made locally, following centrally given guidelines, and will result in one single grade.

As stated above, the purpose of the summative assessment is to find out to what extent the student is able to understand and generate or bring about meaning; that is which **level of communicative competence** he/she has reached. Assessment should be based on (R'94: chap.3.4) to what extent the pupil is able to¹³:

- master correct grammar, vocabulary and pronunciation (linguistic competence)
- interpret and use a appropriate language in various contexts (socio-linguistic competence)
- grasp and express the meanings and connections of spoken and written English (discourse competence)
- employ suitable alternative ways of communication to compensate for lacking language skills (strategy competence)
- use their knowledge of the foreign cultures in their interpretation of text and in meetings with other people (socio-cultural competence)
- and willing to and capable of use his/her linguistic competence (social competence)

¹² Quoted from the English version of the R'94 curriculum

¹³ My own translation

However, we have to keep in mind that language learning consists of both knowledge and skills (R'94:chap.2), that the six targets listed above constitute the objectives of the course, and that the final assessment of upper secondary students aims to show to what extent they have reached these targets (R'94:chap.3.2).

2.6 Exam test VF2210 – the construct

The term 'construct' refers to what such a test should be like; the platform for the design, so to say. What kind of tasks and topics will have to be included to give the test in question 'construct validity' (Alderson 2002: Ch.1) – to make sure it measures what it was intended to measure? The question is especially interesting since the VF2210 is locally, or more often regionally, made. To secure equal measurement of the students, all test-makers of parallel tests around the country need to agree on this construct. The Norwegian Board of Education gives, as far as I have been able to find out (R'94:Chap.3) and (SUE/Vg -95-005), the following rather general guidelines¹⁴ for local exam test design:

- VF2210 is a local exam
- Each county is responsible for the making of the tests in accordance with the centrally given directions.
- The exam test VF2210 has to be a combination of written and oral testing on the same day, resulting in one single grade
- The written part may last from 1 to 3 hours. The oral examination may last 20-40 minutes.
- The curriculum in each subject constitutes the directions for the content of the exam test. For VF2210 this means the six targets.
- More than one target must be subject to testing.
- The exam test should, as far as possible, be designed in such a way that the pupil's holistic competence will be assessed.
- The test must be designed so as to give the pupil a possibility to show his/her competence.
- The amount of tasks given has to be realistic in relation to the time available. There must be time for reflection and deepening.
- Common general subjects should be related to the area of study of the pupil, i.e. by giving different prompts to choose from.
- There may be an organized preparation session the day before or earlier the same day.
- The exam test has to be taken individually
- The pupils are allowed to use bilingual dictionaries

¹⁴ My own translation

According to the Exam Secretariat (SUE/Vg-97-021:6) ‘the exam is a final assessment in which as many competence elements as possible should be included at the same time’ and the following aims are relevant for VF2210¹⁵:

The exam test should:

- provide a basis for assessing the student’s holistic competence as defined in the aims and targets
- be as ‘wide’ as possible, including both subject-related and common general targets
- let the students show their competence in ‘real life’ situations and through ‘real life’ tasks
- let the students show their ability to find and utilize information
- let all students perform to the best of their ability by including tasks on different levels of difficulty
- be realistic in relation to time frame and give time for reflection and deepening

In addition, the new 1999 rating guide for oral assessment (SUE/vg -99-022) points out the following issues to be considered¹⁶:

The assessor should:

- be aware he/she is not in the role of the teacher; his/her duty is to evaluate, not to correct the answers
- assess ‘open’ tasks with an ‘open’ mind
- evaluate - make a value judgement
- be aware that ‘real life’ tasks call for ‘real life’ assessment
- be in the role of receiver or user of the message, and evaluate whether or not the message bring about useful and relevant information (focus on function more than form)

The examinee should:

- show, not reproduce knowledge
- be given the possibility to discuss and evaluate content while answering
- show what he/she has understood, not what he/she remembers

Altogether, when adding up the information given in the different documents, we do get some idea of what this exam test should be like, but a lot is still open to interpretation and discussion. The basic competences may be assessed through general proficiency testing, but the fact that the final assessment of the students aims to show to what extent they have reached the targets of the curriculum (R’94:chap.3.2) calls for including achievement testing in the exam assessment. In the next chapter, two cases of previous research within the area will be accounted for and to some extent discussed.

¹⁵ My translation

¹⁶ My translation

3 Theoretical reflections and previous research

3.1 Teaching English in the vocational classes

To my experience, most teachers seem to agree that students' motivation is of vital importance for the result of the work done in the vocational language classroom. R'94 curriculum for English, chapter 1, states that 'The pupil's own motivation is a major factor in all language learning. In their upper secondary school training, their interest in English can be sharpened if their learning is closely related to their interests in other subjects'.

Yet, there seems to be a gap between the intentions in the Norwegian curriculum and current practice in the average vocational language classroom (Ulriksen 2002:84). The core curriculum stresses the shift from spotlighting the teacher and his/her contribution, to considering class dialogue to be the focal point. The students ought to be active participants in their own learning process, as the goal of second language learning is to develop a high degree of communicative competence. Ulriksen observed again and again the teacher by the blackboard doing her things while students were busy doing their things, and there was no dialogue between the two. She witnessed several examples of 'ancient' teaching methods that clearly didn't work well in the 21st century vocational classroom. These observations correspond by and large to my own experience. The 'audio-lingual' classroom with the teacher being the 'orchestra-leader', or even methods originating in the 'grammar-translation' period is in many cases still the rule.

Researchers now see the classroom not as much as a place where the language is taught, but as one where opportunities for learning of various kinds are provided through interactions that take place between the participants. This applies clearly to the vocational student as well. Current thinking about L2 learning also underlines the importance of teaching to the students' personal needs, interests, previous knowledge and experience and the significance of engaging the students themselves actively and personally in the learning process (Simensen 1998:136).

Oral skills

R'94 stresses that the students should work with and develop on the four basic language skills. Oral skills in the vocational branches are, intentionally or unintentionally, to a large

extent neglected, while great emphasis is put on reading and writing (Ulriksen 2002:85). 'To put such emphasis on reading and writing is a thing of the past, leading all the way back to the Grammar-Translation period', Kjersti Ulriksen claims. She found that very little English was actually spoken in the classroom; most of the communication between the teacher and her students was in Norwegian. This is not at all in line with the current view on the role of the teacher as Vygotsky's 'more experienced other' (see section 2.3) being a knowledgeable participant in the classroom dialogue. Vocational students seem to be more 'at home' with oral language than with written (Korsvold 2000:199), and they express clearly, when asked in Ulriksen's study, their discontent with the lack of oral activity in the classroom.

The attention paid to communicative language teaching is to a large extent the result of new knowledge of the processes involved in language learning. According to Stephen Krashen's influential 'input hypothesis'¹⁷, the learner acquires new or more language only by being exposed to comprehensible input. Not only reading, but also listening has its natural place in the communicative language classroom. The teaching should ensure that a variety of text types are used as listening comprehension material, such as conversations, interviews, debates, radio plays, news and weather forecasts, lectures and poetry, short stories and novels. Sometimes there may be good reasons for using a written transcript alongside listening (Simensen 1998:147). According to Simensen, the most valuable listening comprehension 'material' is the speaking produced by the linguistically competent and communicatively proficient teacher or by native speaker visitors. The 'message' is then normally adjusted by the speaker to fit the level of understanding of the listener, and comprehension is helped by here-and-now aspects of the situational context. Simensen also refers to the possibility of repeating, paraphrasing and elaborating the message to help the understanding when comprehension problems occur. This is in fact close to what happens in authentic, real life situations.

ESP¹⁸: Vocationally oriented English

In his article "Teaching Vocationally Oriented English" teacher, researcher and textbook writer Olav Talberg (Talberg 2004), seems to be very much in line with current views, even though he claims that his basic teaching strategies have not changed over the years. Based on

¹⁷ Part of the American linguist Stephen Krashen's "Monitor theory" of second language acquisition, in Krashen (1983) "The Natural Approach".

¹⁸ ESP = English for specific purposes

his experiences through 25 years of teaching English for students in the technical and industrial areas of study in upper secondary schools, he underlines the importance of meeting these students' needs as to prepare them for 'working life'. Teaching vocationally oriented English is basically about designing a language programme that ensures that students develop language skills relevant for, and related to, what they will need in the future as professional and skilled workers, according to Talberg.

Many language teachers express their frustration as to having to cope with the ESP vocabularies of the different areas of study. 'Language teachers in vocational classes should not become a teacher of the subject matter, but rather an interested student of the subject matter' is a statement expressed by Hutchinson and Waters (in Ulriksen 2002:90). Talberg encourages English teachers in vocational courses to seek cooperation with vocational teachers to see what is going on in order to feel less alienated. He personally did just that when being a new teacher in a vocational course, and was offered a lot of material – manuals and brochures in English as well as in Norwegian - by his fellow technical teachers to use in his English classes (Talberg 2004). He also activated and made use of the student's technical knowledge by asking them to explain in English the technical matters he himself was not quite familiar with.

As he wanted to focus on developing the students' overall communicative skills, his students were asked to present and explain – orally as well as in writing – subjects such as 'generation and distribution of electricity'. During a school year, the students did on average five to six oral presentations in front of the class and handed in a similar number of written presentations. Translation and isolated grammar exercises were seldom used, but the student were often asked to explain technical subjects of their own choice. Vocabulary exercises were used when a new topic was introduced. I will return to the teaching of vocabulary in the next section.

According to the article, Talberg's students worked in small groups during the majority of the classes, reading texts out loud to each other and collaborating on doing exercises and assignments. In fact, the students used English actively for about 30 minutes of each 45 minutes class. Working in pairs, they gradually became quite used to being 'challenged', and they just loved getting a chance to demonstrate their knowledge to their teacher. Talberg's point was that by focusing on vocational subjects, the students would develop their

vocationally oriented language as well as their general English language skills. And this was basically what he experienced.

Two periods a week insufficient

The current organisation of English in the vocational areas of study – spread out over two or three consecutive years – has become subject of some criticism, for a number of reasons. Teachers maintain that such an organization is not efficient (Gurholt 1997:23). To acquire language skills, students need a study programme that is quite intensive over a relatively short period of time. The common view among teachers (Talberg 1999:10) is that the module of 5 periods a week for one year should also be offered to students in vocational courses, preferably in the second year to make possible utilization of the students' experience from their vocational area. The possibility of doing so has been strongly reduced since the R'94. Kjell Gulbrandsen of the Norwegian Exam Secretariat recognises the problem (Gulbrandsen 1997:24), especially when it comes to students from vocational areas of study wanting the complete 5 period course. He agrees that even if $2+2+1=5$, this organization represents a disadvantage for the vocational student at the final exam. It seems that the new curriculum of 2006 to some extent will improve the situation. I will return to this in section 8.

The monolingual classroom

Another important issue, that of the monolingual classroom, is stressed both by Talberg, by the students in the study of Ulriksen and by Aud Marit Simensen (Simensen 1998:243). Establishing a monolingual classroom means that the teaching itself and the organisation of work in the classroom should take place in the target language. The teacher should resist the temptation to switch into the students' L1, and the students should be expected to speak in the target language. If the teacher is 'permissive' with regard to this principle, the students may choose the easiest way. A negative consequence is the tendency revealed in Ulriksen's study of students claiming to be bored and losing interest in what is going on in the classroom because they experience it as unimportant and not at all challenging. This may look different from the teachers' point of view. To my own experience, based on conversations with fellow ESP teachers, the tendency of switching into Norwegian arise from the lack of motivation among the students. It is not easy to 'role play' an authentic EFL situation in the classroom if there is a constant argument going on between the teacher and the group as to whether or not they need to cooperate and participate in the classroom dialogue and activities.

3.2 Teaching and testing vocational vocabulary

The assessment of L2 vocabulary has been affected by changing perceptions of the importance of vocabulary in language acquisition. From one perspective it is self-evident that vocabulary knowledge and skills are indispensable for successful communication in a foreign language (Read 1997); this applies especially to ESP, when the focal point is to be able to understand and use the English language within a specific vocational area. However, in the recent history of language teaching and research, vocabulary has tended to be taken for granted, with the main focus being on the development of functional communication skills. In her study of a widely used textbook for English in the vocational courses, Anne Karin Korsvold (Korsvold 1997:58) found only five tasks dealing with vocational vocabulary, and claims this to be a serious shortcoming. The discourse dimension of vocabulary tasks was found to be totally absent.

Still, there seems to be a renewed interest at present in the teaching and learning of vocabulary. Psychologists talk about the mental lexicon, semantic networks, types of lexical knowledge and cognitive depth of processing, while linguists study collocations and corpora. Because of the communicative competence concept, you just can't do without words. When it comes to ESP teaching, the most useful vocational vocabulary deals with job processes, procedures and tools, and in order to understand and use the technical vocabulary, you also need to know what the terms mean and what they refer to (Talberg 2004).

Teaching a vocational vocabulary is not normally regarded as easy. Many teachers find it a tough challenge to teach a vocabulary they are not quite familiar with themselves (Ulriksen, 2002:48). Being an English teacher in a vocational school, you may have to change area of study more than once, from *Health Care* one year to *Building and Construction* the next. To my own experience, and also according to Talberg, this brings about a tendency among teachers to leave out the vocational syllabus altogether, or at least to spend as little time as possible in the danger area.

To most vocational students this is a loss. The new vocabulary connected to their area of study is about the only thing which is not pure repetition from previous years. Research (Ulriksen, 2002:58) has shown that one of the reasons students give for being bored in the vocational English classroom is that everything is repetition. Not that learning a new

vocabulary, with all the work it takes, necessarily would make them shout with joy, but to my own experience, they normally find it useful. According to Ulriksen's study, students want more oral activity in class, and learning words, elaborating and consolidating the learning may very well be an oral activity. Talberg (Talberg 2004) refer to similar experiences.

Vocabulary acquisition is an incremental process according to the theory on the subject. Vocabulary teaching means more than just introducing new words; it also includes nurturing partially known vocabulary along to the point where learners can use it at will. Mere exposure to language and practice with functional language will, according to Norbert Schmitt (Schmitt 2000), not ensure the acquisition of an adequate vocabulary. Best practice according to current theory includes both a principled selection of vocabulary and an instruction methodology that encourages meaningful engagement with the chosen words over a number of recyclings. One way to do this is to introduce the new words through texts, and then elaborate, expand on and consolidate these words through classroom activities.

An important implication for vocabulary teaching is the conception of vocabulary in itself, not as single words but as larger stretches of words – the multi-word units. Schmitt (2000:97) gives the following definition: 'The words take on aspects of a single entity, that is, a string of words act as a single lexeme with a single meaning'. Such MWUs are prefabricated units, phrases or idioms that may be stored in the mind as stretches and chunks. They are related to functional use, are essential for pragmatic competence and help both reception and production.

One consequence of the communicative focus in language teaching is an interest in the concept of *fluency*, and in what might contribute to fluency in learner language. In addition to temporal variables, vocabulary does play an important role, but not in terms of single words. According to David Wood (Wood 2001:574), automatic processing based on multi-word units can explain many aspects of fluency. The multi-word units can be stored as automatized units in memory, and fluency lies to a great extent in the control of large numbers of formulaic language units and sentence stems. As to the teaching of fluency, input is important for automatization of useful clauses, but reading is weak input for spoken fluency. David Wood (Wood 2001) presents an oral fluency course based on formulaic language units. Pawley and Syder (in Wood 2001: 582) and (in Schmitt 2000:101) assert that memorized chunks form a high proportion of a person's total vocabulary.

As to the question of teaching and testing of vocabulary in the vocational courses in our school system, target 6 in the R'94 syllabus states that “the student should be able to *understand and use*¹⁹ English in contexts that are relevant to his/her area of study”, and this will imply acquiring a new type of vocabulary. To be able to communicate easily about a specific vocational field, you need to know the words, but also the specific MWUs in order to use the words pragmatically correct. Norbert Schmitt underlines that in any well-structured vocabulary program there needs to be the proper mix of explicit teaching and activities from which incidental learning can occur (Schmitt 2000:145). This means that in addition to reading texts from the area of study in question, an explicit teaching of words, e.g. tools and processes, will be needed. As to the MWUs, ‘real life’ conversation in the workshop would be of great value.

As discussed above, a negative “wash-back” effect can easily emerge. To make teachers and students see target 6 as important, it has to be included in the assessment – if it is not tested, it will not be taught. If we think vocabulary is important, then it is worth including a vocabulary component in the final assessment to build positive attitudes toward vocabulary study (Schmitt 2000:163). In classroom teaching and learning of vocabulary, there is a continuing role for relatively discrete test items that show whether, and how well, learners know particularly useful words in the target language. If vocabulary assessment is to have a place in the final assessment of an ESP course under the CLT trend, it will be necessary to concentrate on vocabulary in use and to assess the learner’s ability to cope with the lexical demands of receptive and productive communication tasks. Currently however, assessment relies a great deal on test tools that produce global measures of the test-takers’ performance. We are still in need of good performance descriptors and indicators of progress.

There is an ongoing debate about whether vocabulary should be tested in context (Schmitt 2000:173). In principle, this is a good idea, but it often leads to test items that are expensive in terms of time. According to Schmitt, it is difficult to design formats that truly require engagement with context, which is why words still tend to be presented in isolation or in single sentences. The more test writers wish to measure learners’ ability to actually use words in real-world situations, the further the tests have to move towards embedded, comprehensive, and context-dependent tests. The idea of testing vocabulary in context is a reaction against the

¹⁹ My underlining

discrete-point trend, in accordance with the integrative trend. Schmitt underlines the need for teachers to be clear about the purpose of the test before they can decide upon the most appropriate type.

3.3 Presenting a related study

In 1997, Olav Talberg made a study on behalf of the Norwegian Exam Secretariat, comparing 66 samples of locally made VF2210 exam sets from schools in four different counties. The report was never published, but Talberg was kind enough to let me use it as reference material for the present thesis. The exam assessment of English in the vocational areas of study is, as stated above, the responsibility of the county authorities. Talberg's study was done as a contribution to a quality evaluation of the locally made exam tests, aiming to get a picture of the situation and to see to what extent the local exams are in accordance with the centrally given guidelines. The following is a summing up of Talberg's report.

In most cases, the exam tests had been developed for one single group of students. Only very few were meant to cover more than one area of study. In those cases the exam sets had one or two tasks in common, while the last task was directed against a specific area of study, and designed especially for this group of students.

One of the centrally given guideline documents (SUE/Vg-95-005) states that 'In connection with the carrying out of locally made exams it may be appropriate to allow time to the examinees for organized preparation before the exam test is given²⁰'. The reason why so few, only 14 out of 66 tests, have chosen to do so is not clear, but what is obvious, according to Talberg, is that the formulation in the guidelines has not been interpreted as a strict command. Another reason might be that the employment of organized preparation is new to many teachers, especially in the vocational areas of study. This issue will be discussed further in section 5.3.

