INLAND NORWAY
UNIVERSITY
OF APPLIED SCIENCES

Faculty of Educational Sciences

Stine Lisa Johannessen

Master's Thesis

Oral assessment in the English subject

Teachers' understandings of what to assess

Lektorutdanning i engelsk

2018

Consent to lending by University College Library   YES ☒   NO ☐

Consent to accessibility in digital archive Brage   YES ☒   NO ☐

Acknowledgements

The process of writing this thesis has been exceptionally challenging and exceptionally fun. Now it is complete, and new adventures await. But first, I have people to thank:

I want to thank the teachers taking part in this project. Without you there would be no thesis.

I want to thank my supervisors Heidi Silje Moen and Christina Sandhaug. Heidi, thank you for your impressive attention to structure and your detailed comments. Christina, thank you for your eye-opening comments on my empirical research findings and helpful literature suggestions. I want to thank both of you for answering my emails in late night/early morning moments of despair and for letting me stop by your offices when I did not know what more to write. I could not have done this without you.

I want to thank Henrik Bøhn for his doctoral thesis on which the present thesis is based, and for his suggestions of what to explore. Your work has been invaluable.

I want to thank Knut Øystein Høvik for his immensely helpful comments on the final draft.

I want to thank my family, all eight of you, for bearing with me and for your continuous support. I am lucky to have you.

Cathy, my much older twin and forever partner in crime. What would I have done without you.

And lastly, to my friends, especially: Elisabeth, Torunn, Stefanie, Mari, Matilde, Caroline, and Elise. Without you I would have completed this thesis a long time ago. But it is okay. We had fun. So much fun. Thank you.

Hamar, May 2018

Stine Lisa Johannessen

Table of contents

List of Figures

List of Tables

Abstract

Title: Oral assessment in the English subject: Teachers' understandings of what to assess

Author: Stine Lisa Johannessen

Year: 2018

Pages: 94 (excluding appendices)

Previous research in high-stakes assessment indicate that teachers pay attention to different aspects of students performace in the assessment of oral English, both nationally and internationally (Bøhn, 2016; Iwashita, Brown, McNamara & O'Hagan, 2008). This thesis explores what teachers in the English subject at the upper secondary level identify as important to assess when assessing oral English in the classroom setting. Classroom assessment is in this thesis defined as assessment that may have both formative and summative purposes. This means that assessment in the classroom may be seen as both a tool to enhance learning and a tool to decide on a student's overall achievement. Semi-structured interviews were conducted with three teachers to collect empirical data.

The empirical research findings in this thesis suggest that teachers have similar overall understandings of what to assess, but that there are variation in how they understand the relative importance of more narrow performance aspects. For instance, the teachers in this study held differing opinions of how the relationship between 'language' and 'content' in oral assessment is to be understood. Two of the teachers reported views of 'language' and 'content' being equally important, while the third reported 'content' as more salient to assess. Content, as reported by the teachers, relates to students' abilities to discuss and reflect to a larger degree than being able to convey subject matter knowledge. The teachers further presented relatively similar understandings of the narrower linguistic features, expect from 'fluency', which the teachers reported to understand as both related to 'language' and 'content', as well as being an overarching feature that may affect the overall impression of a student's oral performance.

Norsk sammendrag

Tittel: Muntlig vurdering i engelskfaget: Læreres forståelse av hva som skal vurderes

Forfatter: Stine Lisa Johannessen

År: 2018

Sider: 94 (ikke medregnet vedlegg)

Funn fra tidligere forsking i «high-stakes»-vurdering antyder at lærere legger vekt på ulike aspekter av elevprestasjoner i vurdering av muntlig engelsk (Bøhn, 2016; Iwashita, Brown, McNamara & O'Hagan, 2008). Denne oppgaven undersøker hva lærere i engelskfaget i norsk videregående opplæring identifiserer som viktig å vurdere når de vurderer muntlige ferdigheter i klasserommet. Klasseromsvurdering brukes i denne oppgaven om vurdering som både har formative og summative formål. Dette betyr at vurdering i klasserommet kan sees på både som et verktøy for å fremme læring og som et verktøy for å vurdere en elevs måloppnåelse. Semi-strukturerte intervjuer med tre lærere ble gjennomført for å innhente empirisk materiale.

