The course grade in VG1 and 10th grade English:

How do teachers assess oral proficiency towards the final course grade?

A mixed method study

Alma Cosabic

Masteravhandling i engelsk fagdidaktikk ved Institutt for Lærerutdanning og skoleforskning
Utdanningsvitenskapelige fakultet

UNIVERSITETET I OSLO

Vår 2016

The course grade in VG1 and 10th grade English:

How do teachers assess oral proficiency

towards the final course grade?

© Alma Cosabic

2016

The course grade in VG1 and 10th grade English:
How do teachers assess oral proficiency
towards the final course grade?

Alma Cosabic

http://www.duo.uio.no

Sammendrag

I tiende klasse får elevene to separate karakterer i engelsk – en muntlig og en skriftlig karakter. På VG1 derimot, får eleven en samlet karakter som skal beskrive både elevenes muntlig og skriftlige kompetanse. Hvis man tar utgangspunkt i kompetansemålene fra læreplanen i engelsk (LK06) er det svært få ulikheter fra 10.klasse til VG1, ved unntak av blant annet krav om større selvstendighet og refleksjon. Målet med denne studien var å få et innblikk i hvordan lærere på VG1 og i 10. klasse, ender opp på sluttkarakteren i faget.

Studien har forsøkt å belyse lærernes praksis rundt vurdering av elevenes muntlige ferdigheter, rutiner for vurdering, bruk og vektlegging av vurderingskriterier, hvilke vurderingssituasjoner som blir brukt og hvor ofte det legges til rette for muntlige vurderinger, sammenliknet med skriftlige. Problemstillingen jeg formulerte er: *How do teachers assess oral proficiency towards the final course grade?* I tillegg til problemstillingen, ble tre underspørsmål formulert, for å gi en mer oversiktlig oppgave. Metoden brukt i undersøkelsen er en blanding av kvalitativ og kvantitativ metode, såkalt «mixed method approach».

Datamaterialet ble samlet inn ved å intervjue syv lærere, først ved å besvare en spørreundersøkelse, deretter et oppfølgingsintervju over telefon, hvor jeg har brukt åpne spørsmål. I tillegg til dette, samlet jeg inn datamateriale fra 27 lærere ved å bruke et digitalt spørreskjema. Her brukte jeg i hovedsak spørsmål med kategorier, hvor informantene blant annet måtte rangere svarene etter viktighet.

Hovedfunnene mine er analyserte i oppgaven, der både statistiske og kvalitative data blir presentert. Funnene peker mot ulik praksis og variasjon i bruken av vurderingskriterier og i selve vurderingsprosessen. Dette samsvarer med resultater fra annen forskning på vurdering av muntlige ferdigeter.

Abstract

There are two separate grades in English at 10^{th} grade – one to describe the students' oral proficiency, and one to describe the written proficiency. At VG1 there is only one grade to describe both skills. The competence aims in the English subject curriculum (LK06) on the other hand, are quite similar for both levels.

The aim of this study was to shed light on teachers 'practice in assessment of students' oral proficiency, the procedures around assessment, how teacher emphasize the different assessment criteria, and how often they facilitate oral assessment compared to written assessment. My research question was: The course grade in VG1 and 10^{th} grade English: How do teachers assess oral proficiency towards the final course grade? In addition to the research question, three sub-questions were formulated, in order to provide context. The method used in the study is both qualitative and quantitative methods, so-called "mixed method approach."

The data was collected by interviewing seven teachers, first by answering a questionnaire, then a follow-up interview by phone, where I have used open-ended questions. In addition to this, I gathered data material from 27 teachers using a digital questionnaire. Here I mainly used questions with categories, where informants among others things had to rank the answers according to importance.

The main findings were then analyzed, and both statistical and qualitative data were presented. The findings point toward different practices and variation in both in the use of assessment criteria and in the evaluation process. This is consistent with results from other research on assessment of oral proficiency

Acknowledgements

First of all, I want to express my deepest appreciation to Phil Grey, PhD for giving me advice and for being very patient with me throughout the process of working with this master thesis.

Furthermore, I would like to thank all of my 34 informants who contributed with answers for both my survey and my interview. I appreciate the time you have spent in order to answer all my questions. Without your contributions there would be no master thesis.

I would also like to thank my family and friends who have encouraged me at times when I wanted to give up. Most of all, I thank my mother for being such a great support through this, as with everything I go through, for inspiring me and for always believing in me. Hvala mama!

I would like to thank my colleagues and students at Landøya School and Bjørnsletta School for giving me inspiration in writing my thesis, for being patient with me while I was studying and providing me good experience. Last but not least, thanks to my former teacher Anne Bunkholt, who without ever knowing it, influenced me to become a teacher myself.

Oslo, May 2016

Alma Cosabic

Table of contents

List of tables

1 Introduction

1.1. Background

When I first began my master study, I had a difficult time deciding what my master thesis would be. I had already worked as a teacher for three years and there were so many topics that I found interesting. When I first met my supervisor I told him about some of my ideas, before he asked me “what is the most challenging thing you have come across, working as a teacher?” The answer was simple – assessment, especially assessing and giving feedback on student’s oral proficiency.

When I talked to my colleagues about this, I discovered that I was not the only one experiencing these challenges. How is it possible to grade the student’s oral proficiency fairly? Which situations are best suitable for assessment? How is it possible to ensure that we make good assessment criteria? I wanted to find answers to all of these questions.

Many articles have been published about the oral examinations; however there are few articles about the final course grade. The Norwegian national newspaper Aftenposten published an article in June 2013 about oral examinations, where they among other things revealed big differences in how oral examinations are carried out from county to county. After this article was published it also became clear that this issue was unknown for Norwegian politicians.

Aftenposten also published an article in March 2016 about a suggestion by the second largest youth political group in Norway, The Norwegian Young Conservatives, to record oral examinations at lower secondary and upper secondary schools. The suggestion is based on the number of complaints on written examinations from 2013, where 11 000 complaints were filed. 1 out of 6 complaints resulted in adjustments for higher grades than what was originally given. Unlike written examinations, it is not possible to complain on oral examinations. The Norwegian Young Conservatives wish to implement the possibility to record oral examinations in order to strengthen the legal safeguards for the minority of students who feel unfairly treated during oral exams.

Assessment in general is a highly debated topic among teachers and politicians, students and parents. Assessment should be as objective as possible and according to the Education Act § 3-1, pupils in public primary and secondary education, and pupils, apprentices and trainees in public secondary schools have the right to be assessed. The objectives for assessment should be known for the pupils and what is being emphasized in assessing his or her competence. The schools administer has the responsibility to ensure that the pupils', apprentices' or trainees' right for assessment is met (my translation) (Education Act, 2009).

Assessment plays a highly important role in teaching. A 2012 study on the role of raters' attitude in the oral assessment, by Shafipoor & Latifa, showed that teachers with different educational backgrounds have completely different attitudes to oral assessment. The results from this study suggested that teachers should have similar rather relevant pedagogical knowledge and experience in order to decide efficiently and fairly in oral assessment. This goes in line with the work of Brown (1995) and McNamara (1996), which proved that rater's background pedagogical knowledge and experience can affect assessment validity (Shafipoor & Latifa, 2013).

Henrik Bøhn conducted a study on performance aspects teachers' pay attention to in EFL oral exams. Among other findings, the results showed that there are inconsistencies in how teachers value the constructs and criteria and that there are variations in how teachers view the significance of content knowledge (SAGE, 2015).

1.2 Related studies

Studies on teachers assessment is far from unknown in this field of research. However, there are few studies related to assessment on oral proficiency and the course grade combined. I was not able to find any studies on this particular topic. Therefore I chose to look into studies on assessment, grading and oral proficiency separately.

Bruce Frey and Vicki L Schmitt's (2010) study on teachers' classroom assessment practices examined classroom assessment practices of 3rd- through 12th-grade teachers in a Midwestern state. In addition to determining the frequency with which specific assessment item formats were utilized, the level of use of selected "best practice" approaches to assessment was considered ("performance-based assessment, teacher-made tests, and formative assessment"). Some of their findings were that essays and written assignments were the most common assessment formats reported. Their results pointed out that there is substantial use of performance-based assessments across grade levels and subject, but traditional paper-and-pencil testing remains the predominant classroom assessment format. In addition, the study showed that female teachers choose performance-based assessment more often than male teachers and that performance-based assessment is used much more frequently by language arts teachers than by those who teach other subjects and is more common at higher levels than at the elementary level. Though teachers design their own classroom assessments, they routinely rely on tests or items written by others. Formative assessment is not common, as only about 12% of assessments do not affect student grades and 3 out of every 4 assessments are administered after instruction is completed. This study is related to my study because it revealed that written assignments are the most common assessment forms, in addition to the fact that teachers frequently use assessment criteria made written by others.

A study on VG1 level oral examinations by Yildiz (2011) showed that there are many variations concerning the design and the assessment of English oral examinations at VG1 level. In her conclusion she also implies that common assessment criteria would make assessment easier, in addition to provide assurance that the results of the examinations are correct. (Yildiz, 2011)

To sum up, several Norwegian and international studies have shown that there are inconsistencies connected to assessment and assessment criteria. Most of the studies I have mentioned above are based on written and oral examinations, however, there is a big possibility that the same problems occur when the teachers decide on the final course grade in English. To prevent inconsistencies connected to assessment, specifically assessment of oral proficiency; it would be useful to do more research on this topic. That is also the reason I chose to write my master thesis on this topic.

1.3 Research statement

My research statement is aimed at teachers in the lower secondary school (10^{th} grade) and upper secondary school (VG1 level). The basis for this study is the impression from working as a teacher for a few years, that there is no clear definition on assessment of students oral proficiency and that there are big differences in what teachers view as important, and how often they assess students before setting the course grade in English.

The purpose of this study is to examine the procedures teachers use when deciding the final course grade in English, whether there are big differences from 10^{th} grade where there are two separate grades in English and VG1 where there is only one grade, and whether something should be done to make it easier for teachers to set the course grade.

Therefore, my research question is:

The course grade in VG1 and 10^{th} grade English:

How do teachers assess oral proficiency towards the final course grade?

To help me answer the research question I developed a couple of subsidiary questions:

- How often are teachers assessing students' oral proficiency compared to other types of assessment e.g. written tests?
- Which criteria are used when assessing students' oral proficiency and who makes the criteria?
- Does classroom size play any role when assessing the student's oral proficiency?

The findings from the questionnaires I am going to present in chapter 6, continued with a discussion of my findings in chapter 7. Following is the theoretical background for my thesis, where I have chosen to include theory regards to speaking in chapter 2, assessment in chapter 3 and finally, classroom size in chapter 4.

2 Speaking

2.1 Introduction

Language ability has traditionally been considered, by language teachers and language testers alike, to consist of four skills: listening, reading, speaking, and writing (Bachman, Palmer: 2010, p. 55). My main focus is oral proficiency; therefore I will provide some theory about the spoken language in this chapter. In order to create context, first I will give a brief historical overview of foreign language teaching. However, it is important for me to mention that speaking is probably the most difficult language ability to assess, according to Luoma (2004). In chapter 3 I will present theory about language assessment.

2.2 Historical overview of foreign language teaching

Language has existed for as long as humans have, to begin with in very basic forms then gradually become complex systems, in constant development and change. Many attempts to explain the different segments of language have been made throughout the years, which is why I will give a brief presentation of the history of language in the following sections. The historical background is mainly based on literature from Simensen (2007) and Luoma (2004).

2.2.1 The Reform Movement

The International Phonetic Association (IPA) was founded in 1886, bringing three important aspects for the later development of linguistics and teaching theory:

1. a recognition of the importance of speech,
2. a scientific approach to the contemporary form of language, and

3. the development of an international phonetic script that could be used in the description of any language.

(Simensen, 2007, p. 25)

Among non-native second language learners, the pronunciation was poor. Several members of the Reform Movement were also members of the IPA. Having a practical tool to offer teachers; a phonetic alphabet and phonetically transcribed texts for the teaching of pronunciation, caused the Reform Movement to gain larger impact.

The reformers thought of speech as more important than written language, they believed that short stories could be used for meaningful oral activities and that teaching in the target language itself would establish direct associations between the things, actions and states talked about. (Simensen, 2007, p. 26)

The “direct method” which refers to the belief in establishing direct associations between L2 words and phrases, and the objects, actions and states referred to, can be seen as the practical manifestation of the Reform Movement. The grammar- translation method, with comprehensive study of the L2 grammar and extensive use of translation exercises, was still at that time the most important teaching method in Norway. However, it efficiency was already then being questioned. (Simensen, 2007, p. 31)

2.2.2 System and control

New principles of selection and grading of the content of L2 courses emerged from Palmer and West’s ideas about the importance of an even distribution of new words in teaching materials, together with repetition. In addition to this, Palmer had the idea that an even progression in teaching could be ensured by moving from the known to the unknown in small steps, and also moving from smaller structural units to larger, to eventually be able to construct all kinds of sentences.

Similar to the direct method, mentioned in section 2.3, ideas of learning from habit-formation gained interest. The idea of learning language by forming habits was a central idea towards the middle of the last century.

The oral method, which included the establishment of a system for the selection of what to teach, the principle of frequency of language items, grading and repetition aimed to develop skills in listening, speaking, reading and writing. The oral method is usually seen as an improvement of the direct method, which was the officially sanctioned method of teaching a second language towards the 1950's (Simensen, 2007, p. 34-39).

2.2.3 Objective science

In the period between 1950 and 1975, the American structuralism had a great influence on second language teaching. One aim of the American structuralism was to make linguistics into an empirical, objective and descriptive science, based on positivist approach. New models for the description of modern languages, appropriate for the analysis of patterns in basic sentences were made, in addition to procedures for comparing languages.

The audio-lingual teaching method was introduced, in addition to controlled exercises and error avoidance (Simensen, 2007, p. 40-57)

2.2.4 Language in context

Since the middle of the last century, there has been a growing interest in oral language, conversations in particular. Comparison of oral and written language has also gained interest, written and spoken language was seen in relation to the transactional function of language (to convey information) and interactional function of language (to establish and maintain social relations).

Dell Hymes, an American anthropologist and sociologist, was highly influential in introducing the communicative competence concept. He defined "communicative competence" as the knowledge of "when to speak, when not, and as to what to talk about with whom, when and where, in what manner" (Hymes 1972 in: Simensen 2007, p. 72).

Chomsky's critical thinking about language learning has made a big impact on current views on language learning. He was critical of behaviouristic stimulus-response and conditioning theory. His fundamental hypothesis was that human beings are born with a language ability, which he refers to as "the language ability device" (LAD), which he later called "universal grammar".

Stephen Krashen's theory on second language acquisition, developed during the 1980's and has been influential on second language teaching. The input hypothesis can be characterized in terms of four points:

1. It relates only to acquisition
2. It claims that the learner acquires new language only by being exposed to comprehensive input (i + 1) , the learner need to understand meaning before being able to take in new language.
3. It maintains that if there is enough comprehensible input, the acquisition will automatically happen.
4. It claims that speaking fluently cannot be taught directly.

(Simensen, 2007, p. 81-82)

Over the last decades, many frameworks of communicative competence have been developed, but the most influential seems to be Canale and Swains attempt from 1980. Their framework consists of four components (competences), which are as follows:

1. Grammatical competence
2. Sociolinguistic competence
3. Discourse competence
4. Strategic competence

(Simensen, 2007, p.105)

In chapter 3 I will discuss Canale and Swains model of communicative competence in further detail.

The Common European Framework of Reference (CEFR) is a document published by the Council of Europe with the purpose to provide “a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe”

One of the most innovative aspects of the framework is the scaled descriptions in terms of “can do” statements (Simensen, 2007, p. 111). I will give a more detailed presentation of the framework in chapter 3. Communicative skills are still to this day one of the most important competences in language teaching and language learning.

2.3 What is spoken language?

According to Luoma (2004) the spoken language can be described through four characteristics: *the sound of speech, spoken grammar, vocabulary, slips and errors* and *processing and reciprocity*. The sound of speech can refer to different features, “individual sounds, pitch, volume, pausing, speed, stress and intonation” (Luoma, 2004, p. 11). Furthermore, spoken grammar differs from written grammar in that it contains word chunks and idea units, more than whole sentences. About vocabulary, Luoma (2004) says that the more complicated vocabulary a person had, the better the language ability is. Further, Luoma (2004) claims that “ordinary” words do not describe limited language use more than advanced language use, because such words are important in spoken language discourse (Luoma 2004). Finally, Luoma (2004) claims that slips and errors in speech often are connected to inattention and not necessarily a characteristic in non-native speech only (Luoma, 2004, p. 19).

2.4 The difference between spoken and written language

Hasselgård et.al (2010) claims that informal speech differs from formal writing in a number of respects.

First and foremost, informal speech is produced in face-to face interaction, where a lot is communicated by body language. The speaker and the addressee(s) are present in the same situation and the language tends to be less explicit than in formal writing (Hasselgård et.al, 2010, p.428).

Informal speech is typically a joint activity, where two or more participants are in constant interplay. Utterances do not need to follow each other in a neat order, and they often overlap.

Furthermore, informal speech is typically a social activity, where the focus is less on the transmission of information than in writing. The term phatic communication has been proposed to characterize the social use of language. Examples are small-talk about the weather and the greeting *How do you do?* , which is not a genuine question asking for information (Hasselgård et.al, 2010, p. 428).

Informal speech is produced there and then, without the possibilities of editing which are normally found in writing. Listeners must decode, at the same time as they plan their own contributions, whereas readers have more time at their disposal and can focus on comprehension. Therefore, spoken discourse tends to be less complex (syntactically), and there is less variation in both structure and vocabulary. In addition, disfluencies, such as false starts, repetitions and silent and filled pauses, are common in speech (Hasselgård et.al, 2010, p. 428-429).

Hasselgård et.al (2010) adds that informal speech, like spoken discourse in general, is carried by the medium of sound. there is an obvious relationship between the graphic system and the sound system, the latter is far more flexible in the meanings it may express, particularly through the system of intonation (Hasselgård et. al., 2010, p.429).

2.5 Speaking skills in LK06 Knowledge Promotion

The Norwegian Directorate for Education and Training has developed a *Framework for Basic Skills* approved by the Ministry of Education and Research. In this framework, five basic skills are included; oral skills, reading, writing, digital skills and numeracy.

In the compulsory and secondary education reform of 2006 the already mentioned five skills were defined as basic to learning in school, work and social life.