The report reveals significant dissimilarities in level of difficulty between the exam sets in question. In his conclusion, Talberg points to some tendencies found in the material²¹:

²⁰ My translation

²¹ My translation

- The examinees are only occasionally allowed an organized preparation period, but if they do get one, they are normally allowed two days and a leaflet of 4 – 11 pages. Talberg emphasizes that the Exam Secretariat sees preparation as an essential condition for testing the student's holistic competence.
- The sample of tests do not include much information about the oral part of the exam, but the tendency is that the oral part consists of listening to a recorded text and answering questions about this text, succeeded by a conversation about a topic of which the student is prepared or one he/she randomly selects during the exam.
- The written part of the exam is normally three hours, but is in some few cases shorter.
- The exam sets normally comprise three to four parts, organized in sections.
- *Comprehension* is used about different kinds of tasks aiming to test the examinee's reading comprehension. The most frequent comprehension task is found to be answering questions. 24 out of 66 have chosen this traditional type. About half of the exam sets have comprehension tasks that do not involve writing skills.
- *Short answer tasks* (about 50 words) mainly refer to tasks like writing a letter and/or writing a summary or translate sentences from a text, and are in most cases vocationally oriented. This section often consists of two tasks, where the second task is one of answering questions.
- The last section, the *long answer task* is normally some kind of composition task, and the examinees get three to five prompts to choose from. The most common wording is 'Write/Tell about...'. The study shows great variety in recommended number of words for this task, but most often the minimum is 150.
- The different sections are generally organized according to topic. The text and the first task may have the character of 'general English', being related to target 5 in the syllabus, but it might just as well be related to a specific area of study in general and/or specific subject matters within the area (target 6). This goes for the other sections as well.

On basis of these tendencies, Talberg has made an evaluation of whether or not the exam tests are in accordance with the aims and targets of the syllabus and the centrally given guidelines. He concludes that the exam sets in this sample meet the specifications *to a reasonable degree*. However, the study illustrates that schools do take local considerations, demonstrated by the variation in amount of work and level of difficulty. A significant variation is found among exam sets made in different schools. Talberg sees this as an argument for regionally made exams, but in such a way that several schools collaborate on making exam tests for each specific area of study to meet the need for a national and uniform standard of the exam sets. He underlines the importance of making tasks 'open' in order to allow all examinees to solve them at their own level.

Some exam sets have a collection of short answer tasks instead of the 'essay task'. The extended use of short answer tasks is, according to Talberg, probably due to the fact that 'composition writing' is known to be unfit as a test item for some of the students in question. ESP teachers experience repeatedly that weak writers may be capable of producing a relatively large amount of English if they are allowed to write three short answers instead of one long answer. Therefore, the exam sets should be designed in such a way that the examinee may choose between these two options, Talberg claims. It is, of course, also important to provide clever students aiming at good grades with tasks suitable to show ability to handle more advanced writing tasks i.e. discuss and/or evaluate. This will be discussed further in section 8.

Finally, Talberg finds it surprising that so few of the writing prompts can be said to be realistic or 'in touch with real life'. He found that little or no information is given to the students about the writing situation, and he admits that this may result from a lack of information to the schools on the part of the Exam Secretariat. Example material showing how such writing tasks may look like have been developed and distributed, but it seems that too little time has been allowed and preferred for discussion of the topic among teachers and test makers in the different schools.

His main conclusion is a warning against letting locally made exams get too 'local', and he recommends that the county authorities to a larger degree call on the schools to cooperate both when making the exam tests and when the test results are to be assessed. At the same time, he underlines the importance of avoiding too strong top management on the test making. A comparison between the results of Talberg's study and mine will be found in chapter 8.

4. The testing of reading comprehension²²

4.1 Theoretical reflections on assessing reading

This section will focus on the testing of reading comprehension by examining text difficulty, text length and the kind of tasks given in relation to the texts included in the exam sets. The main aim is to find out if the assessment of reading comprehension in the VF2210 exam tests really gives the students a chance to perform to the best of their ability.

J. Charles Alderson, probably the world expert on the assessment of reading, points out the need for thorough knowledge about what reading comprehension really is in order to inspect the process and/or the product of a learner's reading. Unfortunately, there is no common agreement on what such a construct should be like (Alderson 2000:111). In our situation, target 2 of the R'94 syllabus to some extent defines the construct for us. It is clearly the *product* of reading that is to be assessed. According to the sub-targets 2a – 2e below we want to measure the students' ability to understand various kinds of texts, to find out to what extent they are able to obtain relevant information from these texts etc. But how do we test reading comprehension and nothing but this specific skill?

Target 2 Comprehension of written English²³

The pupil should become skilful at understanding various types of English texts, according to their nature and structure and the purpose for which they are read.

Main points The pupil should

- 2a be able to understand written presentations of general topics
- 2b be able to understand the main content of texts typical of his/her area of study...
- 2c be able to understand important subject-related texts in detail
- 2d obtain relevant information from a text according to various needs
- 2e understand the message and grasp the significant features of a text

As we can see from the syllabus extract above, the receptive mode of target 2 is underlined by specifying 'understanding' as the main competence. In the first place, we need to define what is meant by 'understanding' a text. Aud Marit Simensen (Simensen 1998:141) defines three levels of text comprehension. The first level is a plain understanding of the surface meaning of the text, 'the lines'. Second level involves an understanding of implications and inferred

²² This chapter is based on a paper written by me in the spring term 2005 (Thorenfeldt 2005)

²³ Quoted from the English version of the R'94 syllabus for English

meanings, the ‘between the lines’-information, while the third level requires an understanding which implies an evaluation of the text on basis of personal knowledge and experience, i.e. the reader’s critical evaluations ‘beyond the lines’.

These levels are reflected in different types of questioning in relation to texts, one assumed to be more demanding than the other. Pearson and Johnson (discussed in Alderson 2000:87) identify three types. Textually explicit questions are those where both question information and the correct answer are found in the same sentence. Textually implicit questions require respondents to combine information across sentences, while script-based questions require readers to integrate text information with their background knowledge since the correct responses to the questions cannot be found in the text itself. The level of information being required by a question – local or global – is clearly crucial to determine the difficulty of test items.

Textually explicit questions are those where both question information and the correct answer are found in the same sentence. Textually implicit questions require respondents to combine information across sentences, while script-based questions require readers to integrate text information with their background knowledge. If, at all, we are to make use of comprehension questions, we clearly need to be aware of these different levels of difficulty when making them. Even in production-independent test tasks like e.g. multiple choice items, we need to be aware that less proficient readers often are ‘word-bound’ (Alderson 2000:Ch.3), and consequently will have far more trouble finding global than local information in a text.

According to Bloom’s taxonomy (discussed in Weigle 2002: 63), which has obviously been quite influential on Norwegian guidelines for assessment²⁴, literal level is considered somewhat ‘lower’ in the hierarchy than critical understanding. Alderson (Alderson 2000: Ch.1) emphasizes that the empirical justifications for value judgements between the levels of reading in foreign language education is very slim. He even questions whether the ‘between and beyond the lines’ types really test reading comprehension at all, since they rely on problem-solving abilities and information outside the text. He points out that we should endeavour to keep ‘reading’ separate from ‘reasoning’ when choosing test techniques.

²⁴ Reference to this taxonomy will be found in most guidelines for assessment, the so-called ‘sensorveiledninger’ following the exam sets.

Consulting theory on the assessment of reading (Alderson 2000:Ch.7), (Simensen 1998:ch.6), we find that there are several techniques for testing reading comprehension without simultaneously involving and testing the production skills. Among the more common and often used we find:

- Multiple choice items, of which there are many different types
- Validating a series of statements in relation to a text (true/false or yes/no/doesn't say)
- Connecting tasks of different kinds/matching lists or phrases
- Choosing from a 'heading bank' for identified paragraphs/sections of a text
- Identification of writer's view/attitudes/claims by tick-off tasks (yes/no/doesn't say)

In addition, we find other types of tasks requiring just a minimum of production:

- Fill in tasks like Cloze-tests (with few gaps, many gaps or gaps + multiple choice items)
- Short-answer questions
- Sentence- or table completion

An interesting approach is that of multiple methods or techniques employed on any one passage of a text. According to Alderson, it is now generally accepted that it is inadequate to measure the understanding of a text by only one method, and that good reading tests are likely to employ a number of different techniques. Alderson even suggests that when test-takers share a first language, like they do in most Norwegian schools, it might in fact be better to ask the questions in that language. This view is not in line with the idea of the monolingual classroom (see section 3.2).

For a long time there has been an issue whether or not to use authentic texts in foreign language reading. If choosing to use authentic text without adaptation, we have to make sure that the material is not linguistically or conceptually too difficult (Alderson 2000:Ch.2), and not alienated from its original context. To create a fair testing situation, it is important to aim at texts on topics that are known to be equally familiar or unfamiliar to all candidates (Alderson 2000:103). Most studies of reading tests show, according to Alderson, that the choice of text had a marked effect. This might be due to greater or lesser readability of texts or it might be a content effect. The influence of the question set and the passage on which the comprehension is assessed are equally important, and the difficulty of a reading test also depends upon the relationship between text and items. The use of pretextual activities

(Simensen 1998:Ch.6) related to the topic of the texts may be used for this purpose. This will be further touched upon in relation to organized preparation in section 5.3.

In the next section I will discuss briefly the concept of readability and give a presentation of two readability indexes which I have chosen to use when ranking the texts, before analysing and characterizing texts and comprehension tasks found in the exam sets.

4.2 Estimating readability

Texts may be more or less accessible when it comes to readability. Researchers have long been concerned to identify what features make text readable, in order to adjust text difficulty to the intended readership, and this has been especially important in educational contexts. (Alderson 2000:71). According to Alderson, syntax complexity and lexical density are very important in this connection, but it is clearly not very practical to have to analyse texts for such features, so indices have been developed to allow rough estimates, i.e. the two readability indexes presented below.

4.2.1 The LIX index

LIX is a concept which has been developed to facilitate the process of measuring the readability of texts. It was developed in the 1960s by pedagogy researcher Carl-Hugo Björnsson, and is still one of the most popular readability index systems. The name LIX is an abbreviation of the Swedish word 'läsbarhetsindex', meaning index of readability.

Readability is defined as the total of such linguistic qualities in a text that make it more or less difficult to access for the reader. Typographical and other external qualities are not included. According to this definition readability means about the same as linguistic level of difficulty. But according to Björnsson (Björnsson 1968:Ch.10) this does not imply that easy means readable or that a difficult text is bad. A difficult text may be eminently readable, provided it is read by a qualified reader. From this perspective only such texts are bad that are not appropriate for the reader.

LIX gives factual information that anybody can obtain by a set of calculations. LIX is defined as the number of words in each sentence + the percentage of words longer than six letters:

$LIX = Lm + Lo$, where Lm is the average number of words per sentence, and Lo is the percentage of words longer than six letters. You start by counting the number of words altogether in the text. Then you divide this number with the number of sentences (= Lm). Then you count how many words there are in the text that are longer than 6 letters. You divide this number by the number of words altogether, and multiply by 100 (= Lo). Finally you add $Lm + Lo$, and get LIX.

The LIX results may be divided into different levels as shown below. I have not been able to find examples of each level²⁵ of the LIX scale listed anywhere, but according to Björnsson, children's books will normally have LIX less than 25, and text of law above 55-60. LIX above 50 is considered to be a difficult text to read for most people. Björnsson underlines that one should not attribute much importance to the differences or deviations in one or two units in LIX, and he points out that LIX is only one detail in an entity estimate.

LIX index

- 25	very easy
26 – 35	easy
36 – 45	medium
46 – 55	difficult
56 -	very difficult

To estimate the readability level of the exam texts in this study, I have subjected them both to a LIX test and an on-line FOG²⁶ test. In addition, I have compared text length, content and legibility²⁷, and finally examined the tasks given to the students in relation to the texts. By this, I intend to make possible a comparison between the test samples in a fairly reliable way.

4.2.2 The FOG index

In order to increase the reliability of the present study, I have compared the LIX results to the results of the FOG index, which is especially designed for English texts. The FOG Index (Alderson and Urquhart 1984, in Alderson 2000:71) is a readability test devised by the American Robert Gunning, now professor of English at Oxford University, in 1952²⁸. The name FOG is probably an abbreviation of 'Formulae of Gunning' but I have not been able to get a confirmation of this. In 1944, Gunning formed Robert Gunning Associates, the first

²⁵ Level of readability is based on L1 reading

²⁶ See 'The FOG index' below

²⁷ The term *legibility* is defined as referring to typographical and other external qualities

²⁸ Source: <http://www.impact-information.com/impactinfo/newsletter/plwork08.htm>

readability consultants. His firm worked to improve the readability of the news bulletins from organizations like the United Press, The Wall Street Journal, and Newsweek. Along with the extensive work done by Rudolf Flesch, Gunning's firm improved the readability of newspapers by bringing them down from the 16th to the 11th-grade level, where they remain today.

The Gunning FOG Index gives the number of years of education that a reader hypothetically needs to understand a definite paragraph or text. It implies that short sentences written in plain English achieve a better score than long sentences written in complicated language. Gunning's FOG formula is easy to apply and to remember:

1. You start by finding the average length of sentences in terms of words in a passage, that is, divide the number of words by the number of complete sentences.
2. In the same passage, find the percentage of polysyllabic words, meaning the words with more than two syllables. That is, divide the number of words with more than two syllables by the number of total words in the passage. Do not count as polysyllabic words that are combinations of short easy words like bookkeeper or butterfly. Do not count verb forms that are made polysyllabic by the addition of -ed or -es like "created" or "trespasses".
3. Add the results of 1. and 2. Multiply the sum by 0.4. The result is the FOG. If a passage has a FOG of 12, this means the level of difficulty requires at least a 12th-grade reading skill. According to this, 12 on the index will correspond to graduation level in the Norwegian upper secondary vocational school.

Fog	index:	Examples:
<- 12	Easy	<i>The Reader's Digest</i> , most popular novels, <i>Time</i> , <i>Newsweek</i> , popular magazines
13 - 16	Medium	<i>The Guardian</i> , <i>The Times</i> , technical documentation, professional prose
17 - 20	Difficult	Technical journals, professional prose
20 ->	Very diff.	Books of law

4.3 Analysing and characterizing texts and comprehension tasks

This section will sum up the findings from my analysis of the assessment of reading comprehension in the written part of the exams. I have studied the texts on which the assessment is based, and the comprehension tasks given in relation to these texts. The tests fall into two groups: Those testing reading comprehension based on a single appendix text and those testing on basis of leaflets of varying size. For the latter type, there will be no readability measures, as the leaflets all consist of texts on different readability levels.

4.3.1 Reading comprehension based on single texts

The majority of the tests belong here in the first group, where the testing of reading comprehension is based on a single text, in most cases unknown to the students on beforehand, and found as an appendix to the exam test. All the texts in the sample are on general subjects in line with Target 5²⁹.

Text A: Media /Reporting the news

Number of words:	630
Number of "long" ³⁰ words:	165
Number of sentences:	57
Number of normal ³¹ pages:	1.6
LIX:	37 = Medium
FOG:	15
Type of tasks:	Tick out true/false + correcting the false statements
Preparatory material:	No
Production skills required:	To some extent in one out of two tasks
Possibility of showing comprehension without writing:	Yes
Reading 'between/ beyond the lines' ³² required:	Yes, required by one of the items
Style ³³ :	Informative, adapted for educational use. The text has a good structure with 12 short paragraphs and the key-words emphasized. There are not too many low-frequency words, but some are specific to the media field.
Tasks:	In the validation ³⁴ task, all the statements except one can be compared directly to the text even if some of them are textually implicit. The task does not require reading comprehension beyond the first level except for number 6, which requires beyond the lines-knowledge. Four of the statements are false, and consequently need correcting. The corresponding statements from the text are quite easy to find, and

²⁹ Knowledge about the English-speaking world (USA and The Great Britain)

³⁰ "long" words = words with more than six letters

³¹ A *normal page* is defined as having 2400 signs

³² Explained in section 4.1

³³ Comments on style and lay-out are based on my own evaluation

³⁴ By *validation task* I mean 'tick off' tasks like true/false or right/wrong

can be used directly when answering. Since one of the answers cannot be found in the text, there should clearly have been a 'doesn't say' alternative as well.

Text B: Hooliganism – the English disease?

Number of words:	750
Number of "long" words:	186
Number of sentences:	45
Number of normal pages:	1,9
LIX:	42 = Medium
FOG:	19
Type of tasks:	True/false + correct the false statements
Preparatory material:	No
Production skills required:	Yes
Possibility of showing comprehension without writing:	To some extent in one out of two tasks
Reading 'between/ beyond the lines' required:	Yes
Style:	Quite a long magazine feature story divided into four sections and 11 paragraphs. Vocabulary is not too difficult; low-frequency words are explained. Topic: Violence and stereotypes
Tasks:	Validation task of six items, some of them requiring global understanding. Two wrong statements in need of correction, one of them on the first level.

Text C: The Simpsons

Number of words:	377
Number of "long" words:	66
Number of sentences:	25
Number of normal pages:	0,9
LIX:	33 = Easy
FOG:	13
Type of tasks:	Answering six questions in complete sentences
Preparatory material:	No
Production skills required:	Yes, but a lot may be copied from the text
Possibility of showing comprehension without writing:	Not really
Reading between/ beyond the lines required:	No, not really
Style:	Informal, with elements of oral language including some slang words, not too many low-frequency words. An authentic text taken from The Rolling Stone Magazine.
Tasks:	The first five comprehension questions are all on level one (finding information in the text – reading the lines). In fact, so is the last one, even if it is possible to answer it in more depth if you know what the American Dream is all about (i.e. by including 'beyond the lines'-knowledge).

Text D: American students order pizza in class

Number of words:	298
Number of "long" words:	76
Number of sentences:	18

Number of normal pages:	0,8
LIX:	42 = Medium
FOG:	15
Type of tasks:	Answering four comprehension questions
Preparatory material:	No
Production skills required:	Yes
Possibility of showing comprehension without writing:	No
Reading 'between/ beyond the lines' required:	Yes
Style:	Newspaper article from the <i>International Herald Tribune</i> , adapted for educational use with six short paragraphs. The text is short, including elements of direct speech, but does also some few low-frequency words. The classroom situation, cell phones and pizza ought to be familiar.
Tasks:	Only one task: Answering four comprehension questions in full sentences with the examinee's 'own words'. Require comprehension on all three levels, most of them script-based going beyond the lines: ("Can you think of any reasons why schools are less strict about cell phones <i>after</i> September 11?")

Text E: The making of an American heroine

Number of words:	542
Number of 'long' words:	138
Number of sentences:	34
Number of normal pages:	1.4
LIX:	42 = Medium
FOG:	17
Type of tasks:	Eight multiple choice items
Preparatory material:	Yes: leaflet 9 pages/ excerpt appendix to the exam test
Production skills required:	No
Possibility of showing comprehension without writing:	Yes
Reading between/beyond the lines required:	Yes
Style:	Being an authentic magazine text from <i>Vanity Fair</i> , organized in 6 longer paragraphs, formal media language defines the style: many low-frequency words and foreign names. Because of the lexical variation and the Arabic names, this is probably a quite difficult text for Norwegian students.
Tasks:	Test technique employed is completion of 8 sentences by multiple choice items. The task requires comprehension on the second level: textually implicit inference and reading 'between the lines'. The distractors seem to be probable options.