De empiriske forskningsfunnene i denne oppgaven indikerer at lærerne har relativt lignende forståelser av hva som skal vurderes, men at det er variasjon i hvordan de forstår forholdet mellom snevrere aspekter. Lærerne i denne studien ga blant annet uttrykk for ulike oppfatninger av hvordan forholdet mellom «språk» og «innhold» skal forstås i muntlig vurdering. To av lærerne forsto «språk» og «innhold» som like viktig, mens den tredje forsto «innhold» som viktigere å vurdere. «Innhold», som rapportert av lærerne, handler om elevenes evner til å diskutere og reflektere i større grad enn å kunne formidle faktakunnskap. Videre uttrykte lærerne relativt like forståelser av snevrere lingvistiske trekk utenom «flyt», som lærerne forsto som relatert både til «språk» og «innhold»». Lærerne forsto også «flyt» som en overordnet ferdighet som kan påvirke deres helhetlige inntrykk av en elevs muntlige prestasjon.

1. Introduction

Assessment is an act of communication about what we value[1]

In this thesis I will discuss the complex nature of assessment in Norway and implications for oral English classroom assessment. The nature of education in Norway may be characterized as localized and with a high level of teacher autonomy in what to teach and assess (Eurydice, 2008, p. 31). In addition, teacher-based assessment, which is of high importance in Norwegian upper secondary education[2], is both of formative and summative character.

Studies suggest that teachers find it difficult to operationalize the competence aims in the national curriculum into concrete lesson plans and learning objectives, and that "…there does not seem to be a shared understanding of what constitutes adequate, good, and excellent performance in different subject areas" (Nusche, Earl, Maxwell, & Shewbridge, 2011, p. 129). This raises concerns about a lack of consistency in the assessment of student performances, as teachers might not make adequate judgments to promote learning if the learning objectives and assessment criteria are not clear (Nusche et al., 2011, p. 31). The lack of a common understanding of what constitutes different levels of performance may also result in unfair grading of students, which might impact their opportunities for admission to institutions of higher education. Consequently, I have chosen to focus on exploring what teachers understand as important to assess in oral assessment in the English subject at the upper secondary level in Norway.

---

[1] (Boud, 2000, p. 160).

[2] While I employ American writing conventions, this is the term used for this level in official documents.

1.1 Research aim and purpose

The area of oral skills in educational assessment is underresearched both nationally and internationally, regarding both L1 and L2 language instruction (Svenkerud, Klette, & Hertzberg, 2012, p. 37). However, research specifically regarding the assessment of oral English at the upper secondary level in Norway is found in a recent doctoral thesis by Henrik Bøhn (2016). He found that teachers generally agree on the grading of students' oral English exam performances, but that teachers held different opinions on the importance and relevance of different performance aspects. Bøhn (2016, p. 59) found that teachers disagree on the assessment of more narrow performance features, such as which linguistic features are most important to assess. Teachers were also reported to hold different opinions on the relationship between the assessment of language and content, and some teachers also assessed features of student performances not relevant to the competence aims in the English subject curriculum, such as effort (Bøhn, 2016, p. 33). These findings indicate that there is not a shared understanding of what to assess in the assessment of oral English. This lack of a shared understanding of what to assess prompts a discussion on validity and reliability. Bøhn (2016) therefore suggests that more research should be conducted on how Norwegian L2 teachers assess and grade oral English in the classroom, "…as these make up a substantial proportion of the students' final English grade at the upper secondary school level" (2016, p.71). This thesis is a response to Bøhn's (2016) call for further research on assessment of oral English in a classroom setting.

On the basis of this, the overarching aim of this thesis is to discuss why the Norwegian assessment situation is complex and the challenges involved for teachers. As part of this exploration, I will identify what teachers pay attention to when it comes to oral assessment in the classroom setting, and I will compare their understandings with what the English subject curriculum and other defining documents identify as relevant features to be assessed. In addition, I explore how teachers understand the relationship between specific constructs and criteria identified in the English subject curriculum. 'Construct' as well as the term 'criterion' are further defined and explained in section 1.3. The following research questions will serve to guide this thesis:

What do L2 teachers understand as important to assess in oral English classroom assessment at the upper secondary level in Norway?