(KD, 2013)

The English subject curriculum (KL06) is structured into four main subject areas: *Language learning, Oral communication, Written communication, Culture, society and literature.*

2.5.1 Competence aims in the English subject curriculum

The English subject curriculum (LK06) includes specific competence aims after following years: year 2, year 4, year 7, year 10, VG1 and VG2.
In addition to the competence aims, five basic skills are included in the curriculum. The competence aims are, as previously mentioned divided into four main categories: *language learning*, *oral communication*, *written communication* and *culture*, *society and literature*. Because only oral skills are relevant for this study, following is a brief overview of basic skills and competence aims for oral communication.

According to the English subject curriculum, oral skills in English means being able to "*Listen, speak and interact using the English language. It means evaluating and adapting ways of expression to the purpose of the conversation, the recipient and the situation. This further involves learning about social conventions and customs in English-speaking countries and in international contexts. The development of oral skills in English involves using oral language in gradually using more precise and nuanced language in conversation and in other kinds of oral communication. It also involves listening to, understanding and discussing topics and issues to acquire more specialized knowledge. This also involves being able to understand variations in spoken English from different parts of the world."*

KD (2013)

Following is a presentation of the competence aims in oral communication at VG1 and year 10.

2.5.2 Competence aims after year 10

Oral communication

The aims of the studies are to enable pupils to:

- choose and use different listening and speaking strategies that are suitable for the purpose
- understand and use a general vocabulary related to different topics
- demonstrate the ability to distinguish positively and negatively loaded expressions referring to individuals and groups
- understand the main content and details of different types of oral texts on different topics
- listen to and understand variations of English from different authentic situations
- express oneself fluently and coherently, suited to the purpose and situation
- express and justify own opinions about different topics
- introduce, maintain and terminate conversations on different topics by asking questions and following up on input
- use the central patterns for pronunciation, intonation, word inflection and different types of sentences in communication
- understand and use different numerical expressions and other kinds of data in communication

(KD 2013)

2.5.3 Competence aims after VG1

Oral communication

The aims of the studies are to enable pupils to:

- evaluate and use suitable listening and speaking strategies adapted for the purpose and the situation

- understand and use a wide general vocabulary and an academic vocabulary related to his/her own education programme
- understand the main content and details of different types of oral texts about general and academic topics related to one's education programme
- listen to and understand social and geographic variations of English from authentic situations
- express oneself fluently and coherently in a detailed and precise manner suited to the purpose and situation
- introduce, maintain and terminate conversations and discussions about general and academic topics related to one's education programme
- use patterns for pronunciation, intonation, word inflection and various types of sentences in communication
- interpret and use technical and mathematical information in communication

(KD, 2013)

The competence aims after year 10 are similar to the aims after VG, however, the aims after VG1 require a higher degree of understanding and use of the English language, and they are more specific.

2.6 Chapter summary

In this chapter, a brief overview of the history of foreign language teaching was presented, before a definition of "spoken language" was introduced. Communication is one of the most important parts of spoken language, which it has been since the 1960s. The competence aims after year 10 and VG1 also include more specific skills e.g. vocabulary and fluency. The chapter ends with a presentation of the competence aims regards to "oral communication" after VG1 and 10th grade.

3. Assessment

3.1 Introduction

First of all it is important to answer the question: what is the purpose of assessment and testing? According to Simensen (2007), there is not only one answer to that question. The purpose is "to give information to future schools and employees about the level of proficiency of the test taker", "to give feedback to students about how much they have learned from a course taken or syllabus studied", "to give information to students and teachers about problem areas for the purpose of remedial teaching", to give information to teachers and school authorities about the results obtained in relation to the resources spent", to select students for courses with limited enrolment" and "to provide data for the purpose of research on teaching and learning" (Simensen, 2007, p. 252). There are different purposes for language assessment and testing, and different test types are suitable for each.

In this chapter I will present different aspects of assessment, starting with a short overview of the history of language testing, before I move further to explaining the different types of assessment. Thereafter, I will discuss assessment towards the course grade in 10^{th} grade and VG1 English, based on the English Subject Curriculum.

3.2 A brief history of assessment

Brown (1996) claims that the history of language testing can be divided into four historical movements, which are the following:

1. the *prescientific movement*
2. the psy*chometric- structural movement*

3. the *integrative- sociolinguist movement*, and
4. the *communicative movement*

(Brown, 1996, p. 23-24)

The prescientific movement is known for the use of grammar and translation in language teaching. The fairness of that type of language testing can be questioned, because it is based on subjective, not objective scoring. This of course, would not be accepted in the Norwegian school system today, where it is highly important to use assessment criteria that give a picture as detailed as possible of the students' achievement on tests.

With the psychometric – structuralist movement came the need for more reliable tests, which often were formed as multiple- choice tests, making them easier to assess. This movement was under influence by the behavioural psychology at the time (Brown, 2007, p. 24).

The integrative-sociolinguistic movement is known for introducing the need to test more than the correct use of grammar for instance, and that the language must be put in context instead of emphasizing small fractions of the language. (Brown, 2007, p. 24)

The communicative movement introduced the need for more authentic testing of language, where the ability of proper language use is more important than the focus on grammatical control (Brown, 1996, p.24)

3.3 Types of assessment

The term *"assessment",* is a multifaceted term, that for many years "was used to primarily describe processes of evaluating the effectiveness of sequences of instructional activities when the sequence was completed" (William, 2011, p. 3).

Today, however, numerous distinctions concerning assessment exist. Different researchers have come up with different theories about assessment and here I will present some of them. Throughout the next sections I will go through some of the distinctions and concepts in assessment, before I look into what the English Subject curriculum says about assessment. First off, I will begin with arguably the most basic distinction – formal and informal assessment.

3.4 Formal and informal assessment

According to Simensen (2007), we can divide assessment into two categories; formal and informal assessment.

Formal assessment is the type of assessment teachers give in tests and exams.

Informal assessment occurs on a day to day basis, in dialogues between teacher and student or between students, and activities such as question –answer sequences (Simensen, 2007, p.252).

.

3.5. Achievement and proficiency testing

Another separation within the term assessment is between objective and subjective assessment (Simensen, 2007, p.253).

Simensen (2007) divides testing into *proficiency* testing and *achievement* testing. "*Proficiency testing"* aims at giving information about a student's general level of proficiency in a second language, L2 (Simensen, 2007, p.252). While it normally is not related to any specific syllabus studies or course completed in the past, it points to some sort of gatekeeping function, for example the TOEFL, which is a proficiency test taken my many Norwegian students who need to document proficiency sufficient for studies in the US (Simensen, 2007, p. 252).

The purpose of achievement testing, on the other hand, is to give information about the achievement of a student in relation to a course taken, such as English in the high school system. Achievement testing is meant to answer the question: To what extent has the student learned what has been taught? When this type of testing is used at various points during a course it is called *formative testing,* whereas when it used at the end of a course, it is called *summative testing* (Simensen, 2007, p. 252). In the next section I will further explain the terms “summative and formative”

3. 6 Formative and summative assessment

3.6.1 Summative assessment

Simensen (2007) describes summative testing this way: “When achievement testing is used at the end of a course, it is called summative testing. The purpose then is to provide information about the final outcome of the course”. (Simensen, 2007, p. 252). A form of summative testing is the type of testing the student’s go through during exams, where the proficiency of what a student has learned by the end of the course, is being measured. Summative assessment then applies to examination marks and marks awarded for classwork in each subject, so-called overall achievement grades, provided at the end of lower secondary school, 1oth grade and in upper secondary school. This is referred to as final assessment (Udir, 2014)

In Norway we use what Simensen (2007) refers to as criterion-references testing. Criterion-references testing mean measuring an examinee’s performance against certain criteria. Criterion- referenced testing thus means determining to what degree the students have reached the objectives of course or satisfied the criteria set (Simensen, 2007, p. 252). I will go in further detail on criterion-referenced testing in section 3.7.

According to Slemmen (2009) formative assessment can be conducted through observing the students in different learning activities, e.g in dialogs with the students, or through oral and written tasks or products.

According to Wiliam (2011) assessment of learning often equals summative assessment, as assessment for learning equals formative assessment. He continues to argue that Bennett (2009) suggests that “assessments designed primarily to serve a summative function ,may also function formatively, while those designed primarily to serve a formative function may also function summatively” (Wiliam, 2011, p. 10).

Black & William argue that“ from the information that teachers gathers for formative purposes, they should, with selection and re-interpretation, be in a strong position to contribute to a fair summative report on each pupil” (Black and Wiliam, 1998, p.12)

3.6.2 Formative assessment

Formative assessment can be defined as “*practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited* (Black & Wiliam, 2009, p. 9). Formative assessment serves primarily as a tool for progress. Teachers assess their students in many different ways, with many different criteria.

Formative assessment differs from summative assessment in that it is an on-going assessment process. It can take place at any time and it is up to the teacher to decide where, when or how an activity can be used as a formative assessment. For instance, the teacher can choose to view all activity in the class as formative assessment.

3.7 Norm-referenced and criterion- referenced assessment

Brown (1996) separates language testing into two categories – norm-referenced and criterion-referenced tests. The biggest distinctions according to Brown (1996) is that norm-referenced testing is designed to measure global language abilities, where he lists overall English language proficiency, academic listening ability and reading comprehension as examples. Criterion-referenced tests, on the other hand, measure well-defined and specific objectives – objectives specific to a particular course, program, school district or state (Brown 1996, p. 2). Norm-referenced tests will produce scores which fall into a normal distribution, while criterion-referenced tests enable all students to score 100% on an examination – if they have learned all the course material (Brown ,1996, p. 5).

Brown (1996) also states that he does not view any of the test categories as better than the other – they are both equally important, however they acquire the teachers to decide what the aim for the test is, and which type is most suitable.

Below is a table showing the differences in norm – and criterion referenced tests, adapted from Brown (Brown 1996, p. 5)

Table 1: Differences Between Norm-Referenced and Criterion-Referenced Tests

Characteristic	Norm-Referenced	Criterion-Referenced
Type of Interpretation	Relative (A student's performance is compared to that of all other students in percentile terms.)	Absolute (A student's performance is compared only to the amount, or percentage, of material learned.)
Type of Measurement	To measure general language abilities or proficiencies	To measure specific objectives – based language points
Purpose of Testing	Spread students out along a continuum of general abilities	Assess the amount of material known, or learned by each student
Distribution of Scores	Normal distribution of scores around a mean	Varies, usually nonnormal (students who know all the

		material should all score 100%)
Test Structure	A few relatively long subtests with a variety of question contents	A series of short, well-defined subtests with similar question contemts.
Knowledge of Questions	Students have little or no idea of what content to expect in questions	Students know exactly what content to expect in test questions

The table consists of six different characteristics, which are listed in the first column. The table shows that norm-referenced and criterion-referenced test differ in:

- the ways that scores are interpreted,
- the kinds of thing that they are used to measure,
- the purposes for testing,
- the ways that scores are distributed,
- the structures of the tests and finally,
- the student’s knowledge of test question content.

(Brown, 1996, p. 2-3)

Brown (1996) also provides a table matching tests to decision purposes. The table I have included below:

Table 2: Matching Tests to Decision Purposes

Type of Decision				
Norm-referenced			Criterion-referenced	
Test qualities	Proficiency	Placement	Achievement	Diagnostic
Detail of information	Very General	General	Specific	Very Specific
Focus	Usually, general skills prerequisite to entry	Learning points all levels and skills of programs	Terminal objectives of course or program	Terminal and enabling objectives of courses
Purpose of Decision	To compare individual overall with other groups/individuals	To find each student's appropriate level	To determine the degree of learning for advancement or graduation	To inform students and teachers of objectives needing more work
Relationship to program	Comparisons with other institutions	Comparisons within program	Directly related to objectives of program	Directly related to objectives still needing work
When administered	Before entry and sometimes at exit	Beginning of program	End of courses	Beginning and/or middle of courses
Interpretation of Scores	Spread of scores	Spread of scores	Number and amount of objectives	Number and amount of objectives

			learned	learned

As the table shows, Brown (1996) has made a distinction between four kinds of decisions: *proficiency*, *placement*, *achievement* and *diagnostics*. Further, he argues that those are the primary *language testing functions* (Brown, 1996, p. 8).

Brown (1996) claims that proficiency decisions are useful in cases where students’ general level of language proficiency is requested to entry or exit from some type of leaning institution. He also claims that entrance and exit standards for a curriculum are necessary, for adjustment of the level of program objectives to the student’s abilities. It is also useful for the purpose of comparison between programs.

Placement decisions on the other hand, have the goal of grouping together students of similar ability levels, in order to make it easier for teachers to focus on learning point for a specific level (Brown, 1996, p. 9). The two language testing functions described above, are decisions on program-level.

Achievement decisions are decisions about the amount of leaning the students have made. In order to make decisions about achievement and how to improve it, usually involves testing. Brown (1996) argues that achievement tests must be designed with very specific reference to a particular course, and therefore usually directly based on course objectives. Therefore, it will also be criterion-referenced (Brown, 1996, p. 14).

Diagnostic decisions are aimed at fostering achievement by promoting strengths and eliminating the weaknesses of individual students, although the primary concern should be on the entire group of students. That type of testing is criterion- referenced, and has the purpose to help students and teachers, to focus their efforts where they will be most effective (Brown, 1996, p. 15). The latter two decisions are at classroom-level.

3. 8 Canale and Swain's model of communicative competence

Canale and Swain (1980) produced the first and most influential model of 'communicative competence'. Canale and Swain argue that it is necessary to distinguish between "communicative competence" and "communicative performance" (Fulcher & Davidson 2007, 38). They present a model of knowledge, into which sociolinguistic competence is added; however they do not have a model of performance. Fulcher and Davidson (2007) present Canale and Swains model including two components:

1. Communicative competence (a model of knowledge):
   - grammatical competence: the knowledge of grammar, lexis, morphology, syntax, semantics and phonology
   - sociolinguistic knowledge: the knowledge of the sociocultural rules of language use and rules of discourse, and
   - strategic competence: the knowledge of how to overcome problems when faced with difficulties in communication.
2. Actual communication
   - the demonstration of knowledge in actual language performance.

(Fulcher & Davidson 2007, p. 38)

Canale and Swain outline Hymes's notion of a *speech event* in terms of participants, settings, form, topic, purpose, key, channel, code, norms of interaction, norms of interpretation and genre. The speech event is said to be the basis for understanding the rules of language use. (Fulcher & Davidson 2007, p.39). Although Canale and Swain previously had claimed that explicating a theory of performance was impossible, by 1983 Canale (1983a; 1983b) began to make a model (Fulcher & Davidson 2007, p. 39). Canale then ceased to use the term 'performance' and began to use 'actual communication' to mean:

“the realization of such knowledge and skill under limiting psychological and environmental conditions such as memory and perceptual constraints, fatigue, nervousness, distractions and interfering background noises”

With this definition, it is implied that both psychological and contextual variables must be accounted for, when modelling communicative competence and performance (Fulcher & Davidson 2007, p. 40).

3. 9 Types of assessment according to the CEFR

According to the CEFR there are numerous distinctions that can be made in relation to assessment. Following is a table the CEFR used to describe some of the different assessment types.

Table 3: Assessment types in CEFR

1	Achievement assessment	Proficiency assessment
2	Norm-referencing (NR)	Criterion referencing (CR)
3	Mastery learning CR	Continuum CR
4	Formative assessment	Summative assessment
5	Direct assessment	Indirect assessment
6	Performance assessment	Knowledge assessment
7	Subjective assessment	Objective assessment
8	Impression	Guided judgement
9	Checklist rating	Performance rating
10	Holistic assessment	Analytic assessment
11	Continuous assessment	Fixed assessment points
12	Series assessment	Category assessment
13	Assessment by others	Self-assessment

Achievement assessment is the assessment of the achievement of specific objectives- assessment of what has been taught. *Direct assessment* is assessing what the student is actually doing. For example, a small group is discussing something, the assessor observes, compares with criteria grid, matches the performances to the most appropriate categories on the grid, and gives an assessment. In comparison, *indirect assessment* uses a test, usually on paper, which often assesses enabling skills. Direct assessment is effectively limited to speaking, writing and listening in interaction, since it is not possible to see receptive activity directly. *Performance assessment* requires the learner to provide a sample of language in speech, whereas *knowledge assessment* requires the learner to answer questions which can of a range of different item types in order to provide evidence of the extent of their linguistic knowledge and control. *Norm-referencing* is based on placing learners in rank order, their assessment and ranking in relation to other peers. *Criterion-referencing,* on the other hand, is a reaction against norm-referencing in which the learner is assessed purely in terms of his/her ability in the subject, irrespective of the ability of his/her peers.

Self –assessment is a form of assessment where the students actively can participate in the evaluation of their own competence. The main potential for self-assessment is in its use as a tool for motivation and awareness raising: helping students to appreciate their strengths, recognize their weaknesses and orient their learning more effectively. In addition, one of the competence aims in the English subject curricula is to "on own work in learning English" after year 10.

3.9.1 Communicative competence in the CEFR

The CEFR divide "communicative competence" into three components; *linguistic competences, sociolinguistic competences* and *pragmatic competences*.

Linguistic competence is classified by following components; lexical competence, grammatical competence, semantic competence, phonological competence, orthographic competence and orthoepic competence. (Council of Europe, 2001, p. 109)

Lexical competence is “the knowledge of, and the ability to use, the vocabulary of a language, consists of lexical elements and grammatical elements” (Council of Europe, 2001, p.110). Lexical elements include fixed expressions such as idioms and fixed collocations, and single word forms such as auxiliary verbs and conjunctions. The CEFR has included a scale with descriptions of vocabulary range and vocabulary control into six levels –C2, C1, B2, B1, A2 and A1. (Council of Europe, 2001, p.112).

Grammatical competence is defined as *“knowledge of, and ability to use, the grammatical resources of a language”* (Council of Europe, 2001, p.112). The description of grammatical organization involves the specification of; elements (e.g. morphs, morphemes and words), categories (e.g. progressive, past/present/future tense), classes (e.g. conjugations, declensions), structures (e.g. compound and complex words, phrases, clauses), processes (e.g. nominalization, affixation, suppletion) and relations (e.g. government, concord, valency). According to the CEFR, *“grammatical competence is the ability to understand and express meaning by producing and recognising well-formed phrases and sentences”*. (Council of Europe, 2001, p. 112-113)

A further distinction in grammatical competence is drawn between morphology and syntax. “*Morphology deals with the internal organisation of words, whereas syntax deals with the organisation of words into sentences in terms of the categories, elements, classes, structures, processes and relations involved, often presented in the form of a set of rules*. “(Council of Europe, 2001, p. 114-115).