Text F: Jamaican drug mules "flooding" UK

Number of words:	653
Number of "long" words:	183
Number of sentences:	26
Number of normal pages:	1,7
LIX:	53 = difficult!
FOG:	20
Type of tasks:	Ten multiple choice items
Preparatory material:	Yes: Leaflet, excerpt appendix to exam/ may bring and use notes
Production skills required:	No
Possibility of showing comprehension without writing:	Yes

Reading 'between/ beyond the lines' required: Yes

Style: The most difficult of all the texts! Probably a *net paper* article, long and very hard to read! Seems unorganized; sentences are separated into single, very long items with poor coherence and syntactic complexity, extensive use of slang, a heavy lexical load (at least for Norwegian students) and expressions of transferred meaning. Topic: Drug smuggling

Tasks: One task of ten multiple choice items: Students are to choose one out of three alternatives to complete sentences. The distractors are all close to the correct alternative, and the task is quite tricky, requiring very good reading abilities! In my opinion, parts of this task test intelligence as well as reading comprehension. The items are not chronologically sequenced, and require reading between the lines.

Text G: Boy lost university place...

Number of words: 493

Number of 'long' words: 116

Number of sentences: 25

Number of normal pages: 1.2

LIX: 43 = Medium

FOG: 17

Type of tasks: Correcting 7 false statements, give an explanation

Preparatory material: No

Production skills required: Yes

Possibility of showing comprehension without writing: No

Reading 'between/beyond the lines' required: Yes

Style: Newspaper article from the *Daily Telegraph*, slightly adapted and of medium length, divided into three paragraphs. Two expressions are explained. Not particularly easy, but still comprehensible with the vocabulary linked to schooling, computers and business.

Tasks: Two tasks, both 'real life': '*A friend of yours has read the article and tells you about it, but what she says is all wrong. Correct her statements*'. There are seven statements to be corrected, and that can be done by comparing the statements to the text. In the second task the student is asked to: '*explain briefly to Ann why British Engines decided to support Chris*'. This may be quoted more or less directly from the text.

Text H: A party for the Queen

Number of words: 505

Number of 'long' words: 95

Number of sentences: 30

Number of normal pages: 1,2

LIX: 36 = easy/medium

FOG: 16

Type of tasks: True/false, 10 statements

Preparatory material: No

Production skills required: No

Possibility of showing comprehension without writing: Yes

Reading 'between/beyond the lines' required: Yes

Style: Authentic magazine feature story from *The Week*. Three long compact paragraphs, many quotations. Easy according to the readability indexes, but not necessarily so to the students. Five expressions explained.

Tasks: Validation task of ten items, true/false, some ask for textually implicit information. No correction required.

Text I: Young British 'gappers' embrace their adventures abroad

Number of words:	379
Number of 'long' words:	111
Number of sentences:	21
Number of normal pages:	1.0
LIX:	47 = difficult
FOG:	18
Type of tasks:	Explaining a term and answering a question + give own opinion on matter by listing advantages/disadvantages
Preparatory material:	No
Production skills required:	To a large extent
Possibility of showing comprehension without writing:	No
Reading 'between/beyond the lines' required:	Yes

Style: The LIX index indicates that text I is a difficult text. One page, organized in 10 short paragraphs, but not really a good text to read, as it seems to lack fluency³⁵, probably due to the adaptation. Some low-frequency words occur, but the topic is quite familiar. The text is slightly adapted from '*Globetrotter*', and the layout is ok.

Tasks: In this test battery, the students are given two demanding 'real life' tasks which both involve comprehension at all levels. Both tasks require independent written production, and there is no way of showing whether or not you have actually understood the text without having good production skills.

Text J: London's cockney culture looks a bit different

Number of words:	416
Number of "long" words:	82
Number of sentences:	26
Number of normal pages:	1,0
LIX:	36 = Easy
FOG:	15
Type of tasks:	Write a short summary
Preparatory material:	No
Production skills required:	Yes, to a large extent
Possibility of showing comprehension without writing:	No
Reading 'between/ beyond the lines' required:	No

Style: Newspaper article, slightly adapted from the *Los Angeles Times*. Good lay-out, easy according to readability, 7 paragraphs and a lot of direct speech. Difficult expressions are being explained in the text.

Tasks: One 'real life' comprehension task: Helping a less proficient fellow student by writing a short summary of the text. The catch is you need good writing skills as well as reading skills.

³⁵ By *fluency* (in this connection) I mean the 'flow' of the text, the cohesion and the way paragraphs relate to each other.

Text K: The Media has Power

Number of words:	434
Number of "long" words:	111
Number of sentences:	38
Number of normal pages:	1.1
LIX:	37 = Medium
FOG:	15
Type of tasks:	True/false + text production 70-100 words
Preparatory material:	Leaflet (9 pages), may be used on exam day
Production skills required:	No
Possibility of showing comprehension without writing:	Yes
Reading 'between/ beyond the lines' required:	No, not really
Style:	'Messy' lay-out as seven small texts (or actually mostly paragraphs from the same text?) have been placed randomly on the sheet. Each part is in itself easy to read.
Tasks:	Validation task of 8 items. Tick off true/ false, no correction. Easy to compare to the text.

Text L: Politics

Number of words:	363
Number of "long" words:	67
Number of sentences:	30
Number of normal pages:	0.8
LIX:	31 = Easy
FOG:	14
Type of task:	True/false
Preparatory material:	Leaflet 15 pages, excerpts as appendix to exam test
Production skills required:	No
Possibility of showing comprehension without writing:	Yes
Reading 'between/ beyond the lines' required:	Yes, to some extent
Style:	Copy of text and validation task from an English textbook commonly used in vocational courses. Well organized text, but poor lay-out due to the adaptation of textbook material.
Tasks:	Validation task: right/wrong, 10 items, no correction. Reading on 'between the lines' necessary in some cases.

Text M: The mole people

Number of words:	614
Number of "long" words:	118
Number of sentences:	34
Number of normal pages:	1.5
LIX:	37 = Medium
FOG:	15
Type of tasks:	Answering five comprehension questions + find wrong statements among five and explain what is wrong
Preparatory material:	No
Production skills required:	Yes
Possibility of showing comprehension without writing:	Not really

Reading 'between/ beyond the lines' required: Yes

Style: Authentic magazine text from *Club*, divided into four quite compact paragraphs. Typical 'exam test' layout. Straightforward language, quite easy to but some expressions could have been explained.

Tasks: Two tasks, both involving writing: Five comprehension questions, the last one 'beyond the text', and five statements of which the students are to pick out the false ones and explain what is wrong with them. Two are false, and all are easy to find in the text.

Text N: Maniac Driver in 147mph Case

Number of words: 463

Number of "long" words: 97

Number of sentences: 24

Number of normal pages: 1,1

LIX: 40 = Medium

FOG: 16

Type of tasks: True/false + correct the false statements + answering three questions

Preparatory material: No

Production skills required: Yes, in two out of three tasks

Possibility of showing comprehension without writing: In one out of three tasks

Reading 'between/ beyond the lines' required: Yes

Style: Authentic (?) newspaper article (source not given) of six paragraphs. 'Reporting' newspaper language, sometimes quite difficult because of syntactic complexity and lexical density.

Tasks: One validation task of six items including correction of false statements. In addition three comprehension questions of which one goes 'beyond the text'.

Text O: Gap years

Number of words: 425

Number of "long" words: 87

Number of sentences: 19

Number of normal pages: 1,0

LIX: 43 = Medium

FOG: 20

Type of tasks: Writing an e-mail

Preparatory material: No

Production skills required: Yes

Possibility of showing comprehension without writing: No

Reading 'between/ beyond the lines' required: Yes

Style: Internet text with a photo included, maybe an advertisement for a student exchange organization? Difficult according to the readability indexes, but mostly present tense language. Some expressions could have been explained.

Tasks: 'Real life' task in which the student is to write an e-mail message connected to the text: 'introduce yourself and explain your plans'. This might as well have been a writing task!

Text P: Accommodation in London

Number of words: 320

Number of "long" words: 97

Number of sentences:	19
Number of normal pages:	0,9
LIX:	47 = Difficult
FOG:	18
Type of tasks:	Writing a booking
Preparatory material:	No
Production skills required:	To a large extent
Possibility of showing comprehension without writing:	No
Reading 'between/ beyond the lines' required:	Yes
Style:	Three authentic advertisements for accommodation in London. Difficult to read for several reasons: The second most difficult text according to LIX, many low-frequency words (at least to students) even if they are being repeated, and mostly because of the lay-out with very small letters and vague print.
Tasks:	'Real-life' task of making a booking for a hotel room in London. Test a lot more than reading comprehension.

4.3.2 Reading comprehension based on a preparatory leaflet

In this group, the testing of reading comprehension has been based directly on a leaflet which the students have had time in advance to read through. The leaflets vary in size, difficulty, lay-out, content and genres included. The students may take notes, 2 pages by hand, from the preparation period and bring to the exam.

Leaflet Q: Immigration and minority groups in GB

Number of pages:	6
Type of tasks:	Six multiple choice items + answering three questions
Preparatory material:	Leaflet, one weekend for preparation
Aid during exam:	Leaflet (without notes in it) available + one single sheet of notes from the preparation period
Main topic:	Immigration and minority groups in Great Britain
Production skills required:	Yes
Possibility of showing comprehension without writing:	Yes, in one out of two tasks
Reading 'between/ beyond the lines' required:	Yes
Style:	Beautiful lay-out, clear print in colour, enough space between the different texts. Tempting to read, mostly straight forward textbook language. Different genres, but mainly factual information. Photos and cuttings from magazines, and adapted internet-texts.
Tasks:	The first task implies completion of sentences in six multiple choice items. Most of them test global understanding, as the same answers may be found in several of the texts. In the second task the students are to answer three comprehension questions, of which two go 'beyond the lines'.

Leaflet R: Typically American

Number of pages:	4
Type of tasks:	Writing three different kinds of statements

Preparatory material: Leaflet, one weekend for preparation
 Aid during exam: Handwritten notes from the preparation period
 Main topic: Typically American
 Production skills required: Yes, to a large extent
 Possibility of showing comprehension without writing: No
 Reading 'between/ beyond the lines' required: Yes
 Style: Blurred text (scanned magazine article?) in black & white. One single long text (three pages of which one is reduced in size to fit in) discussing American society. Difficult to read because of the lay-out, the language is OK. In addition we find one page of photos connected to American culture.
 Tasks: Not a reading task at all, but a test of background knowledge. The students are asked to write three sentences expressing a positive view of Americans, three expressing negative views and discuss the role of religion in the life of Americans. The leaflet is not available during the exam!

Leaflet S: Terrorism

Number of pages: 15
 Type of tasks: Completing sentences through five multiple choice items + making questions to three answers
 Preparatory material: Leaflet, two days for preparation
 Aid during exam: Leaflet (without notes in it) available + handwritten notes (one sheet) from the preparation period
 Main topic: Terrorism
 Production skills required: Yes, in one out of two tasks
 Possibility of showing comprehension without writing: Yes, in one out of two tasks
 Reading 'between/ beyond the lines' required: Yes
 Style: 15 pages in black & white, OK lay-out, many different genres: Feature story, two caricature drawings, letters to the editor, President's remarks, a newspaper article, factual texts from textbook and internet, memory plate, poem.
 Tasks: Five multiple choice items relating to one specific, quite difficult feature story from net paper *The Nation*. Language seems difficult according to readability as well. Require good reading skills on different levels. In the second task the student is asked to make questions to five given answers.

Leaflet T: Two American heroes

Number of pages: 6
 Type of tasks: Answering four comprehension questions
 Preparatory material: Leaflet, one day for preparation
 Aid at exam: Leaflet available + handwritten notes (two pages) from the preparation period
 Main topic: Two American heroes (King + Jordan)
 Production skills required: Yes
 Possibility of showing comprehension without writing: No
 Reading 'between/ beyond the lines' required: Yes
 Style: Good readable lay-out in black & white, two factual texts and the lyrics of a song. Language seems OK according to my experience with readability testing.
 Tasks: Answering four questions, one of them 'beyond the lines'

Leaflet U: My culture, your culture

Number of pages:	4
Type of tasks:	Making five questions/answers + three explanations of at least two sentences
Preparatory material:	Leaflet + Video: 45 min, one day for preparation
Aid during exam:	Leaflet available + handwritten notes (two pages) from the preparation period
Main topic:	'My culture – your culture'
Production skills required:	Yes
Possibility of showing comprehension without writing:	No
Reading 'between/ beyond the lines' required:	yes
Style:	OK lay-out, black & white, many photos and illustrations. Mostly amusing and factual texts, comparing Norwegian and American culture. Not too difficult according to my experience with readability testing, even if the first text is a bit 'unorganized' and informal.
Tasks:	Two 'real life' tasks: The reading comprehension task is making a "Trivial pursuit"-card with questions + answers about the USA (within five categories, facts from the texts) + a listening/viewing comprehension task using information from a video: make three choices and explain why on each of them. Requires reading on the first level only.

4.4 Discussion of the findings relative to reading comprehension

The accuracy of these results cannot be completely guaranteed for. The relative difficulty of a text is due to a lot more than what is included in a readability index; indexes show symptoms, but they do not explain the whole difficulty of the text. In addition to lexical density, we need to take lexical variation and lexical sophistication into consideration (Schmitt 2000:Chap.9), and there is still much more to the picture. Alderson draws attention to text content, familiarity with the topic and the vocabulary, text type/genre, text organisation, sentence structure, lexis, text typography, layout and the relationship between verbal and non-verbal text as factors that might influence the reading process (Alderson 2000:Chap.2). He concludes that the more concrete, imaginable and interesting a text is the more readable.

Since the calculation methods do not offer ways of handling all the elements Alderson call attention to, variations will occur when using different tools. I have, as mentioned above, based my study primarily on the LIX index. As far as I've been able to find out³⁶, a LIX score between 30 and 40 seems to be the most suitable for 17 years old students in EFL vocational classes. Eight of the texts in the sample are within this range. Three texts are far too difficult according to readability, and the rest is difficult, but acceptable because of familiar topics and

³⁶ By comparing to the FOG-index. See section 4.1

simple syntax. The authentic ‘news reporting language’, characterized by syntactic complexity, lexical density, omissions and abbreviations, differ a lot from the well organized texts and the more or less simplified language the students are used to meet in their textbooks, and stand out as especially difficult. Actually, according to the FOG index, all these texts are too difficult. The web sites³⁷ suggests a FOG between 11 and 13 as suitable, while the exam set texts all score between 13 and 20.

In the sample material we find merely authentic texts, but most of them are adapted for educational use. Being a relatively high-stakes test, we may ask if the difference in readability between the texts will have any consequences for the students’ test result. The following categories have to be seen only as indications as they are based solely on the indexes, but they do give some clues about a relative ranking between the texts in question.

Table 1: Ranking of texts according to the readability indexes:

Category 1 - Easy:	Text L	LIX = 31	FOG=14
	Text C	LIX = 33	FOG=13
	Text J	LIX = 35	FOG=15
.....			
Category 2 - Medium:	Text H	LIX = 36	FOG=16
	Text K	LIX = 37	FOG=15
	Text A	LIX = 37	FOG=15
	Text M	LIX = 37	FOG=15
	Text N	LIX = 40	FOG=16
.....			
Category 3 - Difficult	Text D	LIX = 42	FOG=15
	Text E	LIX = 42	FOG=17
	Text B	LIX = 42	FOG=19
	Text G	LIX = 43	FOG=17
	Text O	LIX = 43	FOG=20
.....			
Category 4 - Too difficult:	Text I	LIX = 46	FOG=18
	Text P	LIX = 47	FOG=18
	Text F	LIX = 53	FOG=20

³⁷ I have in this connection among other web-sites accessed (www.usingenglish.com)

As it appears, the testing of reading in the VF2210 final exams is typically done in one out of two ways. Either they refer to a preparatory leaflet³⁸ which the students are supposed to have studied on their own on beforehand, or they refer to a text in appendix, either to some extent familiar from pre-reading activities or completely unknown to the students. Tests based directly on the leaflet will not be analysed in detail here, but they do add to the picture of unequal conditions for the students in question. The leaflets vary in size from 4 to 15 pages, some requiring quite extensive reading, and even if they create an inspiring foundation for good readers, they represent just another disadvantage for weak readers. Being dependent on the student's own preparation at home, they tend to increase the differences. To my experience, some will get all the help they need during the preparation period, while others have no one to ask and maybe not even a proper environment to work in, so they just resign and drop the preparation altogether, leaving them unable to cope with this part of the final exam test. I will return to this issue under section 5.3

In most cases, however, the testing of reading comprehension is based on a one page text on a current general subject with some kind of reading comprehension task related to it. Basically just two of the main points of target 2 in the syllabus, 2a and 2d, are subject to testing. Keeping in mind that EFL vocational classes normally spent a lot of time reading subject-related texts and learning vocational vocabulary³⁹, it is surprising to find that none of the main points of target 2 are being tested. Not one single example is found of using a vocationally related text as a starting point for the testing of reading comprehension even if the National Exam Secretariat encourages such a practice (Talberg 1997:13).

There is a striking variety in text difficulty among the different tests, and a considerable variety in text length in such a way that the most difficult texts are also the longest. No traceable correlation is found between level of difficulty and the use of authentic material. There are also significant differences in the demand for written production in the test tasks, and furthermore in the level of reading comprehension required from test to test. Among the 21 exam sets I have studied, only *four* of them test reading comprehension without involving written production skills. Another six have both kinds of tasks, while as many as *twelve* provide no possibility to show comprehension without producing written answers.

³⁸ See section 3.2

³⁹ See also p.72

The table below lists the different kinds of task types found. Some of the exam sets combine two or three types:

Table 2: Different types of comprehension tasks

Type of tasks	Frequency
Comprehension questions	7
Multiple choice tasks	4
Validation task: True/False	6
Correcting false statements	5
Giving explanation	3
Writing a short text	5
Summary	1
Making questions to given answers	1
Making 'Trivial Pursuit'-cards	1

According to Alderson⁴⁰, very few of the test tasks assess reading comprehension exclusively. Either they require good production skills or they test reasoning and problem-solving abilities in addition to reading, or both. Test techniques found are completion of sentences by 'multiple choice', validation tasks in the form of 'true/false' (with correction of the false statements), explanation of expressions from the text, comprehension questions, stating of main ideas and writing of short paragraphs inspired by the text. Obviously the students' ability to express thoughts in writing is taken for granted, and consequently we take it that it doesn't matter in which way we test their reading comprehension. Experience shows something else.

In my opinion, the students deserve the possibility to show their reading comprehension without involving too much written production, as their productive skills are being tested fully in the remaining sections of the exam test. As discussed above, there are many good ways of doing this, and an awareness of the problem among those making the tests might be all that is needed.

⁴⁰ See section 4.1

5 The testing of writing

5.1 Theoretical reflections on assessing writing

Writing has become an essential tool for people on all walks of life in today's global community. It is now widely recognized that writing plays a vital role not only in conveying information, but also in transforming knowledge to create new knowledge. In school settings, the assessment of writing ability is of central importance. But according to J.C. Alderson, author of the preface to Sara Weigle's *Assessing writing* from 2002, it continues to be one of the most problematic areas of language use to assess, primarily because of the subjectivity of the judgements involved.

Due to the Communicative Language Teaching paradigm, writing in second- and foreign language learning has gone from functioning primarily as support and reinforcement of learning in general to becoming important as a skill in its own right (Weigle 2002:1).