- How do L2 teachers understand the relationship between language and content when assessing oral English?

- What are L2 teachers' understandings of the constructs and criteria they consider important to assess?

    o How do L2 teachers understand the linguistic criteria *fluency, grammar, vocabulary, pronunciation,* and *intonation*?

1.2 Thesis structure

The present chapter contextualizes the study by providing a general introduction, defining relevant terminology, discussing official and other relevant documents, as well as presenting a literature review of previous studies relevant to this thesis. Chapter 2 explains the theoretical framework of the thesis, largely based on theories from the fields of educational assessment, applied linguistics and educational theories of learning. These theories provide relevant conceptualizations for understanding what teachers assess, and the consequences of assessment within the context of the classroom. In Chapter 3, the research design is outlined, including the research method, the research questions, the participants, the data collection and the framework for analysis. Here I also consider possible limitations and ethical considerations regarding the empirical research. In Chapter 4, the empirical research findings are presented and analyzed, and in Chapter 5, I discuss the main findings and to which extent these have responded to the research focus and overall aim of this study. Finally, in Chapter 6, I present my concluding remarks.

1.3 Definining terminology

1.3.1 The 'what' to assess: 'construct' and 'criterion'

The question of constructs, or *what* is to be assessed, is essential in language assessment. Fulcher and Davidson (2007, p. 169) state that a 'construct' is an abstract concept that must be defined so that it can be investigated. This means that "it can be operationalized so that it can be measured" (Fulcher & Davidson, 2007, p. 169-170). Constructs usually have a frame of reference (Bachman & Palmer, 2010, p. 213), and the frame of reference in the Norwegian context is national legislation and the national curriculum (LK06) (Bøhn, 2015, p. 1).

However, the operationalization of the constructs is left to the local level (Ministry of Education and Research, 2004, p. 40), which in many cases means the individual teachers.

The teaching and assessment in Norwegian education is to be based on competence aims that make up different subject curricula. The competence aims are broad and general and must be operationalized for both teaching and assessment purposes (Ministry of Education and Research, 2004, p. 40). The Regulations to the Education Act (§3-4) provide general definitions of what constitutes different levels of achievement, such as stating that the grade 4 means "good degree of competence in the subject" (Own translation). However, as the competence aims in the subject curricula are operationalized by teachers locally, teachers' understandings of what the constructs to be assessed are and what constitutes different levels of achievement is of the utmost importance.

'Criterion' is also a term frequently used in literature on language assessment about the 'what' to assess (Broadfoot, 2007; Luoma, 2004; McNamara, 2000), and this must be addressed. 'Criterion' is often used in relation to criterion-referenced assessment, and Glaser and Klaus (1962), explain that this type of assessment "…depend on an absolute standard of quality" (p. 421). However, competence aims cannot be understood as assessment criteria. As competence aims must be operationalized in order to be assessed, this does not correspond with Glaser and Klaus' (1962, p. 421) definition of criterion-referenced assessment as depending on an absolute standard of quality.

Interestingly, as Bøhn also points out (2015, p. 2) Fulcher and Davidson (2007, p. 370) use 'fluency' as an example when referring to a 'construct', which is what Brindley (1991, p. 140) uses as an example of a 'criterion'. More interestingly, Bøhn (2015, p. 2) uses 'construct' when referring to larger concepts such as 'language' and 'content', while Hasselgreen and Ørevik (2018, p. 371) claim that the English subject curriculum is the construct, since students' achievement is based on their overall achievement in the competences that make up the subject curriculum.

To avoid confusion in how to understand 'constructs' and 'criterion' in this thesis, I reserve the use of 'criteria' to mean "…the key aspects of performance" (Brindley, 1991, p. 140). I further follow Bøhn (2015, p. 2) in making a hierarchical distinction to encompass for different levels of operationalizations of the 'what' to assess. In this thesis, I will use 'construct' as an overarching term for the larger concepts, and 'criteria' and 'sub-criteria' for the narrower

aspects of what to assess. To illustrate: "language' is an example of a 'construct', while "grammar" is an example of 'criteria', and lastly "subject-verb concord" is an example of 'sub-criteria'. The purpose of this hierarchical distinction of the terms relates to both how teachers identify the relative importance of what to assess, but also how they report to understand different constructs and criteria, which is relevant for the discussion of empirical research findings. Below is a visual representation of the hierarchical distinction between constructs and criteria:

- Language
  - Grammar
    - Subject-verb concord

Figure 1: Hierarchical distinction of constructs and criteria

1.4 Assessment in Norway

Educational assessment has different purposes in different settings. Since the early 2000s, there has been an upsurge in educational summative assessment practices in the form of national and international large-scale international language testing, which generally have comparative, and diagnostic purposes (Bøhn, 2016; Hopfenbeck, 2014; Kunnan, 2008; Tveit, 2014). At the same time, educational assessment practices with formative purposes have also received increasing attention from researchers, national educational authorities and local school teachers, especially on the role attributed to using assessment in classrooms to promote and enhance learning (Black & Wiliam, 1998a; Sadler, 1989). These different purposes of assessment may be seen as belonging to two different paradigms (Gipps, 1994). Assessment with summative purposes may be seen as having affinities with the 'psychometric paradigm', while formative assessment is more in line with the 'educational assessment paradigm' (Bøhn, 2016; Gipps; 1994). I elaborate on these paradigms and discuss tensions between the two in classroom assessment practices in section 2.1.

1.4.1 Assessment for learning

In Norway, the national policy program known as *Assessment for Learning* (AfL), which draws heavily on the educational assessment paradigm (Bøhn, 2016, p. 4), has been of considerable influence in recent years (Norwegian Directorate for Education and training, 2015, pp. 1-2). The program was initiated by the Norwegian Directorate of Education and Training on behalf of the Ministry of Education, and aims at improving formative learning and assessment practices (Hopfenbeck, Tolo, Florez, & El Masri, 2013, p. 11). The program focuses on how students can use feedback and assessment from teachers and peers, as well as self-assessment to improve their learning, and it emphasizes that feedback should be timely and of high quality. The program was introduced in 2010 as one of several initiated responses to what is known as the "PISA shock" in 2001, where Norway ranked at a mediocre 13th of 30 countries in reading and science. Tveit (2014, p. 221) points out that while achievement studies had revealed similar results in earlier years, it was the PISA ranking that stimulated reform in the Norwegian education system. Other responses were implementation of mapping tests, national tests, and end of school exams (Hopfenbeck et al., 2013, pp. 24-25). The implementation of both formative and summative initiatives to improve competence in Norwegian schools illustrates the importance attributed both learning and measurement in the Norwegian context.

The *Assessment for Learning* program is based on four principles for quality formative assessment, which are outlined in the revised version of the Regulations to the Education Act (Hopfenbeck et al., 2013, pp. 24-26). The four principles emphasize that the purpose of formative classroom assessment is to promote learning, and that students and apprentices learn better when they:

1. *Understand what to learn and what is expected of them*

2. *Obtain feedback that provides information on the quality of their work or performance*

3. *Are given advice on how to improve*

4. *Are involved in their own learning process and in self-assessment*

 (Norwegian Directorate for Education and training, 2015, pp. 1-2).

The main aim of the *Assessment for learning* program has been to integrate these four principles of assessment into everyday teaching practices to improve student learning by making assessment practices more transparent. However, when reviewing formative assessment practices in Norwegian classrooms, an OECD review team (Nusche et al., 2011, p. 56) encountered the view among teachers that formative assessment included a range of small summative tests counting towards a final achievement grade. Teacher-based classroom assessment was used to track student progress and to provide practice for final summative assessments, such as written- or oral exams. Teachers also reported to understand student self-assessment as students grading their own performances, not as reflection on learning. The OECD review team questioned if the conflation of the two purposes of assessment in Norwegian classrooms may reinforce a view among teachers that formative assessment is to be understood as preparation for more summative assessments (Nusche et al., 2011, p. 56).

1.4.2 Classroom assessment

Classroom assessment based on teacher judgment has long been the primary form of assessment in Norway (Nusche et al., 2011, p. 56), and despite the changes in educational assessment practices, it continues to be (Ludvigsen et al., 2015, p. 13). Classroom assessment can include both formative assessment and summative assessment, but these labels represent fundamentally different purposes. Formative assessment "….is the process of identifying aspects of learning as it is developing, using whatever informal and formal process best help that identification, so that learning itself can be enhanced" (Nusche et