Semantic competence *deals with the learner’s awareness and control of the organisation of meaning* (lexical and grammatical and pragmatic). (Council of Europe, 2001, p. 116)

Phonological competence involves *a knowledge of, and skill in the perception, and production of the sound-units (phonemes) of the language and their realisation in particular contexts (allophnes).* (Council of Europe, 2001, p. 116)

Orthographic competence involves a knowledge of and skill in the perception and production of the symbols of which written texts are composed. For alphabetic systems (which all European languages are based on), learners should know and be able perceive and produce; e.g. proper spelling of words and punctuation marks.

Orthoepic competence covers knowledge of spelling conventions, ability to use a dictionary, in addition to knowledge of the implications of written forms and to resolve ambiguity when using in speech (or reading aloud) words first encountered wit in their written forms (Council of Europe, 2001, p.117-118).

Sociolinguistic competence is concerned with the knowledge and skills required to deal with the social dimension of language use - linguistic markers of social relations; politeness conventions; expressions of folk-wisdom; register differences; and dialect and accent.

Politeness conventions vary from one culture to another and are a frequent source of inter-ethnic misunderstanding, often literally interpreted. Examples of positive politeness, negative politeness and impoliteness are given, in addition to the appropriate use of “please”, “thank you” and such (Council of Europe, 2001, p-118-119).

Linguistic markers of social relations include use and choice of greetings, such as “Hello!” and “Goof morning!”, use and choice of address form such as “Sir” (formal) and “John!” (informal), conventions for turntaking and use and choice of expletives (e.g. Dear, dear!, My God!) (Council of Europe, 2001, p.119).

Expressions of folk wisdom refers to fixed formulae, which both incorporate and reinforce common attitudes and make a significant contribution to popular culture. They are frequently used in newspaper headlines and “assumed to be known to all”. These expressions are significant components of the linguistic aspect of sociocultural competence. Proverbs, idioms, familiar quotations, expressions of belief, values and attitudes are examples of such expressions. (Council of Europe, 2001, p. 119-120).

Register difference refers to systematic differences between varieties of language used in different contexts expressed by differences in level of formality, e.g. “Shall we begin?” which shows a neutral level of formality, whereas “May we now come to order, please” shows a formal level of formality. The level of formality is connected to how well we know the person we are addressing (Council of Europe, 2001, p. 120).

Sociolinguistic competence also includes the ability to recognize the linguistic markers of, for example: social class, regional provenance, national origin, ethnicity and occupational group. Such markers of dialect and accent include choices of lexicon, grammar, phonology, vocal characteristics, paralinguistics in addition to body language. The CEFR has also included a scale of *sociolinguistic appropriateness* divided into six levels of competence. (Council of Europe, 2001: 121-122).

Pragmatic competences are concerned with the user/learner's knowledge of the principles according to which messages are: a) organised, structured and arranged ('discourse competence'); b) used to perform communicative functions ('functional competence'); c)sequences according to interactional and transactional schemata ('design competence').

Discourse competence involves the ability of a user/learner to arrange sentences in sequence so as to produce coherent stretches of language. It includes the ability to control the ordering of sentences in terms of topic/focus, given/new, 'natural' sequencing and cause/effect. (Council of Europe, 2001, p. 123). It also includes knowledge about coherence and cohesion, flexibility to circumstances, thematic development and turntaking. Illustrative scales for those are included.

Functional competence is concerned with the use of spoken discourse and written texts in communication for particular functional purposes. A distinction is made between what is referred to as 'microfunctions' and 'macrofunctions'. "Microfunctions are categories for the functional use og single (usually short) utterances, usually as turns in an interaction" (Council of Europe, 2001, p. 125). "Macrofunctions are categories for the functional use of spoken discourse or written text consisting of a sequence of sentences" (Council of Europe, 2001, p. 126).

"Functional competence also includes knowledge of and ability to use the schemata – patterns of social interaction – which underlie communication, such as verbal patterns". (Council of Europe, 2001, p. 126-127). Examples of the interactive communicative activities, formed in pairs are: question-answer, statement-agreement/disagreement, request/offer/apology-acceptance/non-acceptance, greeting- response. (Council of Europe, 2001, p. 127)

3. 10 Assessment in the Knowledge Promotion LK06

According to the Norwegian curriculum, the foundation for assessment should be the competence aims in each subject. The Norwegian primary education has goal-oriented evaluation principles. This means that additional factors as good order and behavior, effort and conditions should not make an impact on the assessment. The competence aims in the curriculum are designed to easily be customized to meet the students' conditions. All students should be able to reach the goals, with different achievement levels.

In the National Curriculum (LK06) there are competence aims for each subject. Summative assessment is then used to describe the students' competence based on these competence aims.

The competence aims are not set for each year. In upper secondary school, the competence aims are stated after each year, however in lower secondary, the competence aims are only stated after 10th grade, and in primary school after the second, fourth and seventh grade.

Assessment in primary and secondary school in Norway is regulated by the Education Act (2009) "The Education Act makes a distinction between assessment *for* learning and assessment *of* learning, where assessment for learning shows what a learner can do, while assessment of learning measures what a learner knows" (Brevik, 2012, p. 133)

3.11 High stakes and low stakes

According to Bachman and Palmer (1997), high-stakes decisions are those that are likely to have a major impact on the lives of large numbers of individuals, or on large programs. Examples of high-stakes decisions about individuals include decisions about admission to academic programs, the awarding of scholarships, and the employment and retention of teachers. In addition to this, high-stakes decisions are not easily reversed, which means that decision errors cannot easily be corrected (Bachman & Palmer, 1997, p. 97).

Low-stakes decisions are decisions which have relatively minor impact on the lives of relatively small numbers of individuals or on small programs, and unlike high-stake they can be reversed quite effortlessly. An example is if a teacher is to use a test to diagnose students' strengths and weaknesses in order to assign them for certain learning activities. It the teacher misdiagnoses some of the students' areas of weakness and then assign them to inappropriate learning activities, relatively few will be affected, and also and also the errors can be easily corrected (Bachman & Palmer, 1997, p. 97).

3.12 Assessment and motivation

Assessment may be directly linked to motivation: "Motivation can be viewed as a condition that causes activity with the individual, directs the activity in a definite direction, and keeps the activity going" (Manger, 2010, p. 280). Manger (2010) further explains that the division between inner and outer motivation is not absolute, and that there can both be a desire for good grades *and* an interest in the topic and/or subject.

Feedback, feed-forward and self-assessment are all examples of formative assessment. When students are given feedback they feel an obligation to their learning, which can motivate them to continue with their work.

3.13 Reliability and validity

According to Bachman & Palmer (1996) reliability is often defined as consistency of measurement. A reliable test score will be consistent across different characteristics of the testing situation. Thus, reliability can be considered to be a function of the consistency of scores from one set of tests and test tasks to another and if the think of test tasks as sets of test characteristics, reliability can then be considered to be a function of consistencies across different sets of test task characteristics (Bachman and Palmer, 1996, p. 19-20). If some raters rate more severely than others, the ratings of different raters are not consistent, and the scores obtained could not be considered to be reliable. Reliability is therefore an essential quality of test scores, and unless test scores are relatively consistent, they cannot provide us with any information at all about the anility we want to measure.

Bachman and Palmer (2010) present two terms related to assessment: *systematicity* and *substantive grounding*. When they refer to assessment that is systematic, they talk about "designed and carried out according to clearly defined procedures that are methodical and open to scrutiny by other test developers" (Bechman & Palmer, 2010, p. 20). A substantively grounded assessment means that it is "based on a recognized and verifiable are of content, such as a course syllabus" (Bechman & Palmer, 2010, p.20). If a test is to be viewed as valid order, two questions have to be answered: what is being assessed and how is it being assessed? Both questions must be included for a test to be valid.

The term construct validity is used to refer to the extent to which we can interpret a given test score as an indicator of the ability (ies) or construct(s) we want to measure. Construct validity pertains to the meaningfulness and appropriateness of the interpretations that we make on the basis of the test scores When we interpret test scores from language tests as indicators of test takers' language ability, a crucial question is "To what extent can we justify these interpretations?" (Bachman & Palmer, 1997, p.21).

Luoma (2004) presents three frameworks specifically relevant for speaking constructs:

1.) the linguistically oriented framework
2.) the communication oriented framework, and
3.) the situation-based framework

The first, focus on vocabulary, pronunciation, grammar etc. which are linguistic characteristics. The second framework focus on communication through tasks with e.g. narrating and explaining, and the third framework includes tasks of communicating in specific areas (Luoma, 2004, p.162-163).

3.14 Chapter summary

The purpose of this chapter was to look into the different aspects of assessment, specifically assessment of oral proficiency. The theory provided points out that there are many considerations to be taken in order to ensure the reliability and validity of assessment and testing. A possibility is to create national assessment criteria with detailed information on how to assess oral proficiency. The Directorate for Education and Training has developed national guidelines for three achievement levels: grade 2 (a low level of achievement), grade 3 and 4 (medium level of achievement) and grade 5 and 6 (high level of achievement, both for written and oral English (KD, 2013). However, the descriptions lack detail and are very general.

4 Classroom size

4.1 Introduction

“Parents and teachers often have strong views about the effects of class size. Probably all of us would take that view that – other things being equal- children are more likely to receive a better quality of education in small classes” (Blatchford and Mortimore, 1994, p. 406). Because this is a view shared by many teachers, policy makers (and parents) there has been conducted many studies upon this particular issue. One of the questions in my questionnaire focuses on class size in relation to oral activities

The research available does not focus on class size specifically linked to oral activity; however there are some interesting examples and indications regards to this.

4.2 The impact of the class size

The research on the effects of class size has been among the more voluminous in educational research with very systematic findings, according to Hattie (1999). Research carried out by Hattie has contributed to the following results:

Achievement, attitude, teacher morale, student satisfaction gains are appreciable in smaller classes, so long as we recognize that small classes mean 10-15, as there are negligible gains between 40 to 20 students per class.

This effect was the same for primary and secondary schools, across all subjects, and across various ability levels.

There is little evidence that instruction methods change when class size is reduced, although a large part of improvement can be explained by improvements in student task engagement. Reducing class sizes from the 30’s to the 20’s is in the right direction, but there is little support for the claim that there are increases in student achievement or satisfaction, or teacher attitude or morale. Only when the class size reduces to 15 or below are there appreciable positive benefits. (Hattie, 1999, p. 9)

Hattie has also concluded that *"Teaching practices that are conducive to successful learning are more likely to occur in smaller rather than larger classes, and these practices do not actually occur more in smaller classes because teachers have been prepared to, and indeed do, work with larger classes using more transmission practices and therefore they are not so equipped to adopt the more effective practices when they are given smaller classes."* (Hattie, 2005 In: Zyngier, 2014)

4.3 STAR project

The Tennessee class size project was a three-phase study designed to determine the effect of smaller class size in the earliest grades on short-term and long-term pupil performance. The study was concluded with compelling evidence that smaller class size helps, at least in early and that the benefits derived from these smaller classes persist leaves open the possibility that additional or different educational devices could lead to still further gains. For example, applying to small classes the technique of within-class grouping in which the teacher handles each small group separately for short periods could strengthen the educational process (essentially a second-order use of small class size). The study also concluded that small classes used jointly with other teaching techniques may add further gains (Mosteller, 1995)

4.3 SAGE project

The Sage (Wisconsin's Student Achievement Guarantee in Education) project enacted by the Wisconsin state law in 1995, began as a five- year pilot program in the school year of 1996-97 to test the hypothesis that smaller classes in elementary schools raise the academic achievement of disadvantaged students. The SAGE project included four reform initiatives: (1) reduction of the pupil-teacher ratio in classrooms to 15:1; (2) establishment of "lighted schoolhouses" that are open longer than the traditional school day; (3) development of rigorous curricula; and (4) refinement of staff development and professional accountability systems to support the class size reduction program. SAGE and comparison school students began first grade with similar reading, language arts and math scores on pre-tests, but by the second and third grades, SAGE students outscored their peers in comparison schools on every test administered by the evaluators. The gap was statistically significant in every subject except reading. The SAGE initiative reduced the gap between white and African American student achievement, with the strongest effect observed during the first grade year. By contrast, the achievement gap increased over time in comparison schools. The differences in achievement outcomes, related to the type of classroom reduction strategy used were not statistically significant. In other words, regular small classes, team-teacher classes, shared space and floating teacher classes had similar, positive benefits for student achievement. According to evaluators, the most significant factor affecting individual student performance on tests was socioeconomic status, but when this variable was accounted for, class size reduction strategies used by SAGE had similar, positive effects (Zyngier, 2014).

4.5 Chapter summary

It is questionable how relevant the two studies mentioned above are for my thesis because both studies concern early stages of children's education. However, if that factor is ignored, both studies conclude with a positive impact of smaller classes. In addition to class size, teaching methods are essential to students' achievement overall. There are many additional factors beside the size of the class when it comes to teaching, and therefore it can be difficult, if not impossible, to know the actual effects of smaller classes.

5 Methods

5.1 Introduction

According to Cohen et.al (2011) there is no single blueprint for planning research. Research design is governed by the notion of 'fitness for purpose' the purposes of the research determine the methodology and design of the research (Cohen et.al, 2011, p. 115). It is however, possible to plan the different stages of the research. Cohen et.al. (2011, p.140) includes five stages when describing the research process:

1. Development and operationalization, including:
    a. a review of literature and commercially produced instruments;
    b. clarification of the research questions;
    c. clarification of methodology and sampling.
2. Instrumentation and the piloting of the instruments

3. Data collection

4. Data analysis and interpretation

5. Reporting (A full report on the findings will include conclusions, implications and recommendations.)

In this chapter I will give a description of how I conducted my research – which choices and decisions I had to make throughout the project. I will to some extent follow the process described by Cohen et.al. with a clarification of research question and methodological choices of my research in section 5.2, research design in section 5.3, data collection in section 5.4,

5.2 Research question and methodological choices

5.2.1 The research question

Any research has one goal – to search for the truth. As already mentioned in chapter 1, I wanted to write a thesis about something I found interesting, challenging, but also useful for my future work as a teacher – which I believe is a very important and sadly not enough appreciated profession. Formulating a good research statement was challenging and I had to adjust it a few times, before I ended up with my final thesis.

The research question is as follows:

The course grade in VG1 and 10th grade English:

How do teachers assess oral proficiency towards the final course grade?

To help me answer the research question I developed a couple of subsidiary questions:

- How often are teachers assessing students' oral proficiency compared to other types of assessment e.g. written tests?
- Which criteria are used when assessing students' oral proficiency and who makes the criteria?
- Does classroom size play any role when assessing the student's oral proficiency?

The third sub- question I included because it is relevant to one of the questions in my questionnaire. I assumed some of the teachers would point to class size having an impact on the amount of available time and resources to assess the students.

5.2.2. Methodological choices

Normally we have to choose between two methods when conducting a research – qualitative or quantitative.

The method we choose depends on what we want our research to show. Quantitative methods are number-based, objective and generalizable, whereas qualitative methods often are text-based, subjective and less generalizable, however both can be valid and reliable.

A qualitative approach usually consists of in-depth interviews or observations. This approach is suitable if you want to form new hypothesis or theory. A quantitative approach, on the other hand, is suitable for testing hypothesis and constructs already existing.

The number of respondents varies between the two approaches. For a qualitative approach, only a few respondents can be sufficient enough, however choosing a quantitative approach, a larger number of respondents is necessary (Cohen et.al, 2011).

There is however, a third option – a mixed methods approach, which I will explain in the next section.

5.2.3 Mixed methods research

Explained by Creswell & Plano Clark (2011):

> *"Mixed methods research is a research with philosophical assumptions as well as methods of inquiry. As a methodology, it involves philosophical assumptions that guide the direction of the collection and analysis and the mixture of qualitative and quantitative approaches in many phases of the research process. As a method, it focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches, in combination, provides a better understanding of research problems than either approach alone."* (Creswell & Plano Clark, 2011, p.5)

The beginnings of mixed methods are dated back to the late 1980s with the coming together of several publications focused on describing and defining what is now known as mixed methods. In 1959, Campbell and Fiske had discussed the possibility of including sources with quantitative information in validation of psychological traits. Others also advocated the use of multiple data sources – both quantitative and qualitative – to conduct scholarly studies (Denzin, 1978), and several well-known figures in quantitative research, such as Campbell (1974) and Cronbach (1975), advocated for the inclusion of qualitative data in quantitative experimental studies (Creswell & Plano, 2011, p. 21).

According to Creswell & Plano (2011, p.21) a number of factors have contributed to evolution of mixed methods research. They explain this by pointing to the complexity of research problems and the need for answers beyond simple numbers in a quantitative sense or words in a qualitative sense. A combination of both forms of data provides the most complete analysis of problems.

I was not certain of which method was the most suitable when I started the process of making this thesis. My biggest concern was which method of data collecting I would choose. Talking to other master students I knew that it would not be easy to find the sufficient amount of respondents if I wanted to interview or observe teachers or students, simply because it is too time consuming. From my own experience of teaching 10th grade English, I monthly received e-mails from students who wanted me to answer questionnaires, where I would refuse to answer those who took more than 10 minutes to finish. The best option for me was obvious – en embedded research design, which I will explain in the next section.

5.3 Research design

Reading through theory about different mixed methods approaches, I figured that "the embedded research design" was the most suitable for my research.

The embedded design is a mixed methods approach where the researcher combines the collection and analysis of both quantitative and qualitative data within a traditional quantitative research design or qualitative research design (Caracelli & Greene 1997; Greene, 2007). The collection and analysis of the second data may occur before, during, and/or after the implementation of the data collection and analysis traditionally associated with the larger design. (Creswell & Creswell, 2011, p. 90)

I wanted to make a questionnaire my respondents would easily access, with both open-ended questions and questions with multiple choices. I also wanted to include interview with open-ended questions where I could get a more in-depth understanding of my respondents' thoughts on my questions, and ensure that they understood my questions the way I had intended them. By doing so, I will end up with results presented both in numbers (quantitative) and words. I will use the answers from the interviews to provide a deeper understanding of the answers from my questionnaire. Then I will present them together. I will go in further detail my data collection in the next chapter.

5.4 Data collection

The basic idea of collecting data in any research study is to gather information to address the questions being asked in the study. In mixed methods research, the data collection procedure consists of several key components: sampling, gaining permissions, collecting data, recording the data, and administering the data collection (Creswell & Plano, 2011, p. 171).

The basic distinction we make between qualitative and quantitative data is that qualitative data consists of information obtained on open-ended questions in which the researcher does not use predetermined categories or scales to collect the data. Here, the participants provide information based on questions that do not restrict the participants' options for responding. In contrast, quantitative data are collected on closed-ended questions based on predetermined response scales, or categories. E.g. a quantitative questionnaire illustrates how a researcher identifies questions and asks participants to rate their answers to the questions on the scale (Creswell & Plano, 2011: 176-177)

As mentioned in the previous section, I will combine qualitative and quantitative data to provide a better understanding of my research.