Whenever the acquisition of a specific language skill is seen as important, it becomes equally important to test that skill, and writing is no exception. Weigle claims that as the role of writing in L2 education increases, there is an even greater demand for valid and reliable ways to test writing ability. In addition to the main aim of target 4, 'Use of written English', we find the following sub-targets:

Target 4⁴¹	The pupils should
4a	be able to take written notes, for instance of key points in oral or written sources, and to report the main contents of written texts, talks, lectures etc
4b	be able to express his/her own attitudes and opinions in writing
4c	be able to write various kinds of texts, such as stories, formal letters, applications, reports, or passages of continuous reasoning.

What does it mean to test writing ability? According to Weigle, a test of writing involves at least two basic components: one or more writing tasks and/or instructions that tell test takers what to write, and a means of evaluating the writing samples that test takers produce.

However, designing a good test of writing involves much more than simply selecting a topic for test takers to write about and then using our own judgement to rank order the resulting writing samples. There are several key questions that need to be considered (Weigle 2002:2):

⁴¹ Quoted from the English version of the R'94 curriculum

- What are we trying to test? That is, how are we defining our construct?
- Why do we want to test writing ability?
- Who will score the test, and what criteria or standards will be used? How can we ensure that raters apply the scoring standards consistently?
- Who will use the information that our test provides?
- What are the constraints (of time, materials, money and labour)?
- What do we need to know about testing to make our test valid and reliable?

To assess language ability in a writing test we must also consider what role background knowledge plays in the test. Should topical knowledge be part of the construct? Bachman & Palmer (in Weigle 2002:45) bring forth three basic options when defining the construct. One option is to specifically exclude topical knowledge, another is to including both language ability and topical knowledge, and the third is defining language ability and topical knowledge as separate constructs. VF 2210 has, as already established (section 2.5), targets of both kinds, and including both seems to be the preferred option.

From psychological research we know that human beings have a definite amount of information processing capacity (Simensen 1998:85-86). In L2 writing, depending on level of proficiency, it may be necessary to use more cognitive resources on questions of language than in L1 writing. Lengthy searches for appropriate lexical and syntactic solutions are time-consuming, thus less attention can be given to content. As stated above (section 2.5.1), VF2210 allow from 1 to 3 hours on the written part and the guidelines emphasize the importance of giving the test takers time for reflection. Assessors need to keep in mind these limitations when evaluating length and content of the written products. In my experience, lots of students find it really hard to generate content 'as ordered'. All aspects of interaction are missing in a test situation like this, and generating content in the absence of a partner is a formidable obstacle to many students. Providing them with a leaflet on the topic and time for cooperation will for some solve the problem, but will for others create new challenges. I will return to this issue below in section 5.3.

Other important guidelines for test making can be found in Bachman & Palmer's 1996' framework for test usefulness, where they list six principles: reliability, construct validity, authenticity, interactiveness, impact and practicality (in Weigle 2002:48). The choice of testing procedures involves finding the best possible combination of the six qualifications

listed, and deciding which qualities are most relevant in a given situation. Weigle draws attention to the fact that the existence of a scoring rubric distinguishes *holistic* scoring from its earlier, less reliable predecessor, *general impression marking*. Good practice in writing assessment involves the existence of a rubric, benchmark scripts and rater training, according to Weigle. Finally, she sees it as rather important to develop tasks that are appropriate for the specific examinee and to avoid biasing test tasks either in favour of or against test takers who have certain characteristics (Weigle 2002:46). In this connection, ‘characteristics’ apply especially to different areas of study.

The importance of defining the construct before testing has been accounted for above. For VF2210, the construct is to a large extent open for discussion, and have to be defined by each individual test maker. However, some features should be easy to agree on: VF2210 should aim to test the students’ ability to use written English for taking notes, reporting main content from different sources, expressing opinions and writing different kind of texts, all under influence of the communicative competence concept, including at least linguistic, sociolinguistic, strategic and discourse competence. Bloom’s taxonomy is recommended as scoring reference. Content is to a certain extent meant to be included, not as requirements for specific facts, but in the form of ‘open’ tasks where the student may include their own specific knowledge in relation to target 5 and 6 of the R’94 syllabus.

The reason why writing ability is being tested here is to meet the need for exam assessment of an upper secondary course in English as a foreign language. The score from the written exam is finally to be included in a single holistic grade. The test will be scored, or rather judged subjectively, by two teachers under time pressure, and the criteria are vague and open to interpretation. To ‘ensure that raters apply the scoring standards consistently’ is not an easy thing to do under the current practice. According to Weigle (2002:59) interrater reliability relies on using a scoring rubric along with model essays, but when the prompts differ like in VF2210, this will not solve the problem. A scoring rubric does in fact exist (SUE/Vg 001-97), but it is not very precise. In this rubric, three categories are defined for the rating,;: *Contents*, *Structure* and *Language*, and a rating scale consisting of four major levels is specified according to a set of objectives: The levels are *Above average*, *Average*, *Below average* and *Objectives not met* (discussed in Simensen 1998:273) The official guidelines of R’94 refer explicitly to the objectives of the course as the criteria for evaluation.

The exam result from the written part will be modified by the oral part of the exam and added to the student's 'Upper Secondary Education Diploma' as a signal to potential future employers and educational institutions of how well the test taker can communicate in English. Taking current thinking about testing writing into consideration (Weigle 2002), it is not easy to make this test valid and reliable. At the moment, the double marking is the best guarantee for reliability.

5.2 VF2210: Prompts and writing instructions

Are we simply 'thinking of a topic' for test takers to write about when making the prompts? VF2210 is a direct writing test, as test takers actually produce a sample of writing. In the following, I will list examples of prompts and writing instructions from last year's examination and take a closer look at wording and genres. Unfortunately, an examination of the students' writing samples, the scoring procedures and the resulting grades will have to remain for further studies.

Table 3: Writing prompts

Formulation	Frequency
Write a letter to a friend telling...	4
Write a (formal) letter to.../ an application for a job...	5
Write about a problem..., / Write a comment .../ Write a story...	11
Write / tell about a person...	12
Write about (topic related to the USA or The Great Britain)...	25
Write about / describe (school, vocational training, future job)...	7
Write about (your future plans/planning a journey)...	5
Write a summary of / Tell about (a film, a book, a poem, a song)...	7
Discuss... / Give your opinion...	11
Write a text / story based on pictures...	4
Why learn English...(discuss)	2
Write a conversation / write a speech...	2

It is of course impossible to give a complete report of all the formulations of prompts found in the material. This table is thus only a record of the most frequent prompts and of some of those that only rarely appear. Many of the prompts fall into more than one category (e.g. write a letter to a friend telling about the USA). As was the case in Talberg's report from 1997, we

find that most prompts have the wording ‘Write about’ or ‘Tell about’ or something similar, directly or indirectly asking the student to show and reproduce knowledge on a given topic.

Talberg found that ‘Write a letter to’ was a frequently used prompt, but this seems not to be the case in this sample. I found only four personal letters among 91 prompts. He characterized the tests as being short of ‘discussion’ and ‘evaluation’ tasks, but that is not my impression: Only two out of 21 exam sets leave out this category. Nine have both a ‘discuss’ and an ‘evaluate’ task, and twenty have at least one ‘evaluate/explain/your opinion’ task included. Talberg also commented on the lack of realistic or ‘real world’ tasks, a principle which has been promoted strongly by the National Exam Secretariat. He defines ‘real world’ tasks as giving the examinees a frame or a situational description under which they are supposed to write. Among the 21 exam sets in my study, attempts are made in nine of them of doing so, at least to some extent. I will quote one writing instruction of this kind as an example:

IMMIGRATION

To the student:

‘You are a participant in an English course in London together with students from many countries.

PART 1

Not all of the students understand English as well as you do, and many of them do not have the same knowledge about immigration as you have.

From your teacher you have got the text *London’s Cockney Culture looks a Bit Different* (see Appendix A), which you are to discuss in groups the next day.

One of the other students understands very little of the text and asks you to write a short summary.

Of course you help your fellow student!

PART 2

The same day you also learn that a British author, John Wilson, around 1850 wrote that *the sun never sets on the British Empire*.

In your opinion, what did he mean by this statement?

PART 3

The last day of this successful course in London you have to show that you have gained a good command of written English and are asked to write a text.

Choose one of the following tasks:

- Write a text for your English teacher about the UK.
- Why is it necessary for you as a Norwegian to learn English? How can it be useful in your future career?
- A day in London.
- personal text based on quotation: ‘Who are the British?’

However, the most common writing instructions found in the exam sets studied are well known from before the Reform – asking the students to ‘write about’ something without reflecting on why and to whom :

- Write an essay on one of the following topics
- Write an essay about one of the following subjects
- Write an essay of at least 150 words on one of the following topics
- Write one of the following essays
- Answer one of the following tasks
- Choose one of the following topics and write an essay

There is probably more than one explanation as to why ‘traditional’ writing tasks still seem to dominate. Talberg points to habit, convenience and practicality. In my opinion, the main reason is that most teachers are in fact not familiar with the official guidelines. Talberg underlines the importance of discussing and concretizing the test examples provided by the Exam Secretariat in classrooms as well as in the language sections at each school, but the reality in many vocational schools nowadays is that the language sections are closed down and the teachers are organized in cross-curricular teams. The expected discussions, the consciousness-rising and the subject-related cooperation among teachers have to some extent ‘fallen out of style’ at the moment; at least this is the case in my school. Talking to language teachers in other upper secondary schools, this seems to be the current tendency.

5.3 Organized preparation

The idea of letting the students prepare themselves for the exam on beforehand has a central role in R’94. In one of the official guidelines (SUE/Vg-95-005) we find that ‘in connection with the carrying out of locally made exams it may be appropriate to allow time for organized preparation before the given exam test⁴²’. An additional information letter (SUE/Vg-97-021) underlines⁴³ that ‘to be able to collect and evaluate new information, take good notes and to cooperate are all part of the student’s holistic competence’, and is thus seen as an appropriate part of the exam. The purpose of a preparation session is, among other things, that the students in advance should become familiar with the topic and/or the given approach, have the

⁴² My translation

⁴³ My translation

possibility of discussing it with others, and to obtain relevant information needed for the individual exam test. In CLT, an ideal is to make students perform on basis of several types of input. Pretextual activities are supposed to ensure that previous knowledge and experience are being activated (Simensen 1998:142+244), and this may help eliciting ideas and content for writing tasks.

In his study, Talberg found that only 14 out of 66 have chosen to include organized preparation. The results from my study show a somewhat higher percentage: 9 out of 21 tests include preparation. In vocational classes, as in many other classes, we find great variety in reading ability, but the poor readers tend to be over-represented⁴⁴, and so are students with no hope of getting any help at home. Good readers may perhaps profit from this kind of written input, but as mentioned before (sections 4.1 and 5.3), the preparation session may represent an additional obstacle to one group in particular, namely those suffering from dyslexia or other severe reading difficulties.

To support the assertion above, I examined the list of students in the Building and Construction courses at my school with respect to documented reading problems, and found that students with statutory rights to special adaptation counts for more than 25% of the lot. For these kids, a 15 pages leaflet and two days for preparation is just another disadvantage. They are, as already mentioned, according to the current system entitled to special adaptation of the test material in the form of i.e. recorded texts, but in practice this will not always be provided for, and even if it is, they will normally not succeed in becoming fully familiar with the texts and thus not prepared to solve the comprehension tasks.

Preparation at school with the teacher at hand, e.g. in the form of a group discussion or oral presentations and conversations, would perhaps improve their situation, but they will still fall short when it comes to taking notes. Poor readers are more often than not poor writers as well, and finding the essence of a text for making good notes is, to my experience, often far beyond their capacity. As we see it now, preparation sessions the way they seem to be practiced just add to increasing the gap between the good and the less proficient readers.

⁴⁴ See the next paragraph

6. Summing up the written part of the exam test

As already stated in the introduction (section 1.3), my main research method has been to examine a sample of 21 exam sets. This section will sum up the findings from my analyses of the written part of the exam tests in total, and will thus include data from chapters 4 and 5. The following paragraphs consist of brief, preliminary descriptions of the 21 sample tests with respect to organization, preparation, extent, readability, types of test tasks, and targets assessed, followed by tables which sum up the findings for each category. The more comprehensive analyses and discussions are found in chapter 8 below.

6.1 Brief descriptions of the tests⁴⁵

Test A

Time available:	2 hours
Preparation period:	None
Preparatory material:	None
Vocational subject:	Compulsory included
Main topic:	News reporting
'Real life' tasks:	No, except for two of the writing prompts
Tasks: 3	Reading comprehension: 8 true/false-statements + correct the false statements. 4 questions to be answered in English about the student's area of study Written production: An essay of at least 150 words - 6 choices (all essays)
Text:	617 words, LIX 37, FOG 15
Targets assessed:	2, 4, 5, 6

Test B

Time available:	2 hours
Preparation period:	None
Preparatory material:	None
Vocational subject:	Compulsory included
Main topic:	Football – hooliganism
'Real life' tasks:	No, except for a vocationally related interview
Tasks: 3	Reading comprehension: 6 true/false-statements + correct the false statements. 5 questions to be answered in English about the student's area of study Written production: An essay of at least 150 words - 6 choices (all essays)
Text:	750 words, LIX 42, FOG 19
Targets assessed:	2, 4, 5, 6

⁴⁵ I have made the analyses in this outline as brief as possible to sum up the findings

Test C

Time available:	2 hours
Preparation period:	None
Preparatory material:	None
Vocational subject:	Not included
Main topic:	'The Simpsons'/ TV / the American Dream
'Real life' tasks:	Yes, but only the writing instructions
Tasks: 2	Text comprehension (+ writing): Answer questions about the text in full sentences Written production – 3 choices (two essays and an application):
Text:	347 words, LIX 33, FOG 13
Targets assessed:	2, 4, 5

Test D

Time available:	2 hours
Preparation period:	None
Preparatory material:	None
Vocational subject:	To be covered at the oral part
Main topic:	Am. students
'Real life' tasks:	Yes, but only the writing instructions
Tasks: 2	Reading comprehension: Answer 4 comprehension questions Written production: – 4 choices: speech/application/essay/story
Text:	298 words, LIX 42, FOG 15
Targets assessed:	2, 4, 5

Test E

Time available:	2 hours
Preparation period:	24 hours – may bring 4 pages of notes to the test
Preparatory material:	leaflet, 9 pages
Vocational subject:	locally made at schools for each area of study and included in the test
Main topic:	The Iraqi conflict
'Real life' tasks:	No, except for some of the writing instructions
Tasks: 3	Reading comprehension: 8 multiple choice items Written production – 4 essays: letter/discuss/retell/ story Vocational English - different for each area of study: e.g.: personal letter/application
Text:	537 words, LIX 42, FOG 17
Targets assessed:	2, 4, 5, 6

Test F

Time available:	2 hours
Preparation period:	24 hours
Preparatory material:	Leaflet of unknown size, excerpt appendix to exam. May bring handwritten notes
Vocational subject:	Compulsory included
Main topic:	Drug smuggling
'Real life' tasks:	No
Tasks: 3	Ten multiple choice items

Text: Written production: 3 choices – discuss/give opinion/essay based on photo
653 words, LIX 53, FOG 20
Targets assessed: 2, 4, 5, 6

Test G

Time available: 2,5 hours
Preparation period: None
Preparatory material: None
Vocational subject: Compulsory included
Main topic: Students and business
'Real life' tasks: Yes
Tasks: 3 Reading comprehension: Correct 7 false statements+ give written explanation
Write letter of enquiry within student's area of study
Written production: Choice between 4 essay questions
Text: words, LIX 43, FOG 17
Targets assessed: 2, 4, 5, 6

Test H

Time available: 2,5 hours
Preparation period: None
Preparatory material: None
Vocational subject: Included as one of the choices for written production
Main topic: The British Monarchy
'Real life' tasks: No
Tasks: 2 Reading comprehension: 10 true/false-statements
Written production: 3 choices – essay/ report/ story based on picture
Text: words, LIX 36, FOG 16
Targets assessed: 2, 4, 5, (6)

Test I

Time available: 3 hours
Preparation period: None
Preparatory material: None
Vocational subject: To be covered at the oral part
Main topic: Taking a gap-year/the young British generation
'Real life' tasks: Yes
Tasks: 3 Text comprehension: Explaining words + writing a list
Written production in connection to the text: 3 choices
Written production (essays) – 3 choices
Text: 375 words, LIX 47, FOG 18
Targets assessed: 2, 4, 5

Test J

Time available: 3 hours
Preparation period: None

Preparatory material: None
 Vocational subject: Not included
 Main topic: London's cockney culture
 'Real life' tasks: Yes, but mostly section one and two
 Tasks: 3 Reading comprehension: write a short summary
 Written production: Give opinion about statement
 Written production: essay – 4 choices
 Text: 416 words, LIX 36, FOG 15
 Targets assessed: 2, 4, 5

Test K

Time available: 3 hours
 Preparation period: 48 hours
 Preparatory material: leaflet, 9 pages, may be used on the exam day
 Vocational subject: Locally made at schools for each area of study and included in the test
 Main topic: Power / The Media
 'Real life' tasks: Yes, to some extent, mostly in the writing instructions
 Tasks: 3 Reading comprehension: 8 true/false statements+ short written production
 Written production – 4 choices: private letter/retell/discuss /article/essay
 Teacher include vocational sub-test: (No examples provided)
 Text: 391 words, LIX 37, FOG 15
 Targets assessed: 2, 4, 5, 6

Test L

Time available: 3 hours
 Preparation period: Doesn't say
 Preparatory material: Leaflet, 15 pages. Excerpt as appendix to exam
 Vocational subject: locally made at schools for each area of study and included in the test
 Main topic: Work and education
 'Real life' tasks: No
 Tasks: 3 Reading comprehension: 10 true/false-statements
 Teacher include vocational sub-test: (No examples provided)
 Written production: Essay – 5 choices
 Text: 365 words, LIX 31, FOG 14
 Targets assessed: 2, 4, 5, 6

Test M

Time available: 3 hours
 Preparation period: None
 Preparatory material: None
 Vocational subject: locally made at schools for each area of study and included in the test
 Main topic: New York's homeless people
 'Real life' tasks: No
 Tasks: 3 Reading comprehension: answering 5 questions+ correct false statements
 Teacher include vocational sub-test: (No examples provided)

Written production: essay – 5 choices
Text: 463 words, LIX 40, FOG 16
Targets assessed: 2, 4, 5, 6

Test N

Time available: 3 hours
Preparation period: None
Preparatory material: None
Vocational subject: Locally made at schools for each area of study and included in the test
Main topic: Youngsters and car accidents
'Real life' tasks: No
Tasks: 4 Reading comprehension: 6 true/false-statements + correct the false statements.
3 questions to be answered
Teacher include vocational sub-test: (No examples provided)
Written production: essay – 4 choices
Text: 463 words, LIX 40, FOG 16
Targets assessed: 2, 4, 5, 6

Test O

Time available: 3 hours
Preparation period: None
Preparatory material: None
Vocational subject: Not included
Main topic: Gap year and what to do with it
'Real life' tasks: Yes, except for some of the writing prompts
Tasks: 2 Reading comprehension: Write an introductory e-mail
Written production: essay – 5 choices
Text: 425 words, LIX 43, FOG 20
Targets assessed: 2, 4, 5