5.5 Sampling

The suitability of the sampling strategy that has been used is important for the quality of the research (Cohen et.al, 2011, p. 143). According to Cohen et.al (2011), judgments have to be made about five key factors in sampling.

1. the sample size;
2. the representativeness and parameters of the sample;
3. access to the sample;
4. the sampling strategy to be used;
5. the kind of research that is being undertaken (e.g. quantitative/qualitative/mixed methods).

Cohen et.al, 2011: 143)

There is no definite answer to how small or big a sample should be, different factors play a role, factors such as the purpose of the study, the nature of the population under scrutiny, the level of accuracy required, the anticipated response rate, the number of variables that are included in the research, and whether the research is qualitative or quantitative. A sample size of 30 is held by many to be the minimum of cases is researchers plan to use some sort of statistical analysis of their data (Cohen et.al, 2011, p.144). Cohen. et. al (2011) represents who types of sampling strategies: *probability samples* and *non-probability samples*. A probability sample is useful if the researcher wants to make generalizations, because it draws randomly from the wider population. Non-probability samples, on the other hand, are samples selective, and derive from the researcher targeting a particular group (Cohen et.al. 2011, p. 153-156). The sampling strategy used in my research is a non-probability sample, more precisely, "purposive sampling", where I have chosen informants with a specific profession, informants who teach English at VG1 and 10th grade.

5.6 Constructing the questionnaire

At the preliminary stage of design, many decisions had to be made. To begin with, I decided the purpose of the questionnaire, how many respondents I would need, before deciding what topics and questions I would include. This is consistent with how Cohen et.al. (2011: 378-

379) explains the approaching of the planning of a questionnaire. The questionnaire would then be used in a survey.

5.6.1 Structured, semi-structured and unstructured questionnaires

According to Cohen et.al (2011: 381) there is a large range of types of questionnaire, however there is a rule of thumb: the larger the size of the sample, the more structured, closed and numerical the questionnaire may have to be, and the smaller the size of the sample, the less structured, more open and word-based the questionnaire may be. Further, Cohen et.al (2011:382) argues that it perhaps is misleading to describe a questionnaire as being ‘unstructured’, as the whole devising of a questionnaire requires respondents to adhere to some form of given structure.

5.7 Types of questionnaire items

There are several kinds of question and response modes in questionnaires. The most common are: dichotomous questions; multiple choice questions; rating scales; constant sum questions; ratio data and open ended questions. Cohen et. al. (2011: 382-383) argue that closed questions prescribe the range of responses from which the respondent may choose, whereas open questions enable participants to write a free account in their own terms, to explain and qualify their responses and avoid the limitations of pre-set categories of response.

There are positive and negative sides to both open and closed questions. Wilson & McLean (1994:21) argue that closed questions are quick to complete and straightforward to code (e.g. for computer analysis), and do not discriminate unduly on the basis of how articulate respondents are. Oppenheim (1992, p. 115) on the other hand, argues that closed questions do not enable respondents to add any remarks, qualifications and explanations to the categories, and that there is a risk that the categories might not be exhaustive and that there might be bias in them.

According to Bailey (1994, p. 120), open-ended questions are useful if the possible answers are unknown or the questionnaire is exploratory, or if there are so many possible categories of response that a closed question would contain an extremely long list of options. On the other hand, open questions can lead to irrelevant and redundant information; they may be too open-ended for the respondent to know what kind of information is being sought, they may require much more time to respond, and thereby make the questionnaire appear discouraging (Cohen et.al, 2011, p. 382).

For my questionnaire I chose both closed, multiple choice, rank ordering questions and open-ended questions. Multiple choice questions are treated as nominal data (Cohen et.al, 2011, p. 384).

In order to make my research as representative as possible I had to find informants from different parts of Norway. Norway is divided into 19 counties and my goal was to find informants from a minimum of 8 counties. Because the size of the area and population of the different counties vary a lot, I found it unnecessary to search for informants from all the 19 counties. E.g. Oslo, which is one separate county, covers an area of approximately 454 square kilometres, whereas Finnmark, a county situated in the north part of Norway, covers an area of entire 48 618 square kilometres. Oslo however, has a much larger population than Finnmark (Berg, Ole T, 2015).

Because I intended to find information from teachers teaching at VG1 level and 10th grade, I also had to find an equal number of informants from both.

Recruiting informants was, not surprisingly, a difficult task. I wanted the informants to reflect on their own practice regarding the assessment of oral proficiency of their students, the assessment criteria they use, and the process of how they ended up with the final course grade. Having this in mind, I formed a questionnaire and an interview guide, with similar questions. The difference is that the questions for my questionnaire consisted of options, whereas the interview consisted of open questions.

Question for online questionnaire:

1. How many oral assessments do you give your students, compared to written assessments during one school year?

2. On this question you have to arrange your answers according to importance. At the top you must put the assessment form you view as the most important when deciding the students' oral proficiency.
   Which assessment forms do you use when assessing your students' oral proficiency in English?
3. How many oral assessments do you base the final course grade in VG1/10th grade on?
4. Which challenges do you face in the process of assessing the students' oral proficiency?
5. Having in mind what you answered in question 4. How do you solve the challenge/challenges?
6. Range the following criteria from 1-5, where 1 represents the criteria you view as the least important, and 5 represents the criteria you view as the most important.
   Which criteria do you use when considering the students' grade?
7. Who forms the assessment criteria?
8. How do you facilitate self-assessment?

In the following chapters I will go through the process of choosing and collecting informants, followed by a presentation of the informants.

5.8 The interview guide

Four main kinds of interview are discussed by Cohen et.al (2011): *the structured interview, the unstructured interview*, *the non-directive interview* and *the focused interview*. In structured interviews, the content and procedures are organized in advance. In contrast, the unstructured interview is an open situation, having greater flexibility and freedom. The non-directive interview includes the minimal direction or control exhibited by the interviewer and the freedom the respondent has to express her subjective feelings as fully and as spontaneously as she chooses or is able. (Cohen et.al, 2011, p. 415)

In the making of my interview guide I formed three thematic parts, the first including questions about "the assessment of oral proficiency" (Appendix 2 (my translation)), which was the part including the highest number of questions. The second part included questions about "assessment criteria" (Appendix 2 (my translation)) and last part included a question about “self-assessment” (Appendix 2 (my translation). In addition to this, the interview guide consisted of background questions and a request for participation in a follow-up phone interview, which was optional.

The interview was made up by open-ended questions only. According to Cohen et. Al (2011, p. 416), open-ended questions have a number of advantages: they are flexible; they allow the interviewer to probe so that she may go into more depth if she chooses, or to clear up any misunderstandings; they enable the interviewer to test the limits of the respondent’s knowledge; they encourage cooperation and help establish rapport; they allow the interviewer to make a truer assessment of what the respondent really believes.

Following are the questions I made for the interview (appendix 1):

1. How many oral assessments do you give your students, compared to written assessments during one school year?
2. How do you teach you students’ in oral proficiency? How often?
3. Which assessment forms/situations do you use when assessing the students’ oral proficiency?
4. How many oral assessments do you base the final course grade in VG1/10th grade on?
5. When you collect all the oral assessments you have on each student before you set the final course grade, how do you weigh the different assessment forms?
6. Which challenges do you face in the process of assessing the students’ oral proficiency?
7. Which criteria do you use when considering the students’ grade? Who forms the criteria? Do you use the same criteria over again?
8. If you are to make a prioritized list, what are the five most important criteria you look for when assessing oral proficiency?
9. How do you facilitate self-assessment?

In the following chapters I will go through the process of choosing and collecting informants, followed by a presentation of the informants.

5.9 The informants

The search for informants began after deciding which counties I wanted to be represented. I searched the county administrations' web pages in order to get contact information for elementary and secondary schools. The amount of contact information varied, most of the secondary schools' web pages provided great details of contact information, including which subjects and levels the teachers were teaching. Then I could contact the teachers directly.

The elementary schools' web pages on the other hand, did not contain the same amount of contact information. Here I had to call the school administration, explain what my study was about and ask them to give me further information. In most cases, even though the school administrators told me they would come back to me, they never did. The amount of e-mails, letters and phone calls I went through and sent, to end up with a total of 27 informants on my online questionnaire and 7 informants who wanted to participate in both an open-ended interview and a follow-up phone interview, was astonishing and not least, very time consuming. The result was a total of 34 informants, representing a total of 15 counties. I was not able to find informants from the four remaining counties. For practical reasons, the interview was send to the informants by e-mail. It would not have been possible to meet the teacher because they are spread around the entire country.

Presentation of teachers

Below is a table showing all my informants. Informant 1-27 answered my questionnaire. Informant number 27-34 answered my interview.

Table 4: Presentation of informants

Informant number	Age	Education	County of residence	Currently teaching at level:	Teaching experience (years)
1	58	Høyere utdanning/universitet 5 år	Østfold	VG1	35
2	51	Høyere utdanning/universitet 5 år	Aust-Agder	VG1	21
3	39	Høyere utdanning/universitet 5 år	Sør-Trøndelag	VG1	11
4	50	Høyere utdanning/universitet 5 år	Sør-Trøndelag	VG1	Ca. 20

5	34	Høyere utdanning/universitet 5 år	Nordland	VG1	9
6	50	Høyere utdanning/universitet 5 år	Aust-Agder	VG1	27
7	44	Høyere utdanning/universitet 5 år	Vestfold	VG1	18
8	46	Høyere utdanning/universitet 5 år	Sør-Trøndelag	VG1	13
9	64	Høyere utdanning/universitet 5 år	Sør-Trøndelag	VG1	39
10	44	Høyere utdanning/universitet 5 år	Nord-Trøndelag	10.trinn	19
11	50	Høyere utdanning/universitet 5 år	Aust-Agder	VG1	26
12	45	Høyere utdanning/universitet 5 år	Vestfold	10.trinn	18
13	31	Høyere utdanning/universitet 5 år	Møre og Romsdal	VG1	5
14	27 år	Høyere utdanning/universitet 5 år	Finnmark	10.trinn	Since Aug. 2013
15	50	Høyere utdanning/universitet 5 år	Vestfold	VG1	21
16	41	Høyere utdanning/universitet 5 år	Telemark	VG1	12
17	52	Høyere utdanning/universitet 5 år	Akershus	VG1	26
18	27	Allmennlærerutdanning og mastergrad i engelskdidaktikk	Akershus	10.trinn	This is my first year.
19	58	Almennlærerutdanning pluss 1.avd spes.ped og 1.avd sos.ped	Finnmark	10.trinn	23
20	42	Høyere utdanning/universitet 5 år	Telemark	VG1	10
21	28	Høyere utdanning/universitet 3 år	Telemark	VG1	1
22	49	Høyere utdanning/universitet 5 år	Rogaland	10.trinn	26
23	41	Høyere utdanning/universitet 5 år	Finnmark	10.trinn	17
24	37	Høyere utdanning/universitet 5 år	Buskerud	10.trinn	10

25	46	Høyere utdanning/universitet 5 år	Buskerud	10.trinn	21
26	45	Høyere utdanning/universitet 5 år	Telemark	VG1	18
27	63	Høyere utdanning/universitet 5 år	Østfold	10.trinn	38
28	46	Allmennlærerutdanning	Hordaland	10.trinn	
29	47	Høyere utdanning/universitet 5 år	Oslo	VG1	20
30	50	Høyere utdanning/universitet 5 år	Rogaland	10.trinn	28
31	51	Høyere utdanning/universitet 5 år	Buskerud	VG1	26
32	24	Høyere utdanning/høyskole 3 år + PPU ettårig	Østfold	10.trinn	1,5
33	35	Høyskole	Møre og Romsdal	10.trinn	9
34	41	Høyere utdanning/universitet 5 år + PhD 2,5 år	Hedmark	VG1	12

5.10 Can the results be trusted?

The following question will be answered in the next sections. First, I will begin with a definition of validity.

5.10.1 Validity

Validity is an important key to effective research. If a piece of research is invalid then it is worthless (Cohen et.al, 2011, p. 179). Onwuegbuzie and Johnson (2006): in Cohen et.al (2011, p. 198), argue that the term 'validity' can be replaced by 'legitimation' in mixed methods research. In order to assess the validity of a research, triangulation can be used. Triangulation may be defined as the use of two or more methods of data collection in the study of some aspect of human behavior. The use of multiple methods contrasts with the ubiquitous but generally more vulnerable single-method approach (Cohen et.al.,2011, p.195).

Because I used a mixed method approach, I could compare the outcome of the questionnaire survey, to those of the interviews. Cohen et.al (2011) argues that in its use of mixed methods, triangulation may utilize either normative or interpretive techniques; or it may draw on methods from both these approaches and use them in combination.

5.10.2 Transferability

Transferability is according to Johannesen, et.al, (2006, p. 200) concerned with whether or not the results can be transferred to other related phenomena.

Although I have gathered information from 34 teachers, it is impossible to generalize the findings to a large population. In order to make generalizations, it would be necessary to conduct a larger research on the topic.

To make generalizations was neither the aim of my research, however I believe the study has contributed to interesting findings that may indicate the need for improvement in how teachers practice assessment of oral proficiency.

5.10.3 Conformability

What is meant by the term ‘conformability’ is whether the results of a study reflect the results of the research, and not the personal opinions of the researcher (Johannesen, et.al., 2006, p. 201). It was impossible to anticipate what the result of the study would be; therefore I believe this study reflects my finding, and not my own opinions.

5.11 Chapter summary

Throughout this chapter, the methodological choices, used in my research, have been presented. A description of the mixed method approach was given, and the data collection, research tools and informants were presented. At the end of the chapter, I gave an overview of validity, transferability and conformability of the thesis.

6 Results and analysis

6.1 Introduction

In this chapter, the findings from my project will be presented. I have divided my finding thematically as follows:

1. Assessment
2. Assessment situations
3. Assessment criteria
4. Who makes the assessment criteria?
5. Challenges in assessing oral proficiency
6. How to solve the challenges
7. Self-evaluation

In this chapter I will only comment on the most important findings. In every chapter I will first comment on the results from the 27 informants who answered the survey questionnaire. At the end of each chapter I will present and comment on the answers from the 7 informants who both answered my questionnaire and went through with a follow-up phone interview.

6.2 Assessment

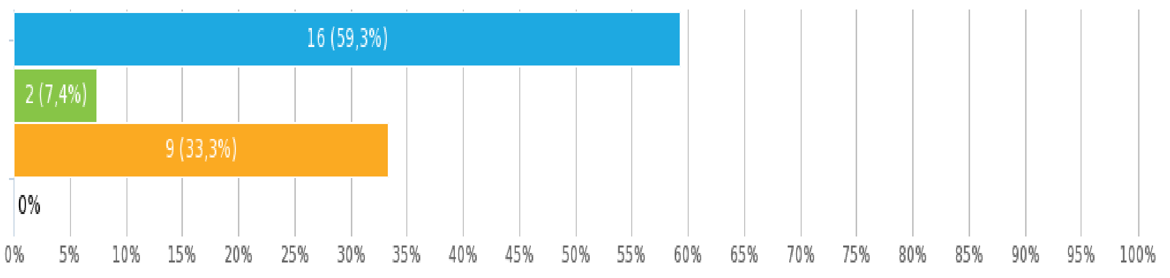
My informants were asked the following question on assessment: How many oral assessments do you give your students, compared to written assessments during one school year?

As previously mentioned, there is only one course grade for VG1 English which is supposed to describe the students' competence at the end of the course. For 10^{th} grade there are two separate grades in English – one for written, and one for oral competence. The purpose of this question was to establish how many oral assessments, compared to written assessments made up the final course grade. I asked both the informants who answered my questionnaire survey this question, as well as the informants who answered my interview. For my online questionnaire, I provided three options to this question; "I give my students the same amount of oral assessment as I do with written", "I give my students oral assessment more often than written" and "I give my students written assessment more often than oral."

No options were provided to the informants who answered my interview. For them – this was an open-ended question. In the table below, the answers from my survey are represented:

Table 5: Results from question 1

Answer	Number of Answers	Percentage
"I give my students the same amount of oral assessment as I do with written"	16	59,3 %
I give my students oral assessment more " often than written	2	7,4 %
"I give my students written assessment more often than oral."	9	33,3 %
"Other"	0	0 %



As the table shows, almost 2/3 of the informants (59, 3 per cent) answered that they give their students an equal amount of oral assessments as written. 33, 3 per cent answered that they more often give their students assessment on oral proficiency, while only 7, 4 per cent answered that they give their students more oral than written assessment.

The informants, who answered my interview, gave more detailed information on this question. Following are the answers from the interviews:

"We practice informal oral assessments weekly, or at least every other week. We have formal assessments approximately once every month, connected to projects the students are working on. Because I teach English at the vocational program, we have relatively few written assessments, about 3 per semester."

"When it comes to formal assessments I usually have two oral and three written per semester. That is the amount of assessments I need to be able to set a final course grade at the end of the school year."

"The students were given more oral activities during the first semester (fall), everything from listening tests to presentations. My plan is to go through an oral examination at the end of April. The tasks are built up systematically with an oral exam in mind."

"I went through with two formal oral presentations during fall, and I am planning to implement 1-2 formal presentations this spring. I conduct more written assessments than oral through one school year. However, I emphasize the importance of engaging in speaking activities in class every English lesson."

"The minimum requirement for oral assessment is 1 out of 4 each semester, or 2 out of 8 during one schools year. (This is the minimum I have set myself as a goal). I wish to become better at this, so that eventually I will be able to facilitate the same amount of oral assessments, as written. Being a recently graduated teacher, I have spent some time adjusting the assessment processes. So far I have conducted more written assessments because it gives me more time to assess and I have the possibility to look over the student's product several times."

"If I add informal oral presentations to the count, I give the same amount of oral assessments as I do with written. By informal oral presentations I mean presentations of homework which the students are not given time to prepare in class. For example, I give two-three students a task - to prepare a short news presentation each English lesson. The next lesson, two-three other students must do the same, until all the students in class have presented. This is something I do often, although I change the topic."

"I give my students more written than oral assessments because of the simple fact that oral assessments are much more time consuming and more difficult to arrange in a class with

many students. I do not have exact numbers, but I give my students somewhere between 4-5 written assessments every semester and no more than two oral assessments every semester."

In the Knowledge Promotion LK06, there is no information about the amount of assessment that is required to be able to set a course grade. One can however argue that it should be 50/50 in 10th grade because there are two separate grades. Also at VG1 level there should be a fair division between oral and written assessment because there is one combined grade for the two skills. It is a possibility that a student has a high achievement in written English and a low achievement in oral English. It a teacher was to put more emphasis on the oral proficiency, one can argue that the overall assessment is not fair for that particular student. To avoid situations as this, teacher should at least be aware of this issue and strive to make the assessment as fair as possible.

My findings show that there are different practices concerning the amount of written vs. oral assessments and that numerous other factors must be taken in consideration, e.g. the size of the class and the time available to conduct oral assessments compared to written assessments. Another factor is that teachers put different emphasis on what they view as assessment situations or not. Some teachers view all oral activity as suitable for assessment, while other teachers have a clear distinction between formal and informal activities. I will go further into assessment situations in following section. On the other hand, the results show that most of the teachers do in fact give the same amount of oral and written assessments and that those who do not have this division are conscious about the limitations when it comes to oral assessment.

6.3 Assessment situations

My informants were asked to arrange different assessment situations according to importance.

I formulated six different assessment situations, in addition to giving the informants the possibility to add other assessment situations by including an empty space to fill in, labeled "other".

The six options, in addition to the one I already mentioned, were: *"formal oral presentation", "group discussion with other students", "topic conversation", "class discussion", "group presentation"* and *"informal oral activities in class".*

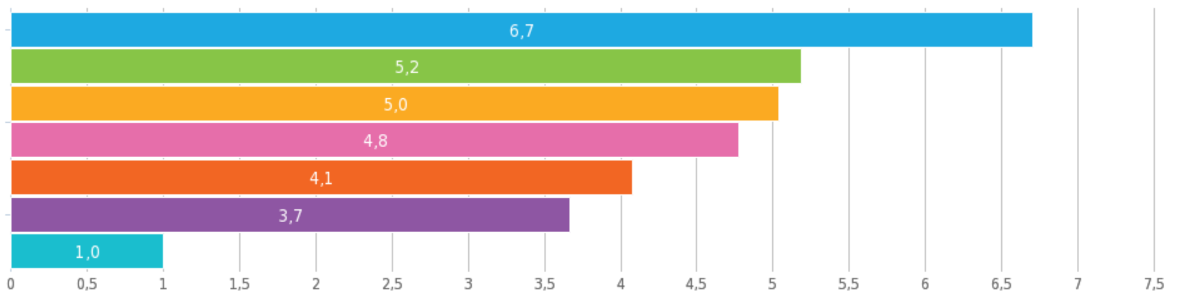
I asked my informants this question because I wanted to find out whether or not they view the same situations as suitable for assessment. I was not certain of whether informal everyday activities were important for all teachers – or how important they are. I also wanted to investigate how important formal presentations are, and if the final course grade mainly is based on such presentations or not.

Only one of my informants ticked off for *"other",* however no explanation was given.

The three assessment situations that were rated on top were: *formal oral presentation*, *class discussion* and *topic conversation*. After this, *"group discussion with other students"* came on fourth place, *"informal oral activities in class"* on fifth place and "*group presentation"* at last. Below is a table showing the results:

Table 6: Results from question 2

Answer	Importance
Formal oral presentation	6,7
Class discussion	5,2
Topic conversation	5,0
Group discussion with other students	4,8
Informal oral activities in class	4,1
Group presentation	3,7
Other	1



The seven informants who answered my interview and follow-up phone interview were not given options on this question.

Some of the answers they provided were:

"First and foremost, we focus on speaking English in my lessons, this goes for every lesson during the entire school year. For most of my students this is no problem, but there are always some students who are at a low level of achievement, and therefore need to use the Norwegian language in order to understand the content of the lessons. I use a lot of discussion in my teaching, where we talk about various topics, either in groups or the whole class together. One of my main goals in teaching is to develop the student's confidence in the subject, which is why I often talk with my students about different teaching lessons and – methods. I emphasize what the students already know, but what they believe is irrelevant, for instance the English language they use in gaming, social media, etc. I make room for the students interests in our lessons and by doing so, I sometimes move to topics who are not necessarily in the English subject curriculum. As long as we are speaking English, I believe all topics are relevant."

"The most common and most important assessment situations I use are reading out loud in class, presentations and conversations."

"Most often dialogue with the class and presentations. I try consciously to promote assurance and give much encouragement to participation in oral activities in class."

"The most important assessment situation/forms are oral presentations and discussions in class or in groups, mainly. Sometimes we also do listening tests where the students listen to a tape of a person speaking English, and then have to answer tasks that show whether or not they have understood the content."

Here I have chosen not to include more than four of the answers from the informants who answered my interview, because of the similarity of the answers.

To sum up, the results from my survey and interviews show that most of the informants view "formal presentations" as the most important assessment situation, followed by discussions both in class and in groups.

In addition, one of my informants specifically mentioned *"reading out loud"* as being the most important assessment situation. Listening to students reading out loud can give a good impression of the students' language abilities. When we observe the students reading, we can listen for intonation, fluency and vocabulary, which all are important skills in oral proficiency. Although only one of my informants mentioned reading out loud, *"informal oral activities in class"* was (as previously mentioned), the fifth most important assessment situation.

Several of my informants have stated that participation in any form of oral activity is something they find very important, sometimes even more important than answering a specific task or having a presentation. As long as the students use the English language and become comfortable with it, they have achieved something.

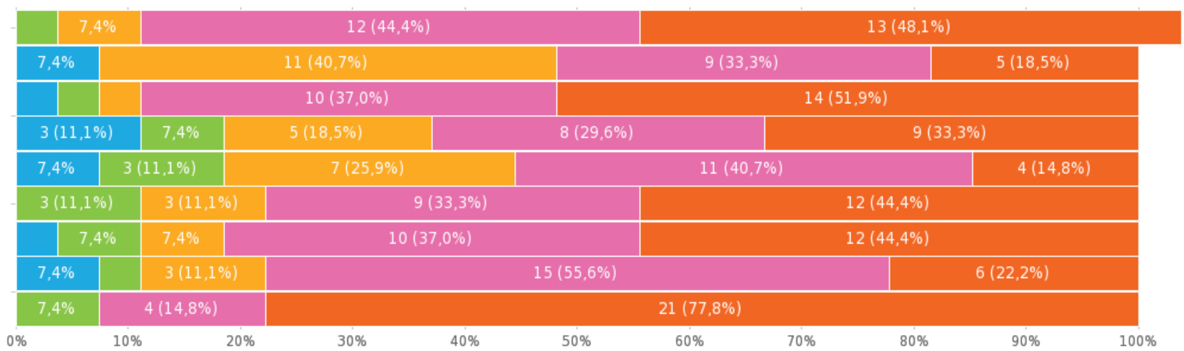
6.4 Assessment criteria

My informants were asked to arrange the following assessment criteria from 1-5, 1 being the least important, and 5 the most important assessment criteria; *vocabulary*, *pronunciation and intonation*, *content/knowledge*, *structure*, *correct use of grammar*, *fluency*, *the ability to answer the topic in accordance to the task*, *independence* and last but not least, *the ability to communicate*. What is meant by the criteria "independence" is the ability to show reflection and not only reproduce something, but put it into your own words.

77, 8 per cent of my informants rated "the ability to communicate" as the most important criteria, 51, 9 per put "content/knowledge" as the most important criteria, and 48, 1 per cent put "vocabulary" on top. "Fluency" and "the ability to answer the topic in accordance to the task" are the forth and the fifth most important criteria according to my informants. The table below shows all the results from the survey questionnaire:

Table 7: Results from question 3

	1	2	3	4	5
Vocabulary	0	1 (3,7 %)	2 (7,4 %)	12 (44,4 %)	13 (48,1 %)
Pronunciation and intonation	2 (7,4 %)	0	11 (40,7 %)	9 (33,3 %)	5 (18,5 %)
Content/knowledge	1 (3,7 %)	1 (3,7 %)	1 (3,7 %)	10 (37,0 %)	14 (51,9 %)
Structure	3 (11,1 %)	2 (7,4 %)	5 (18,5 %)	8 (29,6 %)	9 (33,3 %)
Correct use of grammar	2 (7,4 %)	3 (11,1 %)	7 (25,9 %)	11 (40,7 %)	4 (14,8 %)
Fluency	0	3 (11,1 %)	3 (11,1 %)	9 (33,3 %)	12 (44,4 %)
Answer the topic according to the task	1 (3,7 %)	2 (7,4 %)	2 (7,4 %)	10 (37,0 %)	12 (44,4 %)
Independence	2 (7,4 %)	1 (3,7 %)	3 (11,1 %)	15 (55,6 %)	6 (22,2 %)
The ability to communicate	0	2 (7,4 %)	0	4 (14,8 %)	21 (77,8 %)



As the table shows, there are big variations in what the informants view as the most important assessment criteria. Most of the numbers are even, only “the ability to communicate” sticks out as having by far the largest number of informants rating it on top.

The numbers are also even if we look at what the informants put on fourth place. The only assessment criteria, that sticks out as being rated on fourth place by least of the informants, is not surprisingly “the ability to communicate” - the reason being that most of the informants put that assessment criteria on top.

The three criteria rated on fourth place by most of the informants were: “independence” (55, 6 per cent), “vocabulary” (44, 4 per cent) and “correct use of grammar” (40, 7 per cent).

The results from my interview were as follows:

Table 8: Results from question 4

1	1. The ability to communicate 2. The ability to convey a vocational message 3. The ability to initiate and keep a conversation about everyday topics going 4. Precise and nuanced vocabulary 5. To show knowledge about relevant topics in the English –speaking world.
2	1. To show willingness and the ability to implement knowledge 2. Repeatedly to engage in oral activities in class 3. Good pronunciation and content 4. Fluency 5. Correct use of grammar
3	1. Speech: whether the student has a good pronunciation and speaks clearly and understandable. 2. Content: whether the student has included everything the task asks for, and is able to convey it in a good way. 3. Language: whether the language is suitable for the task and the situation. 4. Structure: whether the presentation has a good and clear structure. It should be easy for the audience to pay attention. 5. Grammar: whether the language is filled with structural and grammatical mistakes.
4	1. Communication 2. Reflection 3. Formulate coherent sentences 4. Pronunciation and intonation 5. Fluency

5	1. To have knowledge about a specific topic (content) 2. To answer in accordance with the task description. 3. To communicate in a good way. 4. To speak fluently. 5. To have a clear structure on the presentation/oral test.

Two of the informants who answered my interview did not rate assessment criteria according to importance; however, they gave other explanations. One informant said that she views all the assessment criteria as equally important, but emphasize that the ability to communicate with others has a special position.

The other informant stated that the overall impression is what counts. For this informant, the assessment criteria are related to each other, which he exemplified by saying that the lack of fluency automatically has a negative influence on the communication.

The results from my interviews corresponds with the results from the questionnaire survey, in that there are variations as to which criteria the informants view as the most important when they assess oral proficiency.

Although the importance of assessment criteria varies, the ability to communicate is placed at top three in four of the seven answers on my interview, or stated that it was the most important.

6.5 Who makes the assessment criteria?

I asked my informants to give me information about who participates in creating the assessment criteria. I wanted to know this because I was interested in finding out if there are different practices throughout the country, regards the development of criteria forms. I believe that assessment criteria for the same subject should be similar, and they can only be so, if some sorts of cooperation among the teachers occur.

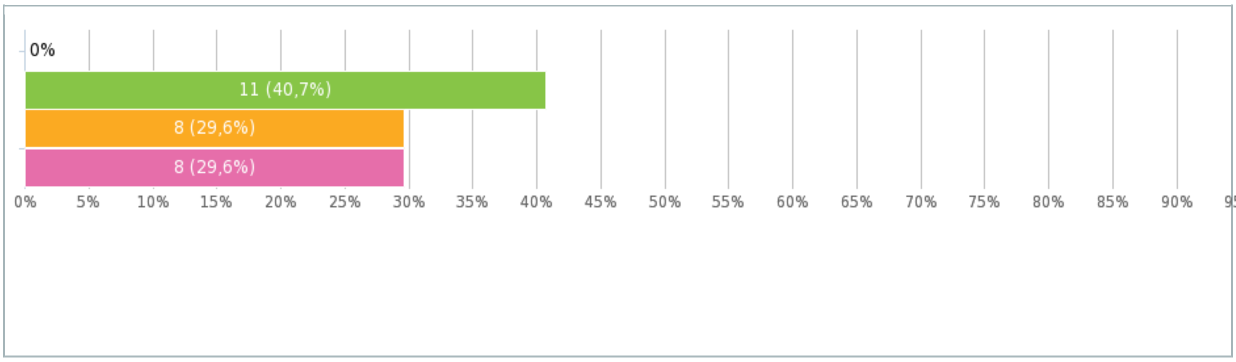
On this question I provided my informants four options: “the school administration”, “the English subject department” and “I form the criteria I give the students”. The forth option was for the informants to fill in their own response by ticking off for “other” and writing an explanation with their own words.

Before I go on to commenting on the results from this question, I will explain what is meant by “the English subject department” (my translation). Most schools in Norway divide their teachers into teams - usually the teachers who teach at the same level work together in a team, regardless of which subject they teach. It is however important to ensure that some cooperation between the teachers who teach the same subject occurs, especially when it comes to examination tasks and assessment. In order to do so, most schools have subject-specific departments with a set amount of meetings throughout the school year. On these meeting the teachers can share experiences in teaching, talk about topics, talk about assessment criteria, share good lesson plans, teaching material etc. for one specific subject. Because there is a good possibility that assessment criteria are made together in these subject departments, I have included “the English subject department” as an option.

Below is a table showing the results:

Table 9: Results from question 5

Answer	Number of answers	Percentage
The school administration	0	0 %
The English subject department	11	40,7 %
“I form the criteria on my own”	8	29,6 %
Other	8	29,6 %



40, 7 per cent of the informants answered that the assessment criteria are made within the English subject department. 29, 6 per cent answered that they form the criteria on their own, while 29, 6 per cent ticked off for "other" and gave an explanation. Some of the explanations were: "*the subject curricula decide which criteria I use*", "*teachers who teach at the same level form the criteria together*", "*I form the criteria using the local/county's criteria descriptors*", "*in cooperation with my colleagues, because it ensures that we use the same criteria on the same level*", "*I form the criteria on my own or together with my students. When it comes to big tests, we form the criteria in the English subject department".*

Out of the seven informants who answered my interview, four stated that they make the assessment criteria in cooperation with other English teacher at their school. The three remaining informants answered that they make the assessment criteria on their own, by either using templates from the local assessment criteria, or inspired by other assessment forms.

One of the informants specifically mentioned that the assessment criteria used by all English teacher at that school is based on the basic skills and the four objectives described in the English subject curriculum.

As the results from both the questionnaire surveys and the interviews show, most schools develop the criteria in cooperation with teachers who teach the same subject, either initiated by the teachers themselves or as part of the agenda in so-called "English subject departments".

Some of the teachers choose to form the criteria on their own, however they answers indicate that they use templates developed either by the Directorate for Education and Training, or by other teachers for inspiration.

The results also show that the development of assessment criteria is not something the school administration is involved in, as none of my informants have chosen that option.

6.6 Challenges in assessing oral proficiency

My next question was about challenges in teaching English. Specifically, I wanted to know what the teachers viewed as the most challenging factors in assessing oral proficiency.

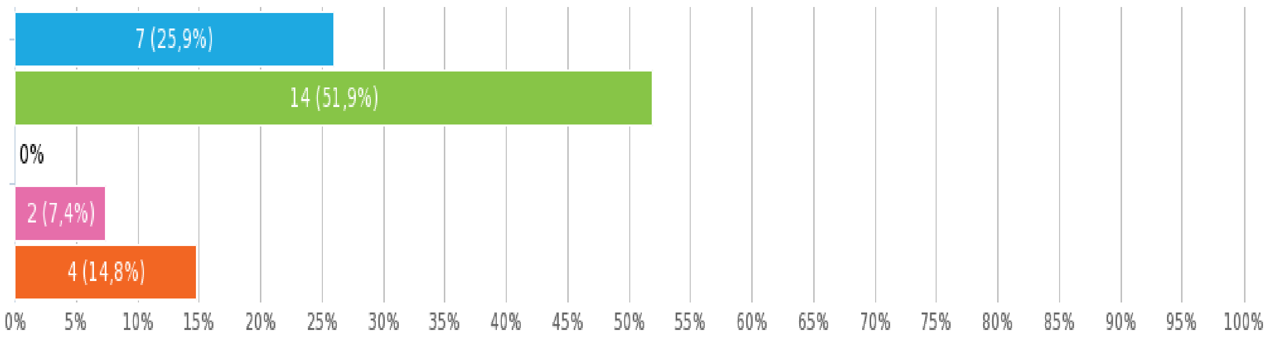
What teachers find challenging when assessing their students, can vary a lot. To make it easier to respond on this question, I provided my informants five options: *"Pupils who do not wish to speak English in front of classmates",* "*not enough time to assess as much as I wish"*, *"increase the pupils' motivation"* and *"make solid assessment criteria".*

The fifth option was to tick off for *"other"* and write an answer by words. I found it necessary to add this option because it is not always easy to anticipate what the teacher may find challenging and what they focus on when answering this question. Because the question before this was about assessment criteria, there is a possibility that some of the teachers would answers "making good assessment criteria" for no other reason than it being the first thing on their mind.

A total of 51, 9 per cent answered *"not enough time to assess as much as I wish".* 25, 9 per cent answered *"pupils who do not wish to speak English in front of their classmates",* whereas 7, 4 per cent answered that the biggest challenge in assessing students oral proficiency, is to *"make solid assessment criteria".* None of the respondents answered that it is challenging to *"increase the pupils' motivation"*, however 14, 8 per cent gave their own explanation to this question. Below is a table with all the answers:

Table 10: Results from question 6

Answer	Number of answers	Percentage
"Pupils who do not wish to speak in front of the class"	7	25,9 %
"Not having enough time to assess as much as I wish"	14	51,9 %
Increase the pupils' motivation	0	0 %
Make solid assessment criteria	2	7,4 %
Other	4	14,8 %



The explanations provided by the informants who ticked off for "other" were the following: *"make good assessment criteria for low-grade students"*, *"to communicate to certain students that it is necessary to do their best in assessment situations"* and *"confidence"*.

One of the informants answered *"it is easier to collect material digitally, because it is not as scary for many students who do not wish to speak in front of the class"*. This informant has not given an answer to the question the way it was intended, rather what can be done do solve the challenges (which was the next question in the questionnaire) – however, it implies that it is a challenge to assess students who do not wish to speak in front of the class.

The results from my interview are as follows:

One of the informants stated that the biggest challenge is having enough time to facilitate formal presentations in class, taken into consideration that the class is big. The same informant pointed to the duration of a lesson, it can be problematic if every student is to hold a 5-6 minute presentation during one lesson which lasts for 60 minutes. Then all the English lessons during that week would be spent on presentations.