Test P

Time available: 3 hours
Preparation period: None
Preparatory material: None
Vocational subject: Compulsory included (very general)
Main topic: Hotels in London/school experience
'Real life' tasks: Yes, except for some of the writing prompts
Tasks: 3 Reading comprehension: Make a booking for a hotel room
Voc: Tell about a project/work placement/excursion
Written production: essay – four choices
Text: 320 words, LIX 47, FOG 18
Targets assessed: 2, 4, 5, (6)

Test Q

Time available: 3 hours
Preparation period: One weekend
Preparatory material: Leaflet, 6 pages
Vocational subject: locally made at schools for each area of study and included in the test
Main topic: Immigration and minority groups
'Real life' tasks: No, except for an accident report and a couple of the writing prompts
Tasks: 3 Reading comprehension: 6 multiple choice items+ answering 3 questions
Example Voc: Write an accident report
Written production: essay – 5 choices
Text: Based directly on the leaflet, no text included
Targets assessed: 2, 4, 5, 6

Test R

Time available: 3 hours
Preparation period: One weekend
Preparatory material: Leaflet, 4 pages
Vocational subject: locally made at schools for each area of study and included in the test
Main topic: Typically American
'Real life' tasks: No
Tasks: 3 Writing 3 different kinds of statements. Example voc: Write a report or an instruction
Written production: essay – 5 choices
Text: Based directly on the leaflet, no text included
Targets assessed: 2, 4, 5, 6

Test S

Time available: 5 hours
Preparation period: 48 hours
Preparatory material: Leaflet, 15 pages
Vocational subject: Compulsory included
Main topic: Terrorism
'Real life' tasks: No
Tasks: 4 Completing sentences through 5 multiple choice items+ making questions to 3 answers. Example voc: write an application. Written production: essay – 5 choices
Text: Based directly on the leaflet, no text included
Targets assessed: 2, 4, 5, 6

Test T

Time available: 2 hours
Preparation period: 24 hours
Preparatory material: Leaflet, 6 pages
Vocational subject: Not included
Main topic: Two American Heroes
'Real life' tasks: No
Tasks: 2 Reading comprehension: Answering 4 comprehension questions
Written production: essay – 4 choices

Text: Based directly on the leaflet, no text included
 Targets assessed: 2, 4, 5

Test U

Time available: 2 hours
 Preparation period: 24 hours
 Preparatory material: Leaflet, 4 pages + video (45 min) about New York
 Vocational subject: Not included
 Main topic: My culture, your culture
 'Real life' tasks: Yes
 Tasks: 2 Reading comprehension: Making five "Trivial Pursuit"-cards +writing 3 explanations
 Written production: essay – 5 choices
 Text: Based directly on the leaflet, no text included
 Targets assessed: 2, 4, 5

Table 4: Length of the written part of the exam

1 hour	2 hours	2,5 hours	3 hours	5 hours
0 tests	8 tests	2 tests	10 tests	1 test

Table 5: Application of a preparation period

None	24 hours	48 hours	1 weekend	Doesn't say
12 tests	4 tests	2 tests	2 tests	1 test

Table 6: Amount of preparatory material

4 pages	6 pages	9 pages	15 pages	Doesn't say
2 tests	2 tests	2 tests	2 tests	1 test

Table 7: Text length in number of words

<= 350	350 – 450	450 – 550	550 – 650	650 =>
2 texts	6 texts	4 texts	2 texts	2 texts

Table 8: LIX score indicating readability of the texts

<= 35	35 - 40	40 - 45	45 =>
3 texts	5 texts	5 texts	3 texts

Table 9: Testing of reading comprehension

Without involving writing	Written production required
11 tests	10 tests

Table 10: 'Real life' tasks

No	Partly	Yes, in the writing instructions	Yes
9 texts	5 texts	4 texts	3 texts

When summing up the findings this way, we will see that the average test has a time frame of two or three hours. 12 tests include preparation, and the preparatory material differs considerably in size. Text length varies from 298 words to 750, and readability according to the LIX index from 31 to 53. Nine tests do not apply 'real life' tasks at all, nine do it to some extent, while three do it consistently. Half of the tests assess reading comprehension by means of written production. The sample reveals, as demonstrated, great variety in text length and difficulty, in demands for written production, in time available to the students on the written part, in design of sub-tests and in targets tested. This is clearly a challenge both to validity and reliability, but they may still, to a reasonable degree, be in accordance with the guidelines.

The fact that the exam tests vary is not a problem per se. It is the clear intention of the Exam Secretariat that teachers and students will not be able to profit by putting their faith in last year's test when preparing for their exam (Talberg 2004). The types of tasks are supposed to vary a lot more from one year to the next than they did before R'94, and those relying on the expectations brought about by previous exams are bound to be disappointed. According to the central guidelines for local exams (SUE/vg – 95-005), more than one target need to be included, but it doesn't say *all* targets need to be tested every year. However, this doesn't mean that parallel tests should differ to a large extent in level of difficulty or in amount of work required.

Comparing Talberg's 1997 results to my own impression based on information gathered in connection with the present thesis, it appears to be less cooperation between schools on the regional test making now than it used to be a few years back, and also less discussion among teachers as to what these tests really ought to be like. This may result from the fact that the idea of a combined exam is not a new phenomenon anymore. Since 1997, the development seems to have gone in the opposite direction from Talberg's recommendations (see sections 3.3 and 8).

6.2 Targets tested

As established above, there are six main targets that constitute the objectives for assessment of the VF2210 exam test. Testing of the basic skills ‘reading’ and ‘writing’ has been discussed in chapters 4 and 5 above. In the following I will give a brief review of the findings, mainly to summarize how the different targets relevant to the written part of the exam are being treated in the sample tests.

Target 4 – ‘the use of written English’ is in all cases seen as the main part of the test. The assessment guidelines suggest that in the weighting between the different tasks in the written part of the exam set, the ‘essay task’ should count for 50% of the grade. All sample tests include this kind of timed impromptu writing (Weigle 2002:69), and written production is more often than not included in the testing of the other targets. For further details on the testing of writing, see section 5.2 above.

Target 2 – ‘Comprehension of written English’ is tested in all the sample tests, normally based on a text of a general topic and in some cases based directly on a preparatory leaflet. The comprehension tasks related to the texts differ significantly with respect to level of difficulty and in the need for writing competence and written production. The testing of reading comprehension will normally count for only 20% of the grade, but some of the tests put some more emphasis on this part. For further details on the testing of reading comprehension, see section 4.4 above.

In all the studied cases there is a possibility for the students to show some kind of knowledge in relation to **Target 5** – ‘Knowledge about the English speaking world’ - but none of the sample tests require exact knowledge or achievement in relation to the syllabus. The curriculum is to a large extent open to making individual choices on classroom level, and a regional ‘county exam committee’ with no access to the specific syllabus of each class will have to make the prompts and tasks fairly general to meet the different needs. The variety of possible foci within this topic makes it difficult to test exact achievement, but the students should at least be given the opportunity to show knowledge in relation to the course they have been in. So-called ‘open tasks’ will to some extent take care of this question, and employment of such tasks are in accordance with the intentions of the National Exam Secretariat. The notion of ‘open tasks’ will be discussed further in chapter 8.

None of the tests in the sample assess **Target 1** – ‘Comprehension of spoken English’ in the written part of the exam, even if a listening comprehension task easily could have been included. One test makes use of a video as preparatory material. Four of the tests are consistently applying ‘real life’ tasks, while another five combine traditional comprehension tasks with ‘real life’ writing tasks, or the other way round.

6.3 The testing of Target 6 – the vocationally oriented English

According to appendix 1 of the R’94 syllabus, more or less half of the course is supposed to focus upon vocational vocabulary and texts in relation to the student’s respective area of study: ‘The amount of material studied in each module under targets 5 and 6 must be balanced, so that neither target exceeds 60% or fall under 40% of the total’. This section will examine to what extent this fact is reflected in the final assessment.

In order to find out how vocationally oriented English is treated in the teaching material, I have examined three of the most widely used English textbooks⁴⁶ for the Building and Construction courses. I found that textbook writers do seem to take the R’94 curricula and syllabuses seriously with respect to this target. ‘Know How’ contains 36 lessons of which 18 are vocationally oriented. ‘@t Work’ is divided into 13 lessons of which 6 are on vocational topics, and in ‘Hands On’ half of the ordinary chapters deals with working life in general, in addition to 10 comprehensive lessons on specific vocational topics, all including both vocabulary exercises and comprehension tasks. Decisions as to which lessons to choose are made on classroom level.

The sample material shows four different ways of dealing with target 6: Seven of the tests leave it out completely, normally with a remark that it should be dealt with in the oral examination. One test includes a vocational topic among the writing prompts and leaves it up to the student to choose it or not. Five tests include a very general task, like a report or a “tell-about-your-future-vocation” task. Finally, some tests leave space for the local teacher to include his/her own vocational task, providing a possibility to let the assessment reflect the teaching and in that way include some achievement testing in the exam set, and as many as

⁴⁶‘Know How’ from Gyldendal – ISBN 82-05-24421-9

‘@t Work’ from Aschehoug – ISBN 82-03-32855-9

‘Hands On’ from NKI Forlaget – ISBN 82-562-3373-7

eight of the sample tests have chosen this strategy. Since my sources are the regional exam boards, only very few examples of locally added vocational tasks have been included in this study, but I list the ones I've got just to show what they might be like:

- You are about to apply for a job at a nursery school. In your application you are asked to write down what qualities you think are important when working as a nursery nurse. You also have to give reasons why you believe these qualities are needed.
- Write a report about an accident that happened at school or at work. (list of possible injuries and a series of pictures included)
- Write a letter to a friend in England where you give some reasons why you want to become a carpenter. You should also describe the practical part of your education so far. (Try to use relevant terminology)
- Either: Write a report from a project, a placement or an excursion you have been part of. Make sure you focus on a task typical for your field.
Or: Write a step-by-step instruction of how you should perform a special task within your special area.
- Read the advertisement below and write an application for a position at *U.S. Navy Telecommunications* (Advertisement with a lot of specifications included)

As we can see from these examples, the average vocational sub-test tends to be just another writing task. Only one test provides vocationally oriented input, and none of them test comprehension of vocationally oriented texts. Compared to the objectives in target 6, where we find that the students should be able to *understand*⁴⁷ and use English on specialized subject matters and on subject-related and inter-disciplinary topics, this seems to be a serious shortcoming. At best, the examples provide possibilities to show knowledge of active vocational vocabulary, *if* the students remember to include it.

In my opinion, an unwanted consequence of the current organization with regional test production is that most of the exam sets test proficiency more than achievement. Target 6 appears to be just right for achievement testing. The syllabus is specific to each course, for most areas of study the related vocabulary is new to the students, and the teacher will to a large extent know what to expect from the student on basis of the teaching. The vocationally oriented English is what most students experience as 'new' compared to the lower secondary syllabus, and consequently something they will expect to be assessed on.

⁴⁷ My underlining

7 The oral part of the test

7.1 Theoretical reflections on assessing oral competence

For assessment and testing of oral proficiency, the criteria seem to be even less defined than for the other skills. Important work in this field was done by the ACTFL⁴⁸ in the 1980s when developing exact procedures for conducting an oral proficiency interview⁴⁹, including explicit assessment criteria and a detailed rating scale. The overall aim of the proficiency interview was to elicit speech samples that could be rated confidently and accurately, and the assessment criteria and rating scales turned out to be extremely fine-grained and detailed (Simensen 1998:271). For the school subject English as a foreign language in Norway, less pretentious assessment criteria are needed. The main aim for the students in question is to obtain good communicative ability, and R'94 calls for testing this competence holistically by a global synthetic judgement. The different aspects involved are weighted intuitively by the assessors. In the general guidelines for oral assessment (SUE/vg-99-022) we find, among other things, rating criteria for comprehension and guidelines for the assessors' behaviour. As to the VF2210 exam, the objectives are 'comprehension' and 'use' of oral English. Below we find the more specified aims of the targets in question:

Target 1 Comprehension of spoken English⁵⁰

The pupil should

- 1a be able to understand everyday speech on general topics
- 1b be able to grasp the main contents of a longer utterance, also when it contains terminology relating to his or her area of study.
- 1c be able to understand an oral instruction
- 1d be able to obtain information relevant to various needs from oral sources
- 1e be able to understand the attitude and intention of a speaker

According to Gary Buck (Buck 1997:66), testing listening is an indirect process and thus listening scores will always be influenced by other skills required for task completion (e.g. answering questions, understanding charts, writing summaries etc). The basic technique for testing listening comprehension is to present an oral utterance in the target language, and check whether the test-taker has grasped some crucial part of it. As to the Vf2210, 'crucial parts' would be grasping the main contents, understanding everyday speech, or obtaining

⁴⁸ ACTFL = American Council on the Teaching of Foreign Languages

⁴⁹ *The ACTFEL Oral Proficiency Interview*. 1989

⁵⁰ The targets are quoted from the English version of the R'94 curriculum

information relevant to various needs from the oral source. Selected responses like ‘multiple choice’ or ‘true/false’ are, according to Buck, the preferred test technique in order to avoid including writing skills. He underlines that it is rather rare nowadays to ask for constructed responses. In communicative listening testing, there seems to be agreement that both texts and tasks should be as realistic as possible.

In addition to the receptive skill of listening, there are two kinds of production competences involved among the oral skills; spoken interaction and spoken production. According to the syllabus extract below, both are represented in the objectives of the course and are therefore inclined to be included in the assessment:

Target 3 Use of spoken English

The pupil should

- 3a master normal forms of communication like conversation and communication on social and job related topics.
- 3b be able to express and explain his or her own attitudes and opinions naturally and easily in conversations and discussions
- 3c be able to give a coherent oral presentation of a topic, i.e. a set of circumstances or an experience
- 3d be able to communicate some prominent features of Norwegian culture

As we will see from the interviews in the next section, both interactional and transactional tasks normally seem to be included in the oral part of the exam: Oral interaction in the form of examinations or conversations between the examinee and the teacher, and oral production in the form of prepared presentations. To help the students perform to the best of their ability, Aud Marit Simensen (Simensen 1998:271) suggests different strategies for eliciting speech:

- Be a good listener, signal interest in what the student says.
- Remember that silence may be constructive – give the student time for reflection
- “Soften” direct questions by the use of modal auxiliaries (Would you like to explain...)
- Introduce a series of topics rather than a series of questions
- Two or more students may be assessed at the same time, but the interlocutor effect⁵¹ may bias the result

According to the central guidelines⁵², there is no opportunity to assess two or more students at the same time in the VF2210 exam, as it is clearly stated that the test has to be taken individually. The criteria by which a good test of speaking can be judged are, according to

⁵¹ Interlocutor effect = the disadvantage a student may have if there are shortcomings on the part of the interlocutor(s) in the group

⁵² SUE/vg – 95-005 (*Sentrale retningslinjer for lokalt gitt og sentralt gitt eksamen*) section 3.2

Glenn Fulcher (Fulcher 1997:75), those applicable to all language tests: reliability, validity and practicality. In the testing of speaking, however, the problems in examining these qualities are heightened by the nature of speech itself. Eliciting a language sample large enough for adequate assessment is time consuming, and scoring will for the foreseeable future depend on the use of expert human judges. In the case of the VF2210 exam, these judges are the language teachers. Fulcher recommends the use of multiple tasks, as it has been increasingly observed that task type has a systematic effect on speaking test scores. According to the curriculum excerpt above, the students should be able to master normal forms of communication, give a coherent oral presentation of a topic and to express and explain their own attitudes and opinions naturally and easily in conversations and discussions. In conclusion, this should call for multiple and varied test tasks and types of oral performance during the exam test.

7.2 Examples based on interviews

In the following, I have cited interviews⁵³ with five randomly chosen teachers from different schools and counties. They were asked to comment on the organization of the oral examination, number of students, assessment of the written products, time consumption, topics, targets tested, vocational English and personal experience with the combined exam.

Teacher A

‘At our school, we normally have a 20 minutes’ oral examination, starting two hours after hand-in of the written part. During these two hours, both assessors read through the six written products. Each of the two assessors has to estimate a temporary grade to be modified by the oral examination. The examination takes place as a combination of assessment of the student’s prepared performance and the conversation between the examiner and the examinee, and it is divided into three parts:

- Student’s own choice of presentation (a syllabus text, a project, a film or an excursion).
- Teacher’s choice of literary texts from the syllabus: The student should read a part and translate it + retell the rest of the story.
- Conversation about the student’s vocational area of study (focusing on tools, working processes and safety).

⁵³ My own translations

When it comes to grading, the student's oral and communicative skills are just as important as his/her written ones. Actually, when in doubt, the oral grade will overrule the written one. This means that the temporary grade from the written part may, in some cases, change considerably after the oral examination. By the end of the day, the two raters have to agree on one grade to express the student's holistic competence in English, comprising all four skills and both targets of content. Sometimes that feels hard to accomplish.'

Teacher B

"In my school, 5 to 7 students constitute one examination group. The students have three hours at their disposal for the written part, 20 minutes for the oral part, and 5 minutes in which we explain our evaluation of the student's total performance and give a grade. We allow two hours between the written and the oral part of the examination, in which we assess the written products. These may be further discussed during the oral examination if the assessors find it serviceable, or if the student himself wants to do so. We test all the targets that the curriculum requires. Since the different schools in the county use different textbooks, digital programmes and other kinds of teaching material, a committee is put together to develop the written part of the common regional exam test. We also develop a preparatory leaflet given to the students 48 hours in advance. Themes have been 'Peace', 'Young People', 'Work and Education', 'Challenges' and 'Activities'. Each individual teacher is responsible for making the vocational part of the test for his/her own group. In my group, the oral part has been carried out as follows:

- Casual 'small talk' to make the student feel welcome and at ease
- Questions to the written part if needed (or wanted)
- The student picks a card containing a syllabus theme, and then we have a conversation around this theme. I endeavour to let the student speak as much as possible, but from time to time I find it useful to ask some questions.
- On the same card there is also an oral vocational task.

As can be seen from my answers, the vocational part is taken good care of. I myself feel comfortable with the current organization of the examination, because I feel that students struggling with the written part get a good chance to show their competence at the oral part. To my experience, the students that are selected to sit for the exam work very well during the 48 hours they have at their disposal, and in my opinion a larger part of the lot ought to have a chance to have this examination'.

Teacher C

‘This year, a group of six students from my class of Electricians (2EL) were selected for the combined exam in English. After the written part, we have two hours, 120 minutes, to assess the six products and that really takes some effort to do. In the written part they were tested in reading comprehension by means of ‘multiple choice’, weighing 20% of the test. Written production about youth culture and drug-related problems constituted 50% of the test and the last 30% was a vocational task, or two tasks to choose from, to be precise: a) ‘Wiring a house’ or b) ‘Power generation, transmission and distribution in the UK and Norway’. The oral part is normally conducted as follows:

- A conversation about the vocational part of the student’s written product as starting point.
- Further conversation and examination on one of the following topics, to be drawn by the examinee:
 - Safety and tools
 - Security systems and security devices
 - Electric and hybrid cars. Advantages and disadvantages.