Another informant stated that the assessment of pronunciation is problematic, in knowing how much it should be emphasized. She therefore gave credit for good pronunciation, and if the pronunciation was not good, she chose to not let it count on the assessment – assuming that the student could make himself/herself understood. The reason for this, she claimed is because many of the students struggle with pronunciation, and would get a low level of achievement.

Another challenge is when students complain on the final grade and you have to explain in detail how you assessed the student's oral proficiency. I use assessment criteria where I tick off the level of achievement in the different categories, in addition to short notes. However, in a class with 29 presentations, it is difficult to remember all the details, except from the short notes. Complains on course grades are definitely a challenge."

One of the informants also mentioned that it is difficult to assess student who are quiet and shy. The same informant pointed out that those students not necessarily are poor students, but students who manage to get good grades.

One of the informants said that he seldom run into problems regarding oral assessments, and that the main reason is the language lab, where the students can practice quietly for themselves.

The rest of the informants thought that the biggest challenge was to get through all of the presentations, because of the lack of time.

As the results show, not having enough time to conduct oral presentations in classes with many students is the absolute biggest challenge, according to the majority of both the informants who answered the survey, and the informants who answered the interview. One of my informants mentioned that it is a challenge to determine how much pronunciation should play a role in assessment, hence that the weighing of different assessment criteria can be difficult to establish.

The same informant mentioned that it can be problematic to give a detailed explanation of the assessment towards a grade in oral proficiency, in cases when a student submits a complaint. This can be problematic because a presentation must be assessed there and then – at the same time as the teacher listens to the presentation, he/she must take notes about the level of achievement on that particular student. If several presentations come after the first one, the teacher will not have much time to go back and fill the assessment form in detail. In this case, assessment of written tasks or tests, give the teacher the possibility to go over the text as many times as necessary, to look at the criteria over and over again to ensure that no mistakes are made. The teacher can keep copies of the texts, and if a student complains on the final grade at the end of the year, it would not be a problem to gather enough documentation to prove that the grade is indeed representative of the student's achievement.

Students who do not want to speak in front of the class, is the second biggest challenge, according to my informants. One of my informants points out that is it not necessarily the poor students who struggle when oral activities take place. On the contrary, good students can struggle just as much with being shy and anxious, especially in a presentation situation where "all eyes" are on them, and when they know that they will be given feedback in some way – usually a grade.

Only one of 34 informants stated that there are no challenges connected to the assessment of oral proficiency.

6.7 How to solve the challenges

In addition to asking my informants to reflect upon challenges they face when assessing students oral proficiency, I asked them what they do to solve these challenges and how they do it.

This question I formed as an open-ended question, requiring the informants to give an explanation in words. The depth and length of the answers varied. Because the informants have described different challenges, I am going to present their explanations according to topics.

6.7.1 How to solve the challenge of not having enough time to assess my students

As previously mentioned, not having enough time to assess the students', was rated as the biggest challenge by most of my informants and all of them provided an explanation of what they think is a possible solution.

Six teachers stated that in order to save time, they practice group presentation or "short" presentations. Short presentations can for instance be to talk about the book they are reading for two minutes, to share something they have read in the newspaper etc. One of the informants points out that group presentations do save time, however, it can be difficult to determine how much each of the students have contributed in the process of making the group presentation, which creates another problem.

Four teachers explain that they solve this by having their students' record audio files. They can record their own presentations as well as reading and send them to the teachers for assessment. In this way, it is not necessary to spend so much time on presentations during the lessons. One of the teachers argues that this makes assessment easier, because one can assess peacefully, without any students around, and one can go back and re-evaluate the recordings if needed.

Three teachers stated that they carry out individual assessments with each student, while they give the rest of the class written tasks they can work with on their own. In that way, every student get the possibility for individual assessments with feedback there and then.

6.7.2 How to ensure good assessment criteria

Some of my informants believe that *"making good assessment criteria"* is the biggest challenge in assessing oral proficiency. They describe the issues and how they deal with them as this:

One of the teacher explain that in order to make good assessment criteria, one must always try to improve them, based on what proves not to be sufficient.

Another teacher stated that the assessment criteria he uses are based on the competence aims and include different topics. The assessment criteria then have to be changed and adjusted to fit the purpose of each individual test.

A third teacher states that the best way to improve assessment criteria is to share ideas and experiences about the topic with other teachers.

6.7.3 How to assess students who do not wish to speak in front of the class

Having a presentation in front of other people is something most people struggle with. It does not get any easier when the presentation has to be in a language which is not your first language, and with a teacher who is supposed to make a judgement on both your language use, the content in your presentation, etc. This is not an easy situation for the teachers either –

they have to be able to assess every student, and it is their job to facilitate assessment situations to fit all the students.

Following are the thoughts my informants had on this issue:

One of the informants argue that this is a complex issue, and that the larger the class, the harder it is to practice oral activity. He also claims that it is difficult to give constructive feedback when you do not have more time for each one of the students.

Two informants explained that they solve this problem by giving the students the option to present in front of a smaller group of students, or in front of the teacher only.

Only one of the informants mentioned the use of “two-teacher system”. Before I go further, I will explain what is meant by the “two-teacher system”. Some schools have carried through with a system where they join two teachers who teach the same subject, to function as “assistants” in each other’s’ class. The teachers can either share the class in two or teach half of the class each, they can function as extra support for each other by for instance sharing responsibilities, one can teach while the other can help the students when they need guidance etc. This is a good resource, especially in classes with many students or in non-homogenous classes, where it is difficult to differentiate the lessons to fit all the students’ levels. Another positive impact of the two-teacher system is that one teacher can concentrate on teaching the class, while the other can take out students for individual assessment and feedback. This is however, only a resource in some schools. The same informant mentions digital presentations with spoken soundtrack or screencasts when every other option is impossible.

Several other informants have mentioned recordings of audio/films, as a solution for students who struggle with anxiety regards to presentations.

Although it can be a challenge to assess those students, it seems like the teachers use good methods to ensure that there are other options for students who struggle with anxiety of presentations.

6. 8 Self- evaluation

The last question I asked my informants was how they facilitate self-evaluation. Self-evaluation is one of the competence aims both after year 10 and VG1. In the English subject curricula it says that after year 10 the student should be able to *"comment on own work in learning English"* and after VG1 it says that the student should be able to *"evaluate own progress in learning English"* (LK06 – English Subject Curriculum). To assess students' oral proficiency can in many ways be more difficult than to evaluate their writing skills. For the students it is probably even more difficult.

I chose to add this question because I found it interesting to see if the teachers actually make their students practice on evaluating their own work, and because I was under the impression that this might be one of the competence aims that teachers do not prioritize. This question was formed as an open-ended question and the answers I will provide in the following table:

Table 11: Results from question 6

1	The students evaluate themselves by using an "evaluation form"
2	We look at some of the criteria given to them before the evaluation, and the students then write a text where they evaluate their own work.
3	Through topic conversations.
4	By varied lessons, for example. Evaluate their own texts by using forms or by own words, evaluate oral activities using assessment criteria.
5	The students have to write logs. In addition we have topic conversations that are based on the students' evaluations of their own achievement in the subject.
6	After every assessment situation/end of lesson. Either written or oral.
7	Fill-in forms for self-evaluation that the students' have to fill out after they

	have worked with a topic. The students also evaluate each other.
8	By asking them question and through topic conversations.
9	I let my students sit in pairs and evaluate their own and each other's texts.
10	I let my students' film themselves with their mobile phones during presentations, so they can (optionally) give me feedback through e-mail of how they think the presentation went.
11	The students evaluate themselves after tests.
12	The students evaluate all written tasks before they get a final grade. The students themselves on smaller tasks as well. There are specific question for self-assessment on everything the students is assessed at.
13	For example through direct follow-up questions about how he/she evaluates the presentation, at the end of the assessment situation, or as reflection through a later topic conversation. Written self-evaluations are also used, of course.
14	For some oral presentations, the students' are provided the possibility to use the same assessment criteria that I use. This has proven to be very demanding, especially for 8th and 9 graders, but also at 10th grade. The students' have difficulties with looking at themselves "objectively", they put too much weight into small mistakes and by large undervalue their achievement in some areas, while overvaluing other areas. This problem occurs even if we together go through the tasks and the criteria, in detail in advance.
15	Often as part of a task.

16	I integrate self-assessment when I give the students' feedback on their work. The students have to assess their own work before they get the grade.
17	They evaluate each other's presentations, and by this, they also get familiar with the criteria. These are used on their own products.
18	I often demand that the students answer a form, either with general questions about their "achievement", or by evaluating their own achievement by ticking off and commenting the assessment criteria, so that they have an understanding of what the assessment is based on. Many students do not use the assessment criteria we go through, which they are handed before every test, as they are intended. However, by focusing on self-assessment, I believe the students get a clearer idea of how they can improve themselves.
19	By evaluating their own and other students' texts. By listening to presentations – ticking off on an assessment criteria form that the students have made in advance.
20	By grading their own work. By writing what they have learned on tags. By filling out forms where they evaluate themselves. I must admit that I am not good at facilitating this.
21	I give the students a form with questions that they have to answer at the beginning of a lesson.
22	In conversation with the student after a performance. I do not use forms. I ask how the student thinks it went. If you have a student several years we will deal about what the student will try to do better next time. In other words: this is a process.
23	My students have good practice at this, orally or with «signs». I seldom use written self-assessment tasks.

	On the other hand, the way I do it saves time and I use it a lot.
24	They evaluate themselves.
25	We facilitate peer- assessment during both written and oral presentations. This promotes self-assessment in form of reflection. The students deliver written self-assessments after bigger tests.
26	Little…
27	The best way to do it is through process-oriented tasks, and by going through the criteria before starting on a new topic. In the middle of the process, the student is given individual feedback. Improvement on some of the criteria motivates and substantiates growth.

As the results show, some teachers practice self-assessment with their students quite often, either by assessment of their own tests, or by practicing peer-assessment. Other teachers admit to not facilitate self-assessment at all, or fairly seldom.

The minority of the informants use forms on self-assessment, but rather integrate it in conversations with the students or as a part of a task.

According to the Regulations to the Act of Education, self-assessment is part of the formative assessment, and the purpose of self-assessment is that the student, apprentice or trainee should reflect upon their own teaching. The student, apprentice or trainee should participate actively in the assessment of their work, their learning progress and development (Forskrift til Opplæringslova§ 3-12).

6.9 Chapter summary

In this chapter I went through the main findings from the questionnaire and interview. Each section presented one of the questions my informants had to answer.

7 Discussion

7.1 Introduction

In the previous chapter I presented the findings from my questionnaire and survey in detail. In this chapter I will focus only on the most important findings, which are in particular relevant in order to answer my research question. I will present findings regards to assessment, classroom size, and the distribution of oral assessments, compared to written assessments. The relevant findings will be discussed with theory from chapter 2 Speaking, chapter 3 Assessment and chapter 4 Classroom size. First, I will have a look at the research question.

7.2 Revising the research questions

The research question for this thesis was as follows:

The course grade in VG1 and 10th grade English: *How do teachers assess oral proficiency towards the final course grade?*

For this research question, three subsidiary questions were made, which I wanted to look at in further detail:

1. How often are teachers assessing students' oral proficiency compared to other types of assessment e.g. written tests?
2. Which criteria are used when assessing students' oral proficiency and who makes the criteria?
3. Does classroom size play any role when assessing the student's oral proficiency?

My findings showed that there are variations when it comes to the distribution of oral and written assessments. Although the majority of my informants have a 50/50 distribution of oral and written assessments, a large number of my informants admit to assess written tasks more frequently. An uneven distribution can in some cases, where a student has a higher level of achievement either in written or in oral skills, cause a student to either get a higher grade or lower grade than what would have been the case if the distribution was even. It is important to remember that for many students the course grade is a high-stake decision, for example for a 10th grade student who wishes to apply for a certain upper secondary school, where the competition is big. This also applies to VG1 students who, for instance, wish to study abroad for a semester or a year, and the school they wish to attend has specific admission requirements. One grade can make a difference.

As for my second question, the findings are presented in detail in the previous chapter. My findings points to variations in which assessment criteria the teachers view as being the most important. *The ability to communicate* is by the absolute majority rated as the most important assessment criteria, however, the numbers are uneven as to which assessment criteria are placed on fifth, fourth, third, etc. place. In practice, that indicates that there are differences as to what different teachers look for and emphasize when they are assessing students' oral proficiency and that they do not weigh the assessment criteria equally. Explicitly, one of the informants stated that pronunciation only counted if the student had good pronunciation, and if the students' pronunciation was poor, it would not have a negative impact on the overall assessment. In the same question I asked my informants who was in charge of making the assessment criteria. The results from that question indicate that even though some teachers make their own criteria, some cooperation between the teachers at the same schools exists. The findings also revealed that the some of the schools use the national guidelines for assessment of the course grade in English, provided by UDIR.

As to my third question, I did not actually ask my informants about classroom size; however I asked them about what they find to be the biggest challenge in assessing oral proficiency. This I did purposely, in order to eliminate the possibility of leading questions. All the results are listed in the previous chapter. The main obstacle in assessment of oral proficiency according to more than 50 per cent (of my informants who answered the survey) was not having enough time to assess in a sufficient amount. Although that does not directly point to being caused by the number of students' in the class, there is a big possibility for that being the reason. When it comes to the seven informants who answered my interview, the message was clearer. Three out of seven explicitly mentioned that the size of the class makes it difficult to go through presentations by each student.

7.3 Assessment

Assessment towards the final course grade in oral English should measure what the student has learned by the end of the course, thus a summative assessment. Assessment towards the final course grade is the formative assessment, where teachers constantly observe, listen, evaluate, give feedback, test and use the information they gather about the students' level of competence in all the different areas of a subject, in order to help them improve and make progress. The summative assessment of a course grade is in my opinion, therefore absolutely connected to the formative, which is in line with what Black and William (1998) say about formative and summative assessment: "from the information that teachers gather for formative purposes, they should, with selection and re-interpretation, be in a strong position to contribute to a fair summative report on each student" (Black & William, 1998, p. 12).

The relationship between summative and formative assessment is inevitable in deciding the course grade in English, unlike the examination grade where the students' achievement is merely based on what he or she presents there and then. Because I wanted to look at the assessment towards the course grade in oral English, I therefore had to gather information about the assessment process throughout the school year.

In the process of ensuring the assessment to be as objective, reliable and valid as possible, the teachers must avoid inconsistencies in which assessment criteria they use, they must ensure that the tests of oral proficiency have a certain purpose and are formed in such a way that what is being tested, is indeed possible to measure. Bachman and Palmer (1997, p. 20) claims that at the same time, we need to recognize that is not impossible to eliminate inconsistencies entirely. What we can do, however, is to try to minimize the effects of those inconsistencies through test design (Bachman & Palmer, 1997, p. 20).

Except from "the ability to communicate" my study has revealed that there are inconsistencies as to which assessment criteria are emphasized by the teachers. The importance of communication is underlined in the English subject curriculum, where it among other things is stated that learning the English language is "making a significant contribution to our communicative abilities" (KD 2013)

According to Luoma (2004) the most common ways to ensure reliable speaking tests are the use of assessment criteria and rater training. Based on the answers from my informants, the different schools have different practices concerning the development of assessment criteria. The results show that the school administration is not involved in the development of the assessment criteria. The criteria are formed within the cooperation of the English teachers, or formed by the teachers on their own. In other words there is no common rule for which criteria are used, and it is up to the teachers to decide what to emphasize when assessing students' oral proficiency. Only two of my informants mention that they use the guidelines for assessment provided by UDIR.

As mentioned in the previous section, my findings also show that although more than 50 per cent of my informants have an equal division of oral and written assessments throughout one school year, 33 per cent admit to have a stronger focus on written assessments and tasks.

In the English subject curriculum this is stated about the overall assessment after year 10: "The pupils shall have one overall achievement grade for written work and one overall achievement grade for oral performance" and about the overall achievement after VG1: "The pupils shall have one overall achievement grade" (KD, 2013). Nothing is stated about how many oral assessments and how many written assessments the overall achievement should be based on. Having in mind that the course grade is what Bachman and Palmer (1997) refer to as high stakes decision for the majority of students, meaning that the course grade is likely to have a major impact on their lives, there should be en equal possibility to show competence through assessment in both oral and written skills.

Challenges in assessment of oral proficiency were one of the topics my informants had to reflect upon. The results of my findings indicate that the size of the class possibly can have an impact on assessment, because 52 per cent of my informants answered that the biggest challenge was not having enough time to asses as much as they wish. Some of the informants explicitly mentioned class size as challenge.

However, that is only a possibility. The lack of time can be connected to many other reasons beside the size of the class. That goes hand in hand with what Hattie (1999) has to say about the impact of the class size, when he concludes that there is little evidence that instruction methods change when class size is reduced, although a large part of improvement can be explained by improvements in student task engagement. Reducing class sizes from the 30's to the 20's is in the right direction, but there is little support for the claim that there are increases in student achievement or satisfaction, or teacher attitude or morale. Only when the class size reduces to 15 or below are there appreciable positive benefits (Hattie, 1999, p. 9) .Factors such as teaching methods and the level of the students in a particular class, can be just as important factors as the size of the class.

A test, like any other type of measurement used to measure, should give the same results every time it measures (if it is used under the same conditions), should measure exactly what is supposed to measure (not something else), and should be practical to use (Brown, 1996, p.185). In order for the assessment of oral proficiency to be valid and reliable, it is important that the test objectives are carefully carried out. The assessment criteria should be specific and formed in such a way that both teachers and students know exactly what is being tested. The results of my study show that this is not the case, when teachers do not use the same assessment criteria and when they weigh the assessment criteria differently. This can be exemplified by one of my informants who stated that she only assessed pronunciation if the pronunciation was good, but if the student struggled with pronunciation, it would not show on the overall assessment.

According to the Regulations to the Act of Education, the final course grade which will be given at the end of the course get a mark on the diploma. The course grade must be based on a wide range of assessments, which shows the students' level of competence in the subject. (Forskrift til Opplæringslova § 3-3 og § 3-16).

Not all situations are suitable for assessment. When my informants were asked which situations they viewed as suitable for oral assessment according to importance, the answers were different. Even though the majority of my informants viewed oral presentations as the most important assessment situation, it became clear to me that some teachers include all oral activities as suitable for assessments, while some teachers only view formal oral activities as suitable for assessment. It seems as the distinction between formative and summative achievement is not clear, taken in consideration that my question was in relation to the course grade, which measures the summative assessment of a student. That goes hand in hand with what Black and Wiliam (1998) view that it may be difficult for teachers to combine the formative and summative roles, although the information teachers gather for formative purposes, with selection and re-interpretation, can contribute to a fair summative report on each student.

Last, but not least, the answers from my informants indicate variations in how and to what extent self-assessment is carried out by the different teachers. Although self-assessment is one of the competence aims in the English subject curriculum, the results of my revealed that there are inconsistencies in the amount of self-assessment given by the different teachers, and also that self-assessment is carried out differently by the different teachers. Some teachers implement self-assessment quite often, as part of a conversation with the student, some facilitates self-assessment through written tasks, and others do not facilitate self-assessment at all.

To ensure the validity of tests towards the final grade in English, the understanding of assessment criteria should be equal. The English subject curriculum should therefore include specific and detailed descriptions of the assessment criteria. Another possible solution for improvement could be to involve the county administrators and school leaders in the process of implementing equal assessment processes and demand that the same process is undertaken by all the teachers.

Another possibility to ensure equal practice in the assessment process is to implement it as part of the teacher training.

7.5 Limitations and validity of the study

7.5.1 Limitations of the study

This study is based on answers from 34 teachers who either teach English at VG1 or 10^{th} grade. The informants have different educational backgrounds as well as teaching experiences, and they represent fifteen out of the nineteen counties in Norway. Compared to other studies at this level, the number of informants is fairly high, and only four counties were not represented. Having this in mind, the number of informants is not representative enough to make generalizations. On the other hand, teachers, students and others may find the study interesting, and it may inspire others to look further into this topic and contribute to more research.

7.5.1 Validity of the study

The validity of this study is restricted to the number of informants interviewed. 34 informants participated in this study, 27 answered a questionnaire survey and 7 answered an interview with a follow-up phone interview.

For the study to be more valid, it should be based on answers from a larger number of informants. The study can however inspire to a larger study of this topic. This study can serve as a starting-point to a future large-scale study about assessment towards the course grade.

7.6 Chapter summary

This chapter provided a presentation of relevant findings, discussed with support from the theory outlined in chapter 2, chapter 3 and chapter 4. The validity and the limitations of this study were discussed.

8 Conclusion

8.1 Have the research questions been answered?

At the beginning of this thesis I presented a research question which I wanted to investigate in this project: “The course grade in VG1 and 10th grade English: *How do teachers assess oral proficiency towards the final course grade?”*

On the basis of the data collected from my informants, I would argue that there are some inconsistencies in the assessment process of oral proficiency towards the course grade in VG1 and 10th grade. Based on answers from my informants, it is possible to argue that there are inconsistencies regard the use of assessment criteria, the significance of assessment criteria and assessment situations and the distribution between oral assessments and written assessment.

After investigating the process of assessment towards the course grade in VG1 and 10th grade oral English, I see some correspondence with results from similar studies of the oral examination, such as for instance Yildiz (2011), who concluded that with pointing to undue variation concerning both format and assessment of the oral examinations and Henrik Bøhn who concluded that there are inconsistencies in how teachers value the constructs of criteria in EFL oral exams, and that there are variations in how teachers view the significance of content knowledge.

8.2 Suggestions for future studies

As previously mentioned, this is a small scale study, and no generalizations can be made from my findings.

It would be interesting to carry out a large-scale study of the topic, to see if the results can be confirmed and if they apply to a wider range of the schools in Norway.

This study concentrated only on answers from teachers. It would also be interesting to investigate the students’ thoughts about the topic.

References

Bailey, K.D. (2007) *Methods of Social Research* (Fourth edition). New York: The Free Press

Bachman, L. F., & Palmer. A. (2010). *Language Assessment in Practice*. Oxford, England: Oxford University Press.

Bachman, L. F., & Palmer. A. (1996). *Language Testing in Practice*. Oxford, England: Oxford University Press.

Berg, Ole T. (2015) Fylker I Norge. I store norske leksikon. Retrieved from: https://snl.no/fylker_i_Norge

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

Brown, J. D. (1996) Testing in Language programs. Prentice Hall, inc.

Bøhn, H. (2015) *Assessing Spoken EFL Without a Common Rating Scale: Norwegian EFL Teachers' Conceptions of Construct.* SAGE open. Retrieved from: http://sgo.sagepub.com/content/spsgo/5/4/2158244015621956.full.pdf

Cohen, L., Manion, L. & Morison, K. *Research Methods in Education.* (7th edt.) USA and Canada: Routledge

Council of Europe (2004). "Chapter 9: Assessing". In: *The Common European Framework of Reference for Languages: learning, teaching, assessment.* Retrieved from: http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Creswell, J. W & Clark V.P. (2011) *Designing and Conducting Mixed Method Research.* United States of America: Library of Congress Cataloging-in-Publication Data.

Forskrift til Opplæringslova. (2009) *Kapittel 3. Individuell vurdering i grunnskolen og i vidaregåande opplæring*, Retrieved from: https://lovdata.no/dokument/SF/forskrift/2006-06-23-724/KAPITTEL_4#KAPITTEL_4

Fulcher, G. & Davidson, F. (2007). *Language Testing and Assessment. An Advanced Resource Book.* London and New York: Routledge.

Greene, J.C., & Caracelli, V. J., (Eds.). (1997) *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms: New directions for evaluation designs. Educational Evaluation and Policy Analysis,* 11(3), 255-274

Hasselgård, H. Johansson, S. Låsvåg, P. (2010) *English Grammar: Theory and use.* Oslo: Universitetsforlaget

Hattie, J. (1999) *Influences of student learning*, University of Auckland. Retrieved from: http://xn--www-rp0a.teacherstoolbox.co.uk/downloads/managers/Influencesonstudent.pdf

Johannessen A., Tufte, P.A. & Kristoffersen, L. (2006 [2002]). *Introduksjon til samfunnsvitenskapelig metode* (Third Edition). Oslo, Abstrakt forlag.

KD (2013) *English subject curriculum*. Oslo: Ministry of Education and Research. Retrieved from http://www.udir.no/kl06/ENG1-03/Kompetansemaal?arst=1858830316&kmsn=-915046718

KD (2013) *English subject curriculum,* Oslo: Ministry of Education and Research. Retrieved from: http://www.udir.no/kl06/eng1-03/Hele/Vurdering/?lplang=eng

KD (2013) *English subject curriculum*. Oslo: Ministry of Education and Research. Retrieved from: http://www.udir.no/kl06/ENG1-03/Hele/Hovedomraader

KD (2013) *English subject curriculum*. Oslo: Ministry of Education and Research. Retrieved from: http://www.udir.no/kl06/ENG1-03/Kompetansemaal?arst=98844765&kmsn=-1974299133

KD (2011) *Vurdering – Standpunktvurdering*, Oslo: Ministry of Education and Research.

Retrieved from:

http://www.udir.no/globalassets/upload/vurdering/kjennetegn/engelsk_kjennetegn_bm.pdf

Luoma, S. (2004). *Assessing Speaking (Cambridge Language Assessment Series).* Cambridge, Cambridge University Press.

Manger, T. (2010). Motivasjon og læring. In T. Manger, S. Lillejord, T. Nordahl and T. Helland, *Livet i skolen* (pp.279-309). Bergen: Fagbokforaget

Mosteller, F. (1995) "The Tennessee Study of Class Size in the Early School Grades." In: *The Future of Children: Critical Issues for Children and Youths.* Retrieved from: https://www.princeton.edu/futureofchildren/publications/docs/05_02_08.pdf

Schmitt, L. V. & Frey B. (2010) "Teachers' Classroom Assessment Practices". In: *Middle Grades Research Journal.* (Vol. 5).United States of America: IAP-Information Age Publishing

Shafipoor, M. & Latif, F (2013) / Procedia - Social and Behavioral Sciences 70 469 – 475

Simensen, A. M. (2007 [1998]). *Teaching English as a Foreign Language: Principles and Procedures. Second* Edition. Bergen: Fagbokforlaget Vigmostad & Bjørke AS

Slemmen, T. (2009). *Vurdering for læring i klasserommet*. Oslo: Gyldendal

Wiliam, D. (2011). What is Assessment for Learning? Retrieved from: http://www.udir.no/PageFiles/Vurdering%20for%20laring/Dokumenter/Bibliotek/2/What%20is%20assessment%20for%20learning%5B1%5D.pdf

Yildiz, L.M. (2011). *English VG1 level oral examinations: How are they designed, conducted and assessed?* (Master's thesis, the University of Oslo, Oslo, Norway). Retrieved from: http://urn.nb.no/URN:NBN:no-29277

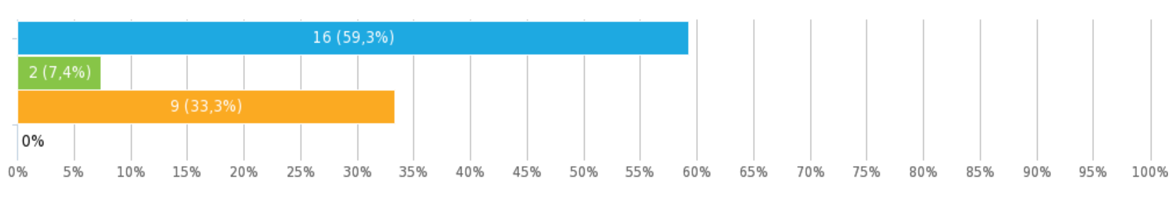
Zyngier, D. (2014) *Class Size and Academic Results*. Evidence Base. Retrieved from: https://journal.anzsog.edu.au/publications/9/EvidenceBase2014Issue1.pdf

Appendix 1: Results from questionnaire

Results from question 1:

Hvor mange muntlige vurderinger gir du elevene dine i forhold til skriftlige vurderinger i løpet av et skoleår?

Svar	Svar	Andel
"Jeg gir elevene omtrent like mange skriftlige og muntlige vurderinger i løpet av et skoleår"	16	59,3 %
"Elevene får flere muntlige enn skriftlige vurderinger"	2	7,4 %
"Elevene får flere skriftlige enn muntlige vurderinger"	9	33,3 %
Annet	0	0 %

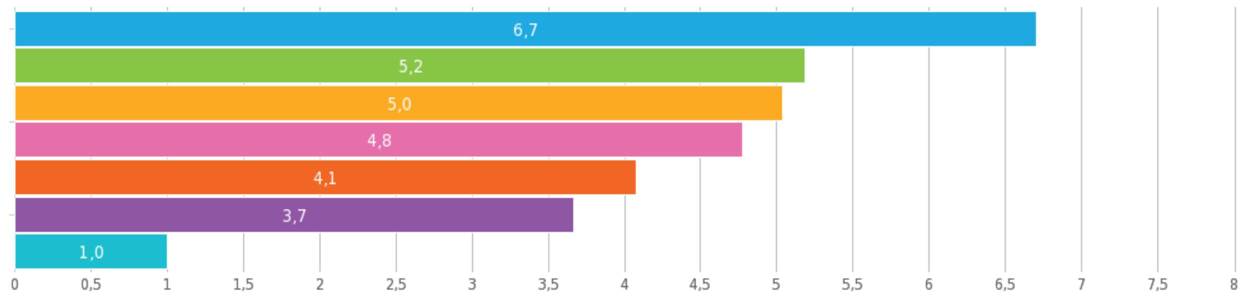


Results from question 2:

På dette spørsmålet skal du rangere alternativene etter viktighet. Øverst rangerer du den vurderingsformen/situasjonen du mener er viktigst for å kunne vurdere elevens kompetanse. Hvilke vurderingsformer/situasjoner bruker du når du vurderer elevenes muntlige kompetanse i engelskfaget?

Svar	Viktighet
Formell muntlig presentasjon	6,7
Klassesamtale/klassediskusjon	5,2
Fagsamtale	5,0
Samtaler i gruppe med andre elever	4,8

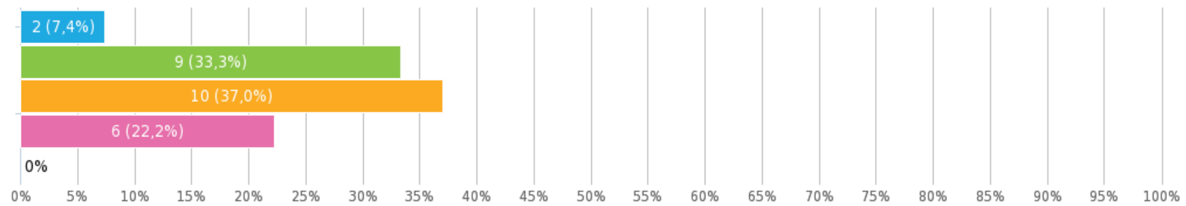
Uformell muntlig aktivitet i timene	4,1
Gruppepresentasjon	3,7
Annet	1



Results from question 3:

Hvor mange muntlige vurderinger baserer du sluttkarakteren i 10.klasse/VG1 engelsk på?

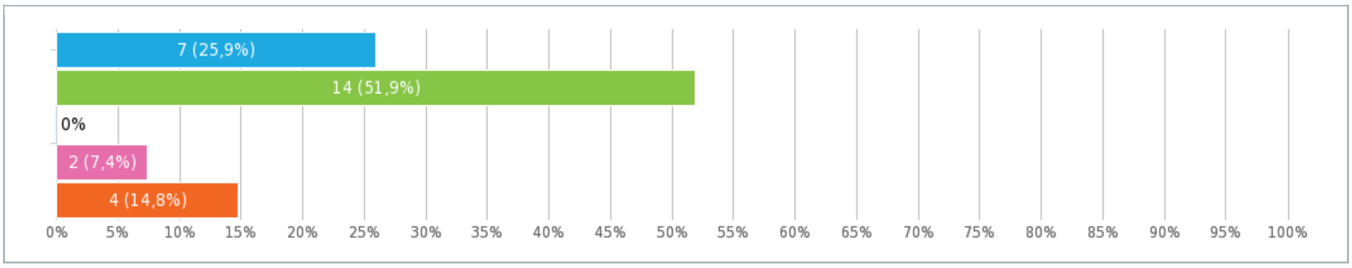
Svar	Svar	Andel
1-2	2	7,4 %
3-4	9	33,3 %
4-5	10	37,0 %
Flere enn 5	6	22,2 %
Annet	0	0 %



Results from question 4:

Hvilke utfordringer møter du på når du vurderer elevenes muntlige kompetanse i engelsk?

Svar	Svar	Andel
Elever som ikke ønsker å snakke foran klassen	7	25,9 %
Å få nok tid til å vurdere i ønsket mengde	14	51,9 %
Øke elevenes motivasjon	0	0 %
Lage gode vurderingskriterier	2	7,4 %
Annet	4	14,8 %



Annet:

- Lage gode vurderingssituasjoner for svake elever
- Enklere å samle inn materiale digitalt der de som ikke liker å snakke i klassen får vist seg frem.
- Selvtillit
- Å få kommunisert til samtlige at de må gjøre et arbeid der de får vist sitt beste når de skal vurd.

Results from question 5:

Ta utgangspunkt i det du svarte på spørsmål 4. Hvordan løser du disse utfordringene?

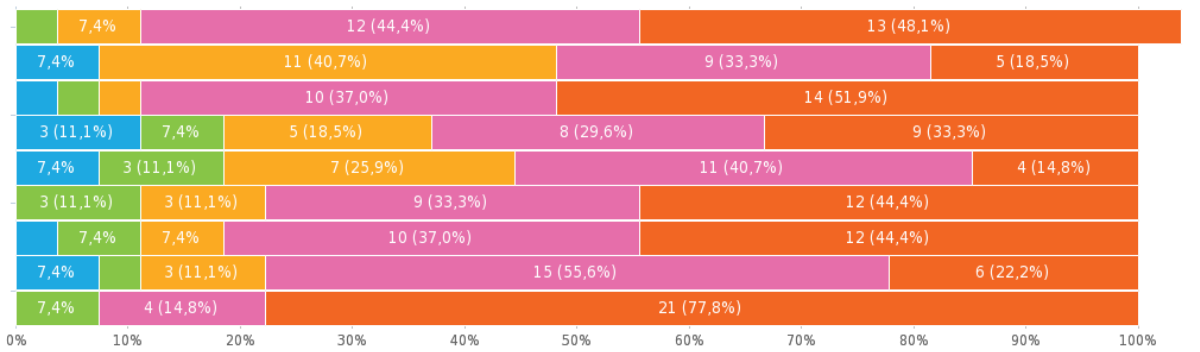
#	
1	jeg setter dem i grupper hvor de snakker sammen og jeg vurderer, jeg lar enkelte fremføre småting og vurderer muntlig varslede lekser
2	Bruker lydfiler de kan sende inn. La de filme seg selv og sende lenken inn til meg
3	Dette er en sammensatt problemstilling. Desto større grupper, desto vanskeligere å få muntlig aktivitet. Sammenhengen mellom muntlige presentasjoner, fagsamtaler og uformell vurdering i klasserommet vektes etter behov, alt etter aktivitet og måloppnåelse/det nivået eleven ligger på. Det er ofte dessuten vanskelig å gi konstruktive tilbakemeldinger når man ikke har mer tid til hver enkelt.
4	Lager flere vurderingssituasjoner, f.eks. framføring for bare meg eller noen få elever, individuell samtale om en tekst mm. Gruppeframføring for bare meg.
5	Elever har individuelle presentasjoner foran lærer/tolærer.
6	Gir mulighet til muntlige presentasjoner for bare lærer eller for få medelever Prøver å lage gode presentasjonsformer.
7	Ber elevene lese inn lydfiler som de leverer, og som jeg som lærer kan vurdere når jeg har mulighet og ro
8	Bruker digitale opptaksverktøy og har individuelle samtaler rundt en tekst, en film, et tema.
9	Lar klassen sitte med skriftlige oppgaver og ta ut 2-3 elever for samtale om tema.
10	Bruker tolærersystem og delte klasser. Elever KAN også levere digitale presentasjoner med innlest lydspor eller Screencasts, hvis alt annet ikke nytter. En og en elev "ut" er det lite tid til, men for noen få kan dette være en løsning.
11	Individuell vurderingssituasjon mens resten av klassen arbeider med skriftlige oppgaver. Muntlige presentasjoner i grupper på 6-7 elever slik at elevene ikke må høre på presentasjoner i 4 timer.
12	Lar elevene presentere på enerom med lærer. Tar elevene ut på gangen til samtale. Lar elevene presentere foran en venn eller to, men lærer hører på.