Two by two of the students had the same oral task. Each examination lasted for 25 minutes, adding 5 minutes for consultation between the two assessors. As already established, we aim at assessing as much vocationally oriented English as possible in the oral part, as so much of the written part focuses on target 5. In my opinion, this form of combined exam works well, but it is certainly demanding for the assessors. You feel an enormous time pressure when assessing in between the written and the oral part. Six candidates is an absolute maximum of what you can handle in such a short time period’.

Teacher D

‘None of my students were selected for exam assessment this year, but here is how we usually do it: The first part is a listening comprehension task. The topic of the recorded text is always vocationally oriented. The examinee is allowed about 20 minutes before the examination itself for listening to the tape. He/she may take notes while listening and bring these to the oral examination. The examination lasts for 20-30 minutes and goes like this:

- The examination starts with a conversation about the recorded text, 5 – 10 minutes
- The examinee may, if he/she wishes to do so, expand the responses from the written part. This is of particular importance for students with specific reading-/writing difficulties. For this, only a few minutes are allowed

- The examination continues with a conversation about the prepared topic, 5 – 10 minutes. The students have on beforehand (before the exam) been asked to pick and prepare a subject-related (vocational) topic. This might be:
 - A cross-curricular or vocational project work
 - The practical training / placement period
 - A technical field within their area of study (e.g. 'the brakes' if the area is Car Mechanics)
- The examinee gives a presentation of a short-story, a novel, a play or another literary work , 5 – 10 min

When deciding on the final grade, the oral and the written part will count equally much. When in doubt, the oral part is the most important. Two assessors, a local and an external examiner, will assess both the written products and the oral performance, and they have to agree on one final grade. The written part is being assessed in between the written and the oral part of the exam and the resulting preliminary grade will be adjusted on basis of the oral examination'.

Teacher E

'This was my first, and so far my only, experience with an oral English exam test. My group consisted of five students from the Health & Care course, and the examinees were allowed about 20 minutes each for the oral examination after a two hours' break subsequent to the written part of the exam. The written products were assessed by me and an external teacher during this break. The oral examination tested both general and vocational English. The weighting here is not determined once and for all; it will have to reflect what we have been working on throughout the year. This time, we chose to put a bit more emphasis on the vocational than on the general English. The oral part of the exam was like this:

- First, the students were asked to tell about their work placement period. They were supposed to include what kind of tasks they had been involved in, whether or not they liked the work, if they could see the use of their school subjects in the practical working situation, whether or not they expected a future career within the area and so on.
- Then the student had to randomly select a card with two texts from the syllabus; one literary text within the general English area and one subject-related text. The student was asked to retell the literary text, and then we had a conversation on basis of what she told, followed by an examination and/or conversation about the vocational text.

In the assessment we focussed upon vocational knowledge and ability to communicate, including vocabulary size and spoken fluency. As mentioned above, this was my first time with this kind of examination, and I have to admit I experience it as rather 'busy', for us examiners as well as for the students. At our school many different groups were having oral exams this day, but the others were oral exams proper (not in combination with

writing). I noticed comments among the other teachers on students having improved their grades (that is to say: got a better grade on the oral exam than on the final assessed attainment). This was also the case with my students, but the results from the written part lowered their grades. In that way the combined exam ‘deprived’ them of the possibility to improve their results. On the other hand, my students ended up with the same grades as they already had in beforehand, implying this was a fairer and more ‘correct’ result. All in all, I guess I am satisfied with the current exam’.

7.3 Summing up the oral part of the exam

A sample of five informants is a fairly thin basis for drawing conclusions, but some tendencies still seem to appear. On basis of the interviews above, supplied by some local guidelines for assessors, it seems that 5 to 6 examinees typically form an exam group. This is probably due to feasibility, since the assessors operate under time pressure. One of the counties holds out in their guidelines that the assessment of the written products ought to be done *after* the oral part. This is not common practice, but may be to the benefit of the student if the written product is weak. When evaluating the oral performance, it will be easy to seek a confirmation of the impression made by the written part if this has been assessed before the oral examination.

As far as I’ve been able to find out from the interviews and from information included in the written tests, all counties have chosen a 20-30 minutes sequence of oral examination later on the same day as the written part. This is in line with the official guidelines (see section 2.6). It is important to notice that the oral part of the final assessment should be something else than a traditional examination in syllabus content (SUE/vg -99-022). According to the targets, listening comprehension and everyday conversation should be included, as well as a possibility for the students to perform a prepared presentation of something they have been working on during the course.

Only one out of five tests seems to have included a listening task, but all of the tests in question include some kind of conversation, and two include a prepared presentation. Another two let the student present their placement period. Still, there seems to be a need for more focus on the variety of oral genres and on the students’ oral performance in terms of ‘real life’

communicative competence. Talberg (Talberg 1997, discussed in section 3.3) found that the oral part normally consists of listening to a recorded text and answering questions about the text, succeeded by a conversation about a prepared or a randomly selected topic. According to the examples above, this seems to have changed.

The weight put on procedural knowledge as compared to declarative knowledge varies a great deal, but in most cases both are seen as important. Based on information to the students found in the written exams, seven of the cases studied test vocationally oriented English as a compulsory part of the oral examination. In all cases, it seems that testing of the literary syllabus from target 5 is a compulsory part of the oral examination, but this is not confirmed by the interviews. Here we find that two out of five exclude the literary syllabus altogether and focus on other sub-targets instead. None of the teachers mention in particular how and if they assess target 1e ‘understand the attitude and intention of the speaker’. The social competence, the so-called ‘will and skill’ to interact, tend to affect the students’ performance anyhow, and will as such be implicit in the assessment.

The oral part of the exam seems to be the only guarantee for achievement testing of all students. It is the only test element that in all cases⁵⁴ will be made and conducted by the students’ local teacher. I hope the teachers are aware of their responsibility in this connection.

My general impression, on basis of the study as well as on my own experience, is that an oral examination is an important part of the final assessment of the students’ holistic competence in English as a foreign language, and that the vocational students in particular profit from being tested orally, and thus deserve to be given this possibility in the final assessment. As will be demonstrated in section 9.2.2 below, the combined exam is about to disappear with the new reform. Under the new system, the students may be randomly selected through multistage sampling for **either** written **or** oral exam testing, and in my opinion, this will make the exam assessment go from bad to worse with regard to fairness. I will briefly touch upon this issue again in the final section.

⁵⁴ Except for the private candidates

8 Discussion

What kind of conclusions may be drawn on basis of these results? Do they provide answers to the questions asked in the introduction? In what ways has R'94 influenced the EFL assessment? Are the changes⁵⁵ for the better or the worse with respect to the vocational students? As demonstrated above, there are differences in the test examples, but are these differences significant for the result of the assessment? Do the tests, despite all differences, assess the students in accordance with the R'94 directions? Does the testing to a reasonable degree reflect the students' achievement, or is the idea of a local exam lost along the way? Does the oral part add anything of significance to the exam assessment? Are all the targets taken into consideration, and does the one single grade the students end up with really say something meaningful about their holistic competence in English? And finally, is test-making on the regional level a threat to fair exam assessment of English for the students in the vocational branches?

Many aspects of R'94 teaching and testing have been discussed above, including the testing of reading comprehension in section 4.4, the testing of writing in section 5.2, organized preparation in section 5.3, and the oral part of the exam in section 7.3. In addition, an account of results from related research has been given in section 3, and some current views on the topic will be presented in section 9. In the following, I will return to various problem areas in order to discuss them in more detail, and I will bring up one or two issues that still haven't been touched upon.

8.1 Analyses and discussion of selected findings⁵⁶

In section 6, an attempt was made to sum up my findings as briefly as possible. On basis of these summaries and the specific results from section 4 and 5, I will now discuss some features and issues of special interest to the present study, and analyse and compare some of the exam sets in more detail.

⁵⁵ See section 8.2 below

⁵⁶ Only the written part of the test is taken into consideration here, as I have no information about the oral parts of these specific exam sets.

The R'94 curriculum and the related guidelines are to a large extent open to interpretation, and with so many test makers designing parallel test on basis on their own interpretations of these documents, differences are likely to occur. Table 1 in section 4.4 illustrates that in the ranking of texts according to readability⁵⁷, three texts stand out as far too difficult, while another three are listed as quite easy. Concentrating on the two easiest and the two most difficult, I will now make an attempt to describe and compare the four exam sets they are part of in order to find out if the differences in level of difficulty will remain when adding the test tasks and the other sub-tests. The easy texts are **text L** and **text C**, and the most difficult ones are **text P** and **text F**, measured according to a readability index⁵⁸.

Exam set L turns out to be a quite comprehensive set, even if it has the easiest text. The test is based on a 15 pages preparatory leaflet, and three hours are allowed for the written part of the exam. The students may not bring the leaflet on the exam day, but 2 pages of handwritten notes are allowed. The examination text has, as far as I can see, not been included in the leaflet, although other texts touch upon the same topic. The text is taken from an English textbook for the vocational branches, and this might explain the low LIX score. The topic, British Politics, is probably not perceived as easy by the students. Comprehension is assessed by the means of a 10 items' right/wrong task, of which some items involve reading 'between the lines'. No written production is required. The writing test is an 'essay' task with five prompts to choose from. No writing instructions to create 'real life' situations are provided, and all the tasks include some kind of reasoning, asking for the student's opinion on or explanation of given topics. No 'write a letter', no 'tell about', no short answer alternative. The last task invites the students to evaluate their teachers and the Norwegian school system! Unfortunately, I have no access to the vocational task, which was meant to be included by the local teacher. All in all, this is not an easy exam set. The comprehension task is practicable, but the set in total requires good reading, writing and reasoning skills.

Exam set C is of a different kind. There is no preparation preceding the exam test, and only two hours allowed for the written part. An attempt has clearly been made to make this a 'real life'-like test, and the topics are all up to date, aiming to catch the students' attention: The Simpsons, Trash TV (and the American Dream) and Olympic Games in Tromsø. The text is an authentic youth magazine feature story about The Simpsons, and comprehension is tested

⁵⁷ See table 1, section 4.4

⁵⁸ The LIX index, see section 4.2.1

by means of comprehension questions. Unfortunately, there is no way of showing comprehension without written production, but most answers may be copied directly from the text, and only one question invites the student to go 'beyond the text': 'Why is Homer compared to the American Dream?' The writing instructions all aim to create 'real life' situations, and there are three different prompts to choose from. Two of them are variations of the 'tell about and explain'-type of tasks, while the last one is a job application. The testing of vocational English is left for the oral part of the exam. This seems to be a fairly easy exam set, even if it relies to a large extent on the student's writing competence.

By comparing the two sets, we find that they differ significantly both with respect to amount of work and in level of difficulty according to my judgement on basis of the analysis. **Exam Set C** is by far the easiest one, and requires less time, skills and effort from the students than **Exam Set L**, even if the latter includes a selected answer task. This applies to organisation as well as tasks and content. I will now analyse the exam sets that include the most difficult texts.

Exam Set P is a three hours' test without organized preparation. It appears to be a 'real life'-oriented test, basing the reading comprehension on authentic accommodation advertisements. As often is the case with 'real life' tasks, comprehension assessment is heavily dependent on writing skills. The students are asked to make a booking for a hotel room based on one of the advertisements, and this is the one single 'comprehension' task. In the second task, which is the vocational alibi, the students are expected to write about a project work or a work placement. Task three is the traditional 'essay' task with four prompts to choose from; one asking for 'personal view', one 'tell about and comment on' and two 'personal reflections' (on the USA or the UK). In conclusion, this is a clean-cut writing test: Three writing tasks, of which one is based on some written input!

The last set, **Exam Set F**, is based on the most difficult text according to the readability index. It is a two hours' test based on a preparatory leaflet with 24 hours for preparation. The students may bring handwritten notes from the preparation session, but not the leaflets. One text is included as appendix to the exam set, and is indeed difficult to read. It appears to be an authentic net paper article, syntactically complex with high lexical density. Reading comprehension is tested by a set of 10 multiple choice items. The distractors are all close to the correct alternative and the items are not chronologically sequenced. The task seems quite

tricky, requiring very good reading abilities and reasoning skills, but at least it does not require written production. Section two is the 'essay' tasks, with three prompts to choose from, all linked to the topic introduced in the text. Two of them are of the 'discuss/give opinion'-type, and one prompt invites the students to 'write a text based on the photograph'. There is no 'short answer' alternative. Section three, the vocational part, will be included locally and can thus not be commented on. A period of two hours is quite a short time for completing all three tasks, especially since the reading comprehension is that demanding. All in all, this test seems to be on the difficult side.

Four exam sets make up a very small sample to draw conclusions from, but the differences that appear do to some extent speak for themselves:

- Poor writers will have better chances of success in some counties than in others
- Poor readers will have better chances of doing well in some counties than in others
- Not all counties test achievement to the same extent
- Readability of text is not a sole indicator of general level of difficulty of the exam sets
- No correspondence is demonstrable between a longer time frame and the inclusion of a third (often vocationally oriented) section
- There is no traceable correspondence between difficulty of text and difficulty of comprehension tasks
- Exam assessment with the VF2210 exam test under the current system is not fair to the students!

My general impression is that validity is taken care of to a reasonable degree, except for the lack of achievement testing in some of the tests, but that reliability and fairness suffer severely. The two seem to have been sacrificed on the altar of practicality and cost effectiveness; considerations that have resulted in the regional test-making in the form that we experience it now. The intentions of cooperation and pedagogical development seem to have been lost along the way, together with the purpose of a local exam. My hope is that the oral part of the exam test will make up for some of the differences between tests demonstrated here, and guarantee the students a reasonably fair assessment after all.

Do I still worry for the vocational students in particular? I do, because it is a fact that many of these students have far better receptive than productive skills, and that they communicate

much better orally than in writing⁵⁹, and I am still not sure if they get the chance to perform to the best of their ability and show all sides of their communicative competence in the final assessment. Good writing skills are, as shown earlier in this study, traditionally seen as the most important of the four basic skills when learning a foreign language. In the Norwegian school subject English, a methodology leaning heavily on the old Grammar-Translation Method have influenced most of the 20th century. Even now, in the middle of the Communicative Language Teaching trend, working with written exercises in connection to texts are by far the most commonly applied teaching strategy. In fact, even more so lately, with the teacher functioning more like a coach than an intermediary of knowledge, and the students, individually or in groups, spending the days doing their written assignments. This doesn't change the fact that many students have far better oral than written skills, and that they prefer oral activity if asked⁶⁰.

VF2210 is, as demonstrated, no exception when it comes to favouring writing skills. Not only do all sample tests rely on essay writing as the best way of showing writing competence, but writing is involved in most other test tasks as well. A student's writing skill may actually still be decisive for the exam result, despite the emphasised role of the comprehension skills in the curriculum and of the oral competences in the assessment guidelines. Section 9 will show that these circumstances may in fact become worse with the new reform of 2006.

The role of ESP, the vocational English, has been discussed and touched upon several times in this thesis, most extensively in section 6.2. The target is clear; the students should understand and use English in relation to subject-related and specialized subject matters. This must, in my opinion, mean more than being able to write a formal letter or an accident report! Since the making of vocational sub-tests most often are left to the local teacher, the sample does not provide many examples, but from those at hand, the vocational tasks seem to be mainly writing tests. English in relation to the specific area of study often introduces the students to a new vocabulary, and my guess is that most of them have a larger passive than active vocational vocabulary at the end of the course. Comprehension tasks based on vocationally related texts would, as far as I can see, meet several assessment targets and provide a chance to test achievement in relation to the teaching.

⁵⁹ This is confirmed by a large scale surveys discussed in the 2005 strategy document '*Språk åpner dører*' from the Norwegian Ministry of Education and Research.

⁶⁰ See section 3.1

There is a significant difference between the academic and the vocational branches when it comes to organization of the EFL school subject: As stated in section 6.3, the amount of material studied in each module under targets 5 and 6 must be balanced, so that neither target exceeds 60% or falls under 40% of the total in the vocational courses. That means about half of the time in the vocational classes will be spent on English in relation to the students' respective area of study. Learning a new vocabulary is time-consuming, and will leave the vocational student with little time to work on the more general English topics compared to students in the more academical branches. This, in addition to the fact that vocational courses get only two periods a week over two years, with no tradition for homework, will make the gap between the two groups of students grow quite big⁶¹ if they are to sit for the same exam. The new curriculum "Knowledge Promotion" will, as section 9 will demonstrate, bring about changes within the area, some for the better and some for the worse.

8.2 Some consequences of R'94 and regional exam design

As expressed above, there has been a change in assessment of English as a foreign language in upper secondary vocational education (Ulriksen 2001:86), and R'94 brought about extensive changes for the vocational students in particular. Before R'94, English in the vocational courses was regarded as one of the vocational subjects. There was no external exam assessment, just the assessed attainment at the end of the course. The syllabus was locally designed to fit local variations, and English was taught only in the advanced courses. With the introduction of R'94 the number of periods was doubled, English was given an academical approach and external exam assessment was introduced. Local exams were chosen in order to let the local variations of the vocational courses be reflected in the exam tests. The county authorities were made responsible for monitoring the test-making, resulting as demonstrated in the present thesis in the common regional production of the local exam tests. A consequence of local exams per se is that there will be variations in the test constructs, but these differences should appear on the basis of pedagogical considerations according to teaching methods, syllabuses and vocationally related needs of the different courses. Significant differences in the tests mainly due to the regional test production are harder to justify.

⁶¹ If the vocational student chooses to add an extra period of English in their third year, they will sit for the same exam as the academic courses. This problem will not be treated any further in this thesis.

The sample tests demonstrate, despite all differences, that the R'94 guidelines are to some extent taken into consideration when tests are made. The possibility of showing communicative competence based on all four skills appears to be extremely valuable for the student group in question, and the main obstacle for better results seems to be the condescending attitude among some teachers and test-makers towards the receptive skills. Writing is still seen as the more important competence as discussed above. The oral part of the exam forms the only real counterweight to written production, and is thus very important to balance the results. I will return to this below.

R'94 advocates in favour of performance-based assessment, focussing on procedural rather than declarative knowledge. The sample tests show that we do not yet in practice fully meet this aim. Far too many writing tasks invite the students to reproduce topical content, and the oral performance tends to be an examination of topical knowledge rather than a test of the students' ability to perform communicative tasks effectively. The exam sets vary, not surprisingly, a lot in this respect. Both Talberg (2004) and Gulbrandsen (1997) argue in favour of 'real-life' tasks⁶², and this concept has to some degree influenced the tests, but not to the extent that the Exam Secretariat intended and expected. Nine out of 21 tests do not include any 'real life' tasks at all, only three tests do it consistently. An argument against using 'real life' tasks for testing comprehension is that this will normally involve written production. As for the oral part, 'conversation' seems to be the most common 'real life' alibi.

In line with R'94, the test should be designed to give the student a fair chance to show his/her competence. The Exam Secretariat emphasizes the importance of making the tasks 'open'⁶³ in such a way that all examinees may solve them at their own level, and the guidelines point out the need for assessors to keep an 'open' mind when assessing 'open' tasks! But 'open' tasks need to be filled. Choosing perspective themselves and/or including topical knowledge of their own specific syllabus may be a challenge for many students, and weak students often find this extremely difficult. The idea promoted by Talberg (Talberg 1997) of always including an alternative to the 'essay' task⁶⁴, i.e. several short answer tasks, seems to me as a brilliant idea.