	Lar elevene spille inn lyd/film til innlevering for muntlig vurdering
13	Lager mange læringssituasjoner der elevene får øve seg i muntlig engelsk i ulike situasjoner og på ulike emner. Disse trenger ikke en formell vurdering, men har læring fram mot senere vurderingssituasjoner som formål.
14	Ingen elever i min klasse har anledning til å undra seg formelle presentasjoner, men en del elever deltar ikke i klassediskusjoner eller uformell muntlig samtale i timene, i hovedsak løser jeg dette problemet ved at jeg snakker direkte med disse eleven i timene. Jeg spør dem direkte spørsmål om temaet vi arbeider med og deres meninger om forskjellige hverdagslige ting. Jeg oppfordrer også disse elevene til å lese høyt i klassen da dette gir mindre direkte press på å formulere egne setninger, men likevel gir meg som lærer mulighet til å høre og vurdere deler av deres muntlige kompetanse - en fordel ved å lese høyt er også at eleven på en enklere måte kan forberede seg på dette i forkant ved egenlesing og ved en eventuelt utvalgt del av teksten som kan særlig øves på.
15	Tar utgangspunkt i kompetansemålene og vurderingskriterier for andre oppgaver. Endrer og justerer. Deler og får fra andre.
16	Jeg snakker med elevene i små grupper og har sjelden den klassiske powerpoint presentasjonen...
17	Jobbe med det mentale. Viktig del av læringsmiljøet. Gode opplevelser får snøballen til å rulle.
18	Det er viktig for meg at elevene ikke opplever at hver engelsktime er en vurderingssituasjon, men den kompetansen de viser i timen kan være positiv med tanke på å bekrefte vurdering på formelle vurderingssituasjoner (f. eks. muntlig presentasjon). Det er ALTFOR få timer i engelsk på ungdomsskolen (kun to klokketimer i uka), noe som begrenser antall vurderinger det er mulig å gjennomføre. Det er derfor viktig å være kreativ og ikke nødvendigvis alltid bruke undervisningstid på å vurdere, men at elevene for eksempel kan spille inn lyd/lage digital storytelling/film som kan vurderes utenom undervisningstid.
19	To økter engelsk i uka på 10.trinn er litt liten tid. Det å ha framlegg foran klassen tar tid når klassene er store. I enkelte tilfeller må jeg la elevene legge frem for meg bare for å komme i mål. Bruker mindre fremlegg og diskusjoner med mindre innlegg foran klassen til å vurdere nivået muntlig.
20	Har individuelle høringer med stoppeklokke. Har felles muntlige innspillinger av svar på spørsmål eller presentasjoner i språklaben. Bruker tid i språklaben. Lager fellesopplegg som klassen drifter selvstendig mens jeg tar ut elever.
21	Lager generelle vurderingskriterier. Bør kanskje ha de smalere, men er vanskelig når det for eksempel er en gruppepresentasjon. Tar ofte med at læreren kan stille spørsmål til stoffet dersom dette oppleves som mangelfullt. Det er også utfordrende at elevene selv føler at de er i stand til å "vurdere" seg selv opp mot andre elever dersom det er muntlige presentasjoner i klasserommet. Igjen, er det viktig å ha gode kriterier for muntlig kompetanse, men jeg synes det er utfordrende.
22	1. Oppgaven må være tydelig i bestillingen. 2. Følge opp arbeid underveis 3. Bruke mine undervisningsfrie timer til å ta ut enkeltelever eller grupper slik at timene i størst mulig grad kan gå til engelsk aktiviteter og ikkei hovedsak brukes til framføringer. Noen ganger er det selvfølgelig lærerikt for elevene å bruke tid på å høre på hverandre også. I store klasser vil dette ta lang tid om for eksempel en framføring tar 4-5 minutter.
23	Bruker mye diskusjoner og gruppesamtaler. Elevene får caser og vurderingskriterier
24	Små korte fremføringer
25	Elever leverer lydfiler eller filmer hverandre
26	Presenterer i gruppe for å spare tid.
27	Hører noen elever på gangen, Hører noen elever oftere enn andre og lytter når eleven øver dialoger.

Results from question 6:

Hvilke vurderingskriterier legger du til grunn i karaktersettingen? Ranger de oppførte kriteriene fra 1-5 hvor 1 er det kriteriet du anser som minst viktig, og 5 er det kriteriet du anser som svært viktig for karaktergrunnlaget.

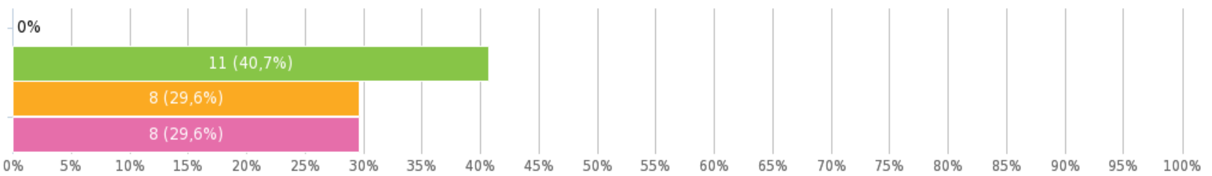
	1	2	3	4	5
Ordforråd	0	1 (3,7 %)	2 (7,4 %)	12 (44,4 %)	13 (48,1 %)
Uttale og intonasjon	2 (7,4 %)	0	11 (40,7 %)	9 (33,3 %)	5 (18,5 %)
Innhold/kunnskap	1 (3,7 %)	1 (3,7 %)	1 (3,7 %)	10 (37,0 %)	14 (51,9 %)
Struktur	3 (11,1 %)	2 (7,4 %)	5 (18,5 %)	8 (29,6 %)	9 (33,3 %)
Riktig bruk av grammatikk	2 (7,4 %)	3 (11,1 %)	7 (25,9 %)	11 (40,7 %)	4 (14,8 %)
Flyt	0	3 (11,1 %)	3 (11,1 %)	9 (33,3 %)	12 (44,4 %)
Eleven svarer i samsvar med oppgavebestillingen	1 (3,7 %)	2 (7,4 %)	2 (7,4 %)	10 (37,0 %)	12 (44,4 %)
Selvstendighet	2 (7,4 %)	1 (3,7 %)	3 (11,1 %)	15 (55,6 %)	6 (22,2 %)
Evne til å kommunisere	0	2 (7,4 %)	0	4 (14,8 %)	21 (77,8 %)



Results from question 7:

Hvem utarbeider vurderingskriteriene?

Svar	Svar	Andel
Skolens ledelse	0	0 %
Fagseksjon	11	40,7 %
Jeg utarbeider kriteriene selv	8	29,6 %
Annet	8	29,6 %



Annet:

- fagteamet i engelsk vg 1
- Læreplanene styrer hva jeg skal legge til grunn
- Fagforum på fylket når det gjelder muntlig, skriftlig: udir
- Jeg utarbeider kriteriene, men ser på skolens / fylkets kriterier til muntlig eksamen
- Lærerne på det aktuelle trinnet utarbeider vurderingskriterier sammen
- Gjerne sammen med kollegaer, for da har alle på trinnet de samme vurderingskriteriene.
- Selv, eller sammen med elevene. I forhold til store prøver lager fagseksjonen dem sammen.
- Vi bruker nasjonale kriterier.

Results from question 8:

Hvordan legger du til rette for egenvurdering?

1	elevene kontrollerer/retter seg selv ut fra en "mal"
2	Vi ser på noen av kriteriene de fikk utdelt før vurderingssituasjonen og skriver en tekst til seg selv som de kan sende til meg der de vurderer eget arbeid
3	Fagsamtaler
4	Ved varierte opplegg, f.eks. kommentere/vurdere egen tekst enten med ord eller skjema, vurdere eget muntlig arbeid etter kriterier
5	Elevene skriver logg, samt fagsamtaler tar utgangspunkt i elevens vurdering av dens ståsted i faget.
6	Etter hver vurderingssituasjon/avsluttet økt. Muntlig eller skriftlig
7	Skjemaer for egenvurdering som elevene fyller ut etter endt arbeid med en tekst. Elevene vurderer hverandres arbeider.
8	Stiller spørsmål og gjennomfører fagsamtaler.
9	Lar elevene sitte i par for å vurdere sin egen og den andres tekst
10	Jeg lar elevene filme seg selv med egen telefon ved framføringer, så kan de (frivillig) melde til meg via læringsplattformen hvordan de syns en presentasjon gikk
11	Elevene vurderer seg selv etter i etterkant av prøver
12	Elevene vurderer alle skriftlige arbeider før de får endelig karakter. Elevene vurderer seg selv på alle mindre arbeider. Det er egne spørsmål til egenvurdering på alt eleven blir vurdert i
13	F.eks med direkte oppfølgingsspørsmål til eleven om hvordan han/hun vurderer egen prestasjon som avslutning/del av en vurderingssituasjon eller, for mer refleksjon, gjennom en senere fagsamtale. Skriftlige former for egenvurdering benyttes selvsagt også.
14	Elevene har ved noen muntlige framlegg muligheten til å bruke de samme vurderingsskjemaene som jeg bruker for å vurdere seg selv. Dette har vist seg å være svært krevende, spesielt på 8. og 9. trinn, men også på 10. trinn. Elevene har problemer med å se seg selv utenifra og objektivt, de tenderer til å legge stor vekt på mindre feil og i stor grad undervurdere seg selv på mange områder mens de overvurderer seg selv på andre. Dette problemet kommer selv om man i felleskap har gjennogått oppgaven, vurderingskriterier og mål detaljert og grundig på forhånd.
15	Ofte som del av en oppgave
16	Jeg integrerer egenvurdering når arbeider gis tilbake.

	- elevene må se på og vurdere egne arbeid før de får karakter.
17	De vurderer hverandre sine presentasjoner og blir dermed trygge på kriteriene. Disse anvendes på eget produkt.
18	Ofte krever jeg at elevene selv besvarer et skjema, enten med generelle spørsmål om deres egen "prestasjon", eller ved at de selv vurderer sin egen prestasjon gjennom å krysse av og kommentere vurderingskriteriene, slik at de får et bevisst forhold til hvordan de vurderes og ut ifra hva. Mange elever bruker ikke vurderingskriteriene vi går gjennom og de får utdelt i forkant like godt som de kanskje kunne ha gjort, men ved å fokusere på egenvurdering tror jeg at elevene selv blir mer bevisste på hvordan de eventuelt kan forbedre seg.
19	Vurdere andres og egne tekster. Høre på foredrag,- krysse av på kompetanseskjema som elevene har utarbeidet på forhånd.
20	At de gir seg selv karakter. At de skal skrive hva de har lært på lapper. At de fyller ut skjema der de vurderer seg selv. Dette er jeg ikke flink nok til å få til.
21	Gir elevene skjema med spørsmål som de svarer på i begynnelsen av en time.
22	I samtale med eleven etter en framføring. Bruker ikke skjema. Spør hvordan eleven syns det gikk. Om du har en elev over flere år vil vi avtale om hva eleven skal forsøke å gjøre bedre neste gang. Med andre ord: dette er en prosess.
23	Det er elevene godt trent i, muntlig eller med "tegn". Bruker lite skr egenvurd. Dermed tar det kortere tid og jeg bruker det ofte
24	De vurderer seg selv
25	Vi gjennomfører kameratvurdering underveis i både skriftlige og muntlige framlegg. Dette fremmer egenvurdering i form av refleksjon. Elevene leverer skriftlig egenvurdering i etterkant av større arbeider.
26	lite
27	Det beste er prosesskriving og gjennomgå kriteriene før start I alle emner. Midt I prosessen får eleven tips som bare gjelder seg. Å forbedre seg på deler av kriteriene motiverer og sannsynliggjør vekst.

Presentation of teachers

Informant number	Age	Education	County of residence	Currently teaching at level:	Teaching experience (years)
1	58	Høyere utdanning/universitet 5 år	Østfold	VG1	35
2	51	Høyere utdanning/universitet 5 år	Aust-Agder	VG1	21
3	39	Høyere utdanning/universitet 5 år	Sør-Trøndelag	VG1	11
4	ca. 50	Høyere utdanning/universitet 5 år	Sør-Trøndelag	VG1	Ca. 20
5	34	Høyere utdanning/universitet 5 år	Nordland	VG1	9
6	50	Høyere utdanning/universitet 5 år	Aust-Agder	VG1	27
7	44	Høyere utdanning/universitet 5 år	Vestfold	VG1	18
8	46	Høyere utdanning/universitet 5	Sør-Trøndelag	VG1	13

		år			
9	64	Høyere utdanning/universitet 5 år	Sør-Trøndelag	VG1	40
10	44	Høyere utdanning/universitet 5 år	Nord-Trøndelag	10.trinn	22
11	50	Høyere utdanning/universitet 5 år	Aust-Agder	VG1	26
12	45	Høyere utdanning/universitet 5 år	Vestfold	10.trinn	18
13	31	Høyere utdanning/universitet 5 år	Møre og Romsdal	VG1	5
14	27 år	Høyere utdanning/universitet 5 år	Finnmark	10.trinn	Siden aug. 2013
15	50	Høyere utdanning/universitet 5 år	Vestfold	VG1	21
16	41	Høyere utdanning/universitet 5 år	Telemark	VG1	12
17	52	Høyere utdanning/universitet 5 år	Akershus	VG1	26
18	27	Allmennlærerutdanning og mastergrad i engelskdidaktikk	Akershus	10.trinn	Dette er mitt første år som lærer.
19	58	Almennlærerutdanning pluss 1.avd spes.ped og 1.avd sos.ped	Finnmark	10.trinn	23
20	42	Høyere utdanning/universitet 5 år	Telemark	VG1	10
21	28	Høyere utdanning/universitet 3 år	Telemark	VG1	1
22	49	Høyere utdanning/universitet 5 år	Rogaland	10.trinn	26
23	41	Høyere utdanning/universitet 5 år	Finnmark	10.trinn	17
24	37	Høyere utdanning/universitet 5 år	Buskerud	10.trinn	10
25	46	Høyere utdanning/universitet 5 år	Buskerud	10.trinn	21
26	45	Høyere utdanning/universitet 5 år	Telemark	VG1	18
27	63	Høyere utdanning/universitet 5 år	Østfold	10.trinn	38

Appendix 2: Outline of the Interview

Spørreundersøkelsen

Bakgrunnsspørsmål:

Hva er din alder?

Hva er din faglige bakgrunn/kompetanse?

Hvor mange år har du vært i læreryrket?

Tema: vurdering av elevenes kompetanse i engelsk

Spørsmål 1

Hvor mange muntlige vurderinger gir du elevene dine i forhold til skriftlige vurderinger i løpet av et skoleår? Hvordan fordeler du disse utover året?

Spørsmål 2

Hvordan underviser du elevene dine i muntlige ferdigheter i engelsk? Hvor ofte?

Spørsmål 3

Hvilke vurderingsformer/situasjoner bruker du når du vurderer elevenes muntlige kompetanse i engelskfaget?

Spørsmål 4

Hvor mange muntlige vurderinger baserer du sluttkarakteren i 10.klasse/VG1 engelsk på?

Spørsmål 5

Når du samler alle muntlige vurderinger du har på en elev før sluttkarakteren skal settes, hvordan vekter du de ulike vurderingsformene? Teller de ulike vurderingsformene like mye? (Eksempel på vurderingsformer/situasjoner: diskusjon i par/grupper, fagsamtale, muntlig presentasjon m.m.)

Spørsmål 6

Hvilke utfordringer møter du på når du vurderer elevenes muntlige kompetanse i engelsk? Hvordan løser du disse?

Tema: vurderingskriterier

Spørsmål 7

Hvilke kriterier legger du til grunn i karaktersettingen? Hvem utarbeider vurderingskriteriene? Bruker du de samme kriteriene flere ganger? Forklar.

Spørsmål 8

Hvis du skal sette opp en prioritert liste, hvilke fem vurderingskriterier vektlegger du mest når du vurderer elevers muntlige kompetanse?

Tema: egenvurdering

Spørsmål 9

Hvordan legger du til rette for egenvurdering?

Kontaktinformasjon

Om du er interessert i å delta i et telefonintervju, vil jeg be deg fylle ut informasjon i tabellen under:

Navn:	
Epost adresse:	
Telefon:	

Appendix 3: Request for participation in interview

Forespørsel om deltakelse i forskningsprosjektet

"Sluttkarakter i engelskfaget"

Bakgrunn og formål

Formålet med denne studien er å finne informasjon om hvordan lærere i engelsk på 10.trinn og VG1 lander på sluttkarakteren i faget engelsk. I 10.klasse gis elevene to separate karakterer i engelsk – en muntlig og en skriftlig. På VG1 derimot, er det kun en karakter som skal beskrive elevenes muntlige og skriftlige kompetanse. Bakgrunnen for dette studiet er egne erfaringer som engelsklærer på 10.trinn og utfordringer knyttet til dette med vurdering som har vekket en interesse hos meg. Dette er et masterprosjekt i engelskdidaktikk ved Universitetet i Oslo, Institutt for lærerutdanning og skoleforskning (ILS) og prosjektet inngår som del av en individuell masteroppgave, med en student og veileder.

Utvalget er basert på hvor i Norge skolene befinner og du forespørres om å delta fordi du underviser i faget engelsk på din arbeidsplass.

Hva innebærer deltakelse i studien?

Deltakelse i studien innebærer at du svar på en spørreundersøkelse som du får tilsendt via epost. Denne spørreundersøkelsen tar ca. 15 minutter å besvare. På spørreundersøkelsen vil du også bli spurt om å delta i et oppfølgende telefonintervju, og dersom du samtykker til dette kan du fylle ut tabellen med kontaktinformasjon. Spørsmålene vil dreie seg om din praksis kring *vurdering* av elevenes kompetanse i faget engelsk. Data vil anonymiseres og registreres i form av notater fra dine svar.

Hva skjer med informasjonen om deg?
Alle personopplysninger vil bli behandlet konfidensielt. Kun student og veileder vil ha tilgang til dine personopplysninger. For å ivareta konfidensialitet til personopplysningene lagres atskilt fra øvrige data slik at ingen andre har tilgang på dette.

Deltakere vil ikke kunne gjenkjennes i publikasjon.

Prosjektet skal etter planen avsluttes i mai 2016. Datamaterialet skal anonymiseres ved prosjektslutt og personopplysninger skal slettes.

Frivillig deltakelse
Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du trekker deg, vil alle opplysninger om deg bli anonymisert

Dersom du ønsker å delta eller har spørsmål til studien, ta kontakt med Alma Cosabic, tlf.: 41211991. Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Samtykke til deltakelse i studien

Jeg har mottatt informasjon om studien, og er villig til å delta

----
(Signert av prosjektdeltaker, dato)

☐ Jeg samtykker til å delta i spørreundersøkelse.
☐ Jeg samtykker til å delta i oppfølgende telefonintervju.