⁶² See section 5.2

⁶³ See section 5.2

⁶⁴ See section 5.2

But tests are only one factor in assessment. Even if the tests were good enough, the assessors may cause unwanted inequalities. If this is the case, it may be due to a lack of sufficient rater training in the wake of R'94. There were in fact some teacher training courses, but neither did they provide enough practical training nor enough workshops or discussions in order to gain a common understanding of new concepts. There were lots of visionary words, but most teachers stayed confused. Confusion creates insecurity, insecurity reluctance, and reluctance a need for retreat; back to status quo! In my opinion, what mainly has stopped R'94 from becoming the success it was expected to be was a failure in getting the teacher brigade eager for it and in line with the new thoughts and intentions.

At present, I finally feel I have to some extent grasped the idea of R'94 assessment, but that insight comes from the intensive reading of more than 100 pages of different guidelines that in some way or other relate to the assessment of VF2210. The assessor's guidelines that are being distributed with the exam sets are normally of two – three pages, very superficial and not fit at all to stimulate new ideas about assessment, and the scoring rubric included doesn't quite make sense without comprehensive background knowledge. How can you expect teachers to change practice if they do not understand the rationale behind the new ideas? In my opinion, there is a need for some kind of user-friendly support to help teachers find their way in the jungle of documents from the national education authorities. What about a web-site (operated by the Directory of Education?) where you may select i.e. the VF2210 and find all the official documents concerning this specific test? I would also like to discover an electronic database with all the former tests from all the different countries together with the example tests from the Exam Secretariat on this site, and an e-mail address where you could ask questions about different test tools, guidelines and exam assessment.

From my point of view, the official guidelines are still far too 'invisible' for the ordinary teacher. I had a hard time myself finding all exciting documents relevant for the present study, and nobody in my school knew where to find them nor had them at hand, except of course for the R'94 curriculum. The reactionary attitudes and reluctance experienced by the Exam Secretariat (Talberg 2004⁶⁵), (Gulbrandsen 1997:25) on the part of the teachers may very well be due to a lack of information. A one day course is definitely not enough for changing direction of thought and practice! There are poor routines in the school system for spreading

⁶⁵ See below section 9.3

this kind of information, and I fear too few actually seek it on their own. It would indeed have been interesting to carry out a large scale survey among upper secondary language teachers, especially in the vocational branches, in order to find out what they really know about, not to say understand of the R'94 intentions, control documents and guidelines. To obtain interrater reliability⁶⁶, which is just as important as reliable tests, a different strategy will have to be chosen when introducing the new reform.

As mentioned above, I found only one scoring rubric⁶⁷ in active use in relation to VF2210. This was very general, and best applicable to written composition tasks. The fact that it doesn't even mention reading or listening comprehension does something about the assessor's focus. The scoring rubric distinguishes, as described above⁶⁸, between four different levels, but not between pass and fail, which is the level where rater disagreement most often occurs. The rubric is clearly based on similar scoring rubrics made for academic writing courses (Weigle 2002:190). It is leaning heavily on Bloom's taxonomy of Educational Objectives in the Cognitive Domain (Bloom et al 1956), and is meant to help the assessor evaluate the student's level of understanding. This taxonomy refers to 'the level of thinking skills or intellectual functioning required to accomplish certain tasks' (Weigle 2002:63), and is hardly earning its predominant role in this connection, putting reproduction at the lowest level and generating ideas at the highest, when the two may be found as juxtaposed prompts in the same test.

In my opinion, this difference in level of difficulty between different writing tasks should be made clear to the students through the instructions, as I believe that only a few of the examinees are aware of the consequences of choosing one prompt for another. They tend to choose according to topic and not task, and will not necessarily see the difference between '*discuss*' and '*tell about*'.

In his report from 1997⁶⁹, Olav Talberg expresses his worries as to the 'professionalism' of the locally made exams, seeing that they tend to be too local at times. He sees this as an argument for regionally made exams, but in such a way that several schools collaborate on making the exam test for a specific area of study to meet the need for a national and uniform

⁶⁶ By this I mean consistency of ratings among different raters, see also section 5.1

⁶⁷ See section 5.1

⁶⁸ See section 5.1

⁶⁹ See section 3.3

standard of the locally made exam sets. Comparing Talberg's report to the results of the present study, it seems that other points of worry have become more predominant. As seen above, most counties now do make regional exam tests, but there is little cooperation found among the schools. Due to practicality, most counties appoint committees of two-three teachers to design a common test for all the schools and all the different areas of study based on propositions from two-three randomly chosen schools each year. It will probably not be easy to meet the principles for test usefulness⁷⁰ as long as the tests are centrally made.

⁷⁰ See section 5.1

9 Current and future trends

9.1 Council of Europe – ‘Framework’

The latest guidelines from the Council of Europe, *‘The common European Framework of Reference for Languages’* (CoE 2004) is an important document describing in a comprehensive manner the competence necessary for communication and the related knowledge and skills. It is developed through a process of scientific research and wide consultation, and provides a common reference base for language learning and teaching in Europe, and a tool for assessment of proficiency. In addition, it is a legitimate child of the communicative movement in its approach.

It has definitely had a strong influence on the Norwegian 2006 reform ‘Knowledge Promotion’, first introduced in *‘Culture for learning’* (St.melding 30, 2003/2004) last year. New curricula and syllabuses for the different subjects are being developed right now, and the curricula for the common general subjects are already published in a preliminary edition (Directorate of Education – *Discussion Document* 2005).

As stated above, ‘Framework’ is a common basis for language syllabuses, curriculum guidelines, assessment, textbooks etc. across Europe, emphasizing what learners have *to learn to do* to use language for communicative purposes. It is concerned with the cultural context of language, and defines six different levels of proficiency. The main intention is to be a common platform of reference in the planning of learning programmes throughout Europe.

According to ‘Framework’, learning programmes and certification can be global, modular (restricted areas, e.g. vocational English), weighted (give learners different profiles) or partial (only “reading” or “oral mode”). The idea of ‘Partial qualifications’ are meant to contribute to plurilingualism in Europe; for example a little of each of the small languages based on the idea of *linguistic diversity in Europe*. It is an aim to preserve the variety of languages and in addition to ensure communication and understanding across states.

Other main areas are related to the planning of language certification, e.g. examinations, assessment criteria and test instruments common to all European countries, and the planning of self-directed learning, in which self assessment, e.g. DIALANG, plays a natural part.

‘Framework’ seeks to provide a common basis for description of objectives. Important principles are transparency, making it possible to compare systems across borders, international cooperation and the development of objective criteria of language proficiency to ensure European mobility.

Skills

‘Framework’⁷¹ provides a new way of seeing the basic skills, for example separating spoken interaction and prepared spoken production. This leads to a new division of main categories as well, substituting the former receptive ⇔ productive-axis with a division into understanding, speaking and writing skills.

Table 11: Basic skills

Traditional:		Framework (CoE 2004:26-27):	
Receptive	Listening	Understanding skills	Listening
	Reading		Reading
Productive	Speaking	Speaking skills	Spoken interaction
	Writing		Spoken production
		Writing skills	Written production

Descriptors

‘Framework’ provides examples of descriptors in terms of “can do” (CoE 2004:25), meant to be objectives and targets for teaching at different levels, and applicable as feedback from testing, i.e.: ‘NN can understand...’ Another area of focus in ‘Framework’ is the phrasing of objectives in self assessment. Self assessment seems to be an important concept in current language learning and teaching theory, but will only have relevance for low-stakes use and is not an option in exam assessment.

⁷¹ All details are based on the CoU 2004 ‘Framework’ document

Competence levels

‘Framework’ has defined six common reference levels of ability (CoE 2004:23-24), from A1 as the beginner level to C1 as complete mastery. The Norwegian Directorate for Primary and Secondary Education (see below) has already employed these reference levels in the new national system for quality assessment, the so-called ‘National Tests’.

Table 12: Common reference levels

A Basic user	A1 Breakthrough
	A2 Way stage
B Independent user	B1 Threshold
	B2 Vantage
C Proficient user	C1 Effective op. proficiency
	C2 Mastery

Communicative language competence

Finally, a radical simplification has been done with respect the classification of communicative competence, from six competence components in the traditional system, down to three in ‘Framework’. The former six competence components are not removed; they have been included in other components (Van Ek and Trim 1990: iv) and (CoE 2004:13).

Table 13: Components of communicative competence

Previous CoE (Threshold level)	CoE Framework
Linguistic competence	Linguistic competence
Sociolinguistic competence	Sociolinguistic competence
Socio-cultural competence	
Discourse competence	Pragmatic competence
Strategic competence	
Social competence (will & skill to interact)	

9.2 The Norwegian ‘Knowledge Promotion’ Reform 2006

The Directorate for Primary and Secondary Education in Norway was established on 15 June 2004. It is responsible for the development of primary and secondary education, and is the executive agency for the Ministry of Education and Research. The Directorate is to develop, organize and implement a national system for quality assessment designed to help to give pupils and apprentices a better general and vocational education, and for preparing about 300 new curricula providing guiding principles for education in the years to come. In addition, they will prepare all final examinations for primary and secondary schools. The new reform is called ‘The Knowledge Promotion’ (*Kunnskapsløftet*), and is a comprehensive reform of the complete primary and secondary school system.

9.2.1 ‘Høringsutkastet’ – the discussion document of the new EFL syllabus

In the discussion document⁷² we find four main areas, seen as four competences. These competences are language learning competence, communicative competence, language & culture, and literature & arts (Directorate of Education – *Discussion Document* 2005: 33).

The syllabus describes five basic skills, as found in Framework as well as indicated above (see section 9.1). These are reading, listening, writing, spoken interaction and spoken production. The basic skills are complemented by mathematical skills (handling numbers, graphs etc in English) and the use of digital tools.

Specifications

There are lots of specifications in this syllabus, in terms of operationalisation statements like: “The student should *be able to* ...” The specifications apply to functions as well as to basic skills. If these are to be the descriptors for assessment, they are indeed very ambitious! However, it seems that *aims* are decisive for the study at all levels and that *means* to a large extent are open for individual choice. There is in other words nothing to be found about teaching methodology, so this seems to be a syllabus of objectives only.

⁷² Based on ‘Læreplaner for Kunnskapsløftet’ – *Discussion Document*. Directorate of Education 15.02.05

Assessment

When looking closer into the descriptions of the upper secondary foundation course, there is nothing to be found about assessment, except for the organization. When it comes to exam assessment, we find that the combined exam test found in R'94 has disappeared. The final assessment comprises the complete 5-period subject, and the students may be randomly selected for either a written examination or an oral examination. The written part will be designed and scored centrally, and the oral part will be designed and scored locally.

9.2.2 'Knowledge Promotion' - The new EFL syllabus

The new curriculum of English, comprising all levels of teaching in primary and secondary education, was published in august 2005 and is to be implemented in the Norwegian school system from August 2006. The intention of introducing a new curriculum seems to be an even clearer focus on all aspects of communicative competence, and to provide a syllabus with fewer details, but with distinct aims and clear objectives. In this way it will be easier for parents and students to know exactly what to learn. It is meant to make education more focused on the result of the teaching, aiming at actual learning, not activity in its own right. The reform all in all put heavy emphasis on the basic skills. As in the preliminary discussion version, we find nothing about methodology in this curriculum.

However, some structural changes have been made in the final version, and there is even less focus on English specific to the different areas of study. The table below (on the next page) is not a quote, but a summary in my own words to make visible the development from the discussion document to the final curriculum.

Table 14: Comparison between the discussion document and the final document

	‘Høringsutkastet’ ⁷³ ,	‘Knowledge Promotion’ ⁷⁴ ,
C O M P E T E N C E S	<p>Four main areas:</p> <ul style="list-style-type: none"> ▪ language learning competence ▪ communicative competence <ul style="list-style-type: none"> linguistic sociolinguistic pragmatic ▪ language & culture (existential/social competence) ▪ literature & arts (knowledge of the world) 	<p>Three main areas:</p> <ul style="list-style-type: none"> ▪ language learning competence ▪ Communicative competence <ul style="list-style-type: none"> Linguistic (5 basic skills) Sociolinguistic Pragmatic Social Sociocultural ▪ Culture, society & literature
B A S I C S K I L S	<p>Three basic areas:</p> <ul style="list-style-type: none"> ▪ Ability to read and to express oneself orally and in writing: <ul style="list-style-type: none"> reading listening writing spoken interaction spoken production ▪ mathematical skills (handling numbers, graphs etc in English) ▪ use of digital tools <ul style="list-style-type: none"> Use a wide selection of digital tools... 	<p>Four basic areas:</p> <ul style="list-style-type: none"> ▪ Reading ▪ Ability to express oneself orally and in writing: <ul style="list-style-type: none"> listening writing spoken interaction spoken production ▪ mathematical skills (handling numbers, graphs etc in English) ▪ use of digital tools <ul style="list-style-type: none"> Use a wide selection of digital tools. Including <i>monolingual</i> dictionaries...
V O O A T I O N A L	<ul style="list-style-type: none"> ▪ write texts on personal, vocational and cultural subjects ▪ use technical and mathematical information in communication ▪ communicate on topics related to different trades and areas of study ▪ write texts about literary and vocationally related topics ▪ describe a working process from idea to product 	<ul style="list-style-type: none"> ▪ use technical and mathematical information in communication ▪ Choose one in-depth cross-curricular topic within own area of study and give a presentation of it <p>Other ‘news’:</p> <ul style="list-style-type: none"> - Focus on mastering a wide <i>vocabulary</i> - The concept of <i>fluency</i> mentioned - Reading English-speaking literature from 1500s to modern times

⁷³ Based on p.33-34 + 38-40

⁷⁴ Based on p.89-91 + 95-96

In the following, I will try to sum up the most striking changes that have taken place from R'94 to the 'Knowledge Promotion Reform':

- A total of 5 periods will be compulsory for all students. As to the vocational courses, it will be comprised by 3 periods a week in the first year and 2 periods a week in the second year. In the assessed attainment at the end of the course, the students will be given one single grade as now.
- The current combined exam test will disappear. Exam assessment will to a large extent be similar to the current 5 periods-a-week-assessment, common to all areas of study, with two separate exams for oral and written competence respectively. The student may be randomly selected for one or both. The written part will be centrally designed and assessed, while the oral part will be locally designed and rated
- There is a renewed interest in reading and vocabulary in the new document. *Reading* is listed as one out of four main areas of basic skills. The other three are
 - *Ability to express oneself orally and in writing*, including the basic skills of speaking, both interactional and transactional, listening and writing.
 - *Mathematical skills*, implying mathematical vocabulary and the competence of utilize information from graphical representation, tables, charts and statistics.
 - *The use of digital tools*, opening for authentic use of the English language.
- English in relation to the different areas of study, however, seems to be a neglected area in the 'Knowledge Promotion'-document. As shown above, the syllabus says nothing about vocational input, and very little about learning a vocational vocabulary. *One* cross-curricular project to be presented in English is all that is required, and it will not be specifically tested in the formal summative assessment, at least not in the written part which is centrally made and common to all students.
- There is a heavy emphasis on language learning strategies and self-directed learning in the new document. The importance of using digital tools is stressed, including the ability to produce multimodal texts. In addition, the students should be able to:
 - use and evaluate different situations, working methods and strategies to learn English
 - compare English to other foreign languages and utilize this knowledge in language learning
 - use relevant and precise terminology to describe formal and structural features of language
 - describe and evaluate their own progress in language learning
- Even more emphasis is given to communicative competence than in R'94, especially to pragmatic concepts. There seems to be a renewed interest in vocabulary as well, but mainly in the general English area, as the students should be able to:
 - master an extensive *vocabulary*
 - discuss the *writer's intention* (views and attitudes)
 - express oneself in a nuanced and situational way, with *fluency*, precision and *coherence*

- pick *appropriate* listening, speaking- and writing strategies
 - take initiative to starting, ending and keeping a conversation going ('keeping the floor')
 - read formal and informal texts of different genres and with different purposes
- As to content, 'background' knowledge about the English-speaking world has become the main target, at the sacrifice of vocationally oriented English. The former focus on the USA and the GB has been extended to comprise the different cultures of all English-speaking countries. Students should be able to:
 - Present and discuss international news and incidents
 - Explain main features of the development of English language
 - Analyze and discuss one movie and a representative selection of literary texts from different parts of the world and different periods in time
 - Discuss literature by and about ancient people in the English-speaking world
 - Make and evaluate own written and oral texts inspired by literature and arts

As we can see from this outline, the new reform will have some important consequences for students in the vocational courses. They will meet exactly the same requirements as to content and attainment targets as students in the more academically oriented courses, and they will be assessed with the very same written exam in the formal summative assessment. The possibility of letting good oral skills compensate for poor writing competence seems to be lost with the disappearance of the combined exam test. So does to a large extent the individual choice of the teacher - student group as to focusing on material and vocabulary specific to their area of study.

There is probably nothing stopping individual groups from focusing on subject-specific material, but it will in the future not be tested in the written exam. Except for one cross-curricular project work, none of the attainment targets focus on ESP. On the other hand, there are many academically oriented targets and requirements to be met. The consequences of these features has been discussed above, and will be commented on by an expert on teaching and learning English in the vocational branches of study below.

9.3 University College Teacher and Researcher Olav Talberg:

‘By focusing on vocational subjects, the students will develop their vocationally oriented language as well as their general English language skills’ (quoted from an interview October 13th 2005⁷⁵)

During my work with this thesis, I have come to consider Olav Talberg, language teacher and researcher at the Faculty of Engineering, Oslo University College, as one of the main experts on English as a foreign language in the Norwegian vocational courses. For that reason, I was eager to get his opinion on the current trends in foreign language teaching and testing, and I asked him for a comment on the new syllabus. This is what Talberg answered:

‘Initially, I would like to say that curricula and syllabuses are education-political documents reflecting what ‘education-politicians’ see as important aims of education. This applies to targets of knowledge and skills in English as well. Within the timeframe awarded to a certain subject, one have to make priorities if there are many different and specific aims to be obtained. An alternative is to make more general formulations of the targets so that they will not appear as excluding any of the involved ‘target areas’. In the latter case, a syllabus may give room for adaptation of the teaching in such a way that ‘excluded’ targets still can be attained within given areas of study. As far as I can see from the new English syllabus, the choice has been made to express the aims in very ‘general’ terms, and in the way I read this syllabus, there is no opening for meeting the more ‘specific’ needs of the different areas of study.

The R’94 English syllabus had clear formulations of targets with respect to skills and knowledge. It gave room for a 40 – 60 % balance between teaching material from ‘vocational’ and more ‘general’ areas. This balance opened for the possibility that language teachers in the different areas of study could make use of teaching material that was adapted to the vocational branch or the area of study of each specific group of students. To my experience, such teaching material was motivating for the students to work with. The fact that many language teachers had difficulties coping with this is a different question. However, syllabuses should not be developed to make it ‘comfortable’ for the teachers to teach’.

I then explicitly asked for his view on the current focus on more general knowledge and communicative abilities even for the vocational branches of study, and Talberg continued:

⁷⁵ My translation

'I am thoroughly sceptical towards the current development for several reasons. Firstly, it seems to me that the 'education-politicians' do not believe in, or are unaware of, the fact that general English skills may be developed through working with 'vocational' or 'vocationally oriented' input. To my experience, based on 20 years of teaching English in the vocational courses, such teaching material rises the motivation among the students in addition to developing their language abilities at least to the same level as texts and exercises of a more general nature do. Secondly, a lot of excellent 'vocationally oriented' fiction exists that is now in danger of being 'excluded' by the teachers. Thirdly, most English teachers in upper secondary school have an academic language education themselves. The general formulations in the new syllabus will reinforce the use of 'academically oriented' teaching material and educational input based on 'general English', and this in turn may feel de-motivating for students in the vocational courses. Finally, the lack of vocationally oriented English input in the vocational areas of study will deprive these students of the possibility of getting to know the use of English typical for their future working life'.

It is often stated that 'the testing should reflect the teaching'. When I asked Talberg if he felt this was the case under the current system, he replied as follows:

'First of all, my view is that the assessment should reflect the individual student's competence (knowledge and skills) of English, as stated in the attainment targets. It is meaningless and far too superficial to claim that 'the assessment should reflect the teaching'. What goes on in the classroom isn't necessarily in accordance with the curriculum; there are plenty enough examples of that, even during the R'94 period. R'94 linked assessment directly to the attainment targets in the syllabus, and descriptors were indicating the different levels of competence. In my opinion, this system worked well – after some time very well – even if some of the 'teacher brigade' always had objections and questions. Whether or not assessment will work in the same way under the new reform 'Knowledge Promotion' – and how it will work – is beyond my knowledge'.

Finally, I wanted Talberg's opinion on future perspectives, and I asked him what he would consider an appropriate and serviceable development of English as a foreign language for the student group in question.

'In my opinion, students in the vocational courses would be better off with the existing R'94 syllabus than with this new one. I have stated the reasons for this view above. Some of the problems in connection with the implementation and carrying out of the R'94 syllabus of English were due, not to the syllabus itself or to the students, but to the attitudes of some in the teacher brigade towards a) the implementation of R'94, b) the personal reluctance from some in the teacher brigade as to teaching English in the vocational areas of study, c) their reluctance against having to get familiar with the different vocational areas of study and d) their own academic background and orientation towards 'General English' and 'Literary Studies'. A new debate and evaluation of English under R'94, extended

by 'examples of good practice', would have supplied the school subject and the new syllabus with a lot of useful and meaningful material. As an example, the text types the students are expected to write could have been extended, and more examples could have been provided to avoid confusion (of which there admittedly has been quite a lot)'.

In conclusion, Talberg seems to worry deeply on behalf of the vocational students. Their specific needs seem not to be taken care of in this reform, and he feels they would have been better off with an improved version of R'94. He underlines that general English skills may be developed through working with 'vocationally oriented' input, and that this kind of teaching material rises the motivation among the students in addition to developing their language abilities at least to the same level as texts and exercises of a more general nature do.

9.4 County Director of Education Ragnhild Stai Amundsen:

‘I see foreign language learning as an entrance to understanding yourself as part of this world’ (quoted from an interview November 8th 2005⁷⁶)

Ragnhild Stai Amundsen, now Director of Education at the County Governors’ Education department in Østfold, has been working with foreign languages and school development all of her life. She has been a teacher of foreign languages, a pedagogical advisor and a county school administrator. She has proved to be a well orientated and curious representative of the educational bureaucracy, and the following is in essence a presentation of her views on some current trends. Initially, I wanted her opinion on the even more academical approach to foreign language learning in the new curriculum for the vocational branches of study.

Amundsen replied:

‘First of all, I’d like to change your point of departure. To talk about ‘vocational students’ as a special kind of students is a thing of the past in my opinion, and does hardly make no sense in our time. The group is not at all homogenous, and there is high time we change the picture of the stereotype ‘vocational working class student’ in our minds. You will often find academically brilliant students in vocational areas of study, and they will want to use their language classes for more than learning a few simple speech acts and some vocationally oriented vocabulary. In fact, I mean this applies to all students. I see foreign languages as school subjects suitable for identity development: Language and culture are equally important and valuable for all students, no matter what their future plans may be! Even the traditional technical and industrial branches have changed. The ‘global village’ has changed the working marked. Young craftsmen will want a wider competence, and they too need to be internationally oriented. I do not think the answer is more ‘vocationally oriented English’, but rather an arena for ‘identity building’ for all. I agree that we gain a lot by basing the teaching on the students’ area of interest, but it seems to me their main interest at the moment is being cool. We will do wrong to many students if we choose to limit the possibilities for certain groups. At least in that way, the new reform is heading in the right direction’.

Her opening line effectively discharged all questions about special needs of the ‘vocational’ students, so I asked her for a general comment on the new reform ‘Knowledge Promotion’:

‘In relation to the reform, the Ministry of Education and Research recently launched ‘*Språk åpner dører*’ (UFD 2005), a strategy document for strengthening the foreign languages in Norwegian primary and secondary education. This document calls on a ‘practical approach’ to foreign language teaching and learning, but in my opinion, the concept of ‘practical’ is given a too narrow definition! If ‘practical approach’ means providing the students with nothing more than the linguistic tools they need to behave like ‘good customers’, I’m deeply worried. I see foreign language learning as an entrance to understanding yourself as part of this world, and to create curiosity and understanding in the student with respect to other people and foreign places’.

⁷⁶ My translation

Based on my worries about the new organization of exam assessment, testing either oral or written competence, I expressed my view on the position of oral competence in current language teaching and assessment. She did not share my concern:

'To my experience, it is not at all being neglected in our schools today. Norwegian students have generally good oral skills in English, probably due to an early learning start. But they are not that good when it comes to writing, and good language skills include both competences. And as mentioned before, I find topical background knowledge to be an essential part of a foreign language school subject. I happen to think that written and oral language need to be taught simultaneously, and that students should work on and develop a language suitable for expressing their minds, showing their personality and being able to communicate with other people in order to get to know them personally and to exchange experiences and feelings. If the school subject has nothing to do with the student's own personal development, it will, I believe, never be 'practical'. All in all, I find the concept of 'practical approach' problematic. If it equals 'simplification', I'm thoroughly sceptical. Who are really the 16-year-olds of 2006? I often experience them to be both 'the cool and uncommitted' and 'the well-informed knowledge-seeking student' at the same time, and their main learning area seems to be outside school. This represents a real challenge for teachers these days.

Finally, I asked for a comment on the subject for the present thesis, the VF2210 examination. She used to be a regional level school administrator herself when R'94 was introduced, and remembers very well the discussion it raised:

'There was in fact ample disagreement about these tests. In my county, the option of regionally made exams was chosen, mainly out of two reasons: A need for regional thinking and cooperation on exam assessment, and a wish to develop the competence on assessment among teachers. The regional thinking helps ease the transition between schools for the students, often relevant in the vocational areas of study. A common exam test may in the future lead to more common teaching practice, and provide a common basis for discussions among schools about exam assessment and teaching in general. We now seem to be heading towards centrally made tests, and I'm not yet sure if this is a gain or a loss'.

In conclusion, Amundsen's point of view represents a very different perspective on foreign language teaching and the vocational student than the one put forward by Talberg. In her opinion there is no reason to see these students as a special group, and her main concern is as to what kind of language teaching the upper secondary student in general will need in the future.

In the next chapter, I will draw attention to some central findings of the study as a whole, and try to come up with answers to the questions asked in the introduction. Finally, I will present a conclusion on my main research question, and point at areas in need of further research.

10 Conclusion

This thesis has grown out of a feeling based on my own experience as a teacher that the present exam assessment in English in the vocational areas of study is not fair, and that the VF2210 test tool is unpredictable, unreliable and not always valid. To what extent have these assumptions been confirmed by the present study?

My sample of tests is, as explained in the introduction, a representative probability sample randomly selected from the population of regionally designed exam sets for the spring term exams in the period 2002 – 2005. With one exception, all counties are represented. This will guarantee that the data collected from the written part of the exam sets are reasonably reliable, and I hope to have attained both internal and external validity of this part of the study. As to the oral part, my data are based on five teacher interviews, some written instructions and my own experience. The sample of teachers was more of an accidental convenience sample (see p.8), as I selected the first five teachers to answer the inquiry distributed. Conclusions drawn from samples this small are normally not reliable, but some tendencies do appear and will be included to supply the overall impression of the VF2210 exam test.

The study reveals that a large amount of work is being done around the country, and that many fairly good tests and tasks are developed that should have been made available to everybody involved in the process. One foreseen consequence of local exam making is that the tests will differ. But the differences seem to occur between the different counties, and not according to the different needs of the groups of examinees, a fact that leads me to the conclusion that other than pedagogical considerations motivate the choices. The collected data demonstrate that the current VF2210 exam assessment seems to be relatively inconsistent, unreliable and unfair. To support the latter claim, the following paragraphs will sum up some tendencies found in the sample tests.

In most cases, the VF2210 exam test appears to be a traditional timed impromptu writing test on subjects general enough to be accessible to virtually all examinees, supplemented by a one page text on a current general subject with some kind of reading comprehension task related to it. In some cases, a vocational writing task is included as well. The test is assessed

subjectively by two teachers under time pressure, normally in the period in between the written and the oral part, and the provisional grade is included in a final global grade after the oral examination. Is this in accordance with the purpose of exam assessment and the R'94 curriculum and guidelines?

To some degree it is, but the intention of a local exam seems to have been lost along the way. The written part of the test is more of a general proficiency test than an achievement test, and could without losing much have been made into a central exam to guarantee same level of difficulty for all students. The intended use of 'real life' tasks varies to a large extent among the exam sets. Some sets try to put the students in authentic situations altogether, while others don't even bother to try. As to the concept of 'open tasks,' there seems to be confusion about what such tasks really are, and many attempts are made to produce such tasks, of which some are clearly in danger of overestimating the students' ability to come up with the expected topical content. R'94 highlights organized preparation as essential for assessing the student's holistic competence, but not many counties have chosen to include a preparation session. Stating the inequality of all this is all there is room for in this thesis.

By turning English into an academical subject even in the vocational courses, R'94 did make the situation worse for a considerable group of students. To some extent this study proves that the assessment of reading comprehension is not always giving the students a chance to perform to the best of their ability, neither do all Norwegian counties test their students at the same level of difficulty. The test sample reveals huge differences in text difficulty, in requirements in relation to level of reading comprehension and in the demand for written production. Production skills are in general attributed more weight than receptive ones in the VF2210 exam tests, and some targets seem to be systematically neglected, especially target 1, probably due to practicality.

A closer examination into some of the exam sets revealed the less fortunate students will have better chances of doing well in some counties than in others, and that not all counties test achievement in relation to the two targets of content to the same extent. Background knowledge about the USA and UK is subject to testing in all the exam sets, but only in general terms and often optional as one of the essay tasks. The literary syllabus is normally left for the oral examination, and so is in many cases the vocationally oriented English.

Assessing the individual student's learning yield in relation to target 6⁷⁷ seems to be best accounted for when the vocationally related task is made by the local teacher. In some cases this target will instead be tested in the written part in the form of an open task. All vocational task examples seem to test production exclusively. The ability to handle vocationally oriented texts is not tested specifically in any of the exam sets. In my opinion, in the written part of the regionally made test the intention of a local exam seems to a large extent to have been lost along the way in the written part of the regionally made test.

This study has, as stated above, not dug very deep into the oral part of the exam, but examples have been given of very different kinds of oral examinations. Some teachers still examine their students in topical content the traditional way, while others keep a focus on more communicative and performance-based aspects. Anyhow, it seems that the oral part is where actual achievement testing does take place, and examples are found where results in the grading of the students' oral performance tend to improve the final holistic grade of the students. The weighting between written and oral competence in the final assessment can only be finally verified after closer studies into this area.

Consequently, there is still work to be done in order to guarantee the VF2210 construct validity and reliability as a test tool, but it will probably be the scoring of the test and the intuitive weighting in the final global assessment that constitute the most important factor for the test result. The problem of interrater reliability has in my opinion not been given enough attention, and in my experience the consistency of assessment among different raters is not good enough. This issue clearly calls for further research.

As demonstrated in chapter 9, a change in the organization of exam assessment will be implemented with the 'Knowledge Promotion Reform'. In my opinion, when centralizing the written part, the locally made oral examination should be made a compulsory part of the exam assessment! I will not predict the consequences of the new curriculum. 20 years of teaching scarcely make me an authority on the subject, but I dare argue that taking away the combined exam is not to the benefit of the students. According to the new curriculum, the students may be randomly selected for either a written **or** an oral examination. This will definitely not improve reliability of the exam assessment.

⁷⁷ Target 6 = 'English in relation to the student's area of study'

In conclusion, the data from the present study reveal that the VF2210 exam test does not in all cases test achievement. The students in the vocational areas are to a considerable extent exposed to general proficiency testing, and I still feel confused as to the intended purpose of the test. I experience a gap between the R'94 intentions, the current practice in teaching, and the actual exam assessment. My initial assumption that the VF2210 exam test is not always a reliable assessment tool has to a large extent been confirmed by the results of the study. Supported by the analyses above, I dare draw the conclusion that exam assessment by means of the VF2210 exam test under the current system is not fair to all students!

11 Bibliography

Alderson, J.C. 2000. *Assessing reading*. Cambridge: Cambridge University Press

Björnsson, C.H. 1968. *Läsbarhet*. Stockholm: Bokförlaget Liber.

Brown, J.D. 1996. *Testing in language programs*. Ch.1 and 2. London: Prentice-Hall International

Buch, G. 1997. The Testing of Listening in a Second Language. In C. Clapman and D. Corson (eds.) *Encyclopedia of Language and Education. Volume 7 Language testing and Assessment*. Dordrecht: Kluwer Academic Publishers

Ellis, R. 1998. *Second language acquisition*. Oxford. Oxford University Press.

Fulcher, G. 1997. The testing of Speaking a Second Language. In C. Clapman and D. Corson (eds.) *Encyclopedia of Language and Education. Language testing and Assessment*. Dordrecht: Kluwer Academic Publishers

Gulbrandsen, K. 1997. "Erfaringar med språkfaga i Reform'94". In *Språk og språkundervisning* no.1, 24-26

Gurholt, P. 1997. "Ett skritt fram og to tilbake?". In *Språk og språkundervisning* no.4, 21-24

Hymes, D. 1972. "On communicative competence". In J.B. Pride & J. Holes (eds.). *Sociolinguistics: Selected Readings*. Harmondsworth: Penguin books.

Korsvold. A.K. 1997. *Discourse competence as seen through the English textbook*. Master thesis. Oslo: University of Oslo.

Pettersen, S. 2001. *The Foundation Course in English: Some aspects of the written exam*. Master thesis. Oslo: University of Oslo.

Read, J. 1997. Ch.10. Assessing vocabulary in a second language. In C. Clapman and D. Corson (eds.) *Encyclopedia of Language and Education. Volume 7 Language testing and Assessment*. Dordrecht: Kluwer Academic Publishers

Schmitt, N. 2000. *Vocabulary in language teaching*. Cambridge University Press

Simensen, A.M.1998. *Teaching a foreign language - Principles and procedures*. Bergen: Fagbokforlaget

Talberg, O. 2004. 'Teaching Vocationally Oriented English'. In *Språk og språkundervisning* no.2, 7-11

Talberg, O. 1997. *Lokalt gitte eksamensoppgaver i engelsk for firetimersfaget*. Report on the basis of an inquiry made for The Norwegian Exam Secretariat. Not published.

Thorenfeldt, A.E. 2004. *The formal summative assessment of English as a foreign language in the vocational courses of Norwegian upper secondary education after Reform'94*. Paper. Oslo: University of Oslo.

Thorenfeldt, A.E. 2004. *The testing of reading comprehension in the summative assessment of English in the vocational courses 2003-2004*. Paper. Oslo: University of Oslo.

Ulriksen, K.2002. *Teaching English in the vocational classroom*. Master Thesis. Trondheim: Norwegian University of Science and Technology

Weigle, S.C. 2002. *Assessing writing*. Cambridge: Cambridge University Press

Wood, D. 2001. "In Search of Fluency: What is it and how can we teach it?" *The Canadian Modern Language Review*. 57 (4). Toronto: Toronto University Press

Curricula and guideline documents

Council of Europe. 1975. *The Threshold Level*. Cambridge: Cambridge University Press

Council of Europe. 1990. *The Threshold Level*. Cambridge: Cambridge University Press

Council of Europe. 2001. *The Common European Framework of Reference for Languages: Learning, teaching and assessment*. Cambridge: Cambridge University Press

Core curriculum for Primary, Secondary and Adult Education in Norway – ENGLISH

1993. (R94) Oslo: The Ministry of Education, Research and Church Affairs

English. Common general subject for all areas of study. Curriculum for Upper Secondary Education - ENGLISH 1993. (R94) Oslo: The Ministry of Education, Research and Church Affairs

Engelsk. Felles allment fag for alle studieretninger. Læreplan for videregående opplæring. 1993. (R94) Oslo: Kirke-, utdannings- og forskningsdepartementet

Metodisk rettleiing. Grunnkurs Engelsk. Felles allment fag for alle studieretninger. 1994. Oslo: Nasjonalt læremiddelsenter.

Metodisk rettleiing. Vurdering i vidaregåande opplæring – 2.utgåve. 1997. Oslo: Nasjonalt læremiddelsenter.

Læreplaner for gjennomgående fag i grunnskolen og videregående opplæring. Midlertidig

trykt utgave – september 2005 Oslo: Utdanningsdirektoratet

Læreplaner for Kunnskapsløftet - Høringsutkast fra Utdanningsdirektoratet 15.02.05

<http://www.utdanningsdirektoratet.no> (accessed 05.03.05)

Språk åpner dører. Strategidokument for styrking av fremmedspråk i grunnutdanningen 2005-2009. 2005 Oslo: The Ministry of Education and Research

http://odin.dep.no/filarkiv/253707/Sprak_apner_dorer.pdf (accessed 05.11.05)

SUE/vg – 95-005. *Sentrale retningslinjer for lokalt gitt og sentralt gitt eksamen.* 1995. Oslo: Statens Utdanningskontor i Oslo og Akershus. Eksamenssekretariatet.

SUE/vg – 97-021. *Om forberedelsesdel til lokal og sentral eksamen i Reform'94.* 1997. Oslo: Statens Utdanningskontor i Oslo og Akershus. Eksamenssekretariatet.

Vurderingsrettleiing for munnleg eksamen i språkfag – førebels utgåve. 1997 Oslo: Statens Utdanningskontor i Oslo og Akershus. Eksamenssekretariatet.

Vurderingsrettleiing i engelsk skriftleg i vidaregåande opplæring. –førebels utgåve. 1994: Statens Utdanningskontor i Oslo og Akershus. Eksamenssekretariatet.

<http://www.utdanningsdirektoratet.no/dav/87d10709b4.pdf> (accessed 22.06.05)

Vurderingskriterium. Vurderingsrettleiing i engelsk – førebels utgåve. 1999. Oslo: Statens Utdanningskontor i Oslo og Akershus. Eksamenssekretariatet.

Web sites

The FOG index: <http://www.usingenglish.com/glossary/fog-index.html> (accessed 15.08.05)

The FOG index: http://en.wikipedia.org/wiki/Gunning-Fog_Index (accessed 15.08.05)

Text content analysis tool: <http://www.internetformidling.dk/index>. (accessed 03.07.05)

Readability: <http://www.textquest.de/readability.htm> (accessed 03.07.05)

Readability: <http://www.readability.info/info.shtml> (accessed 03.07.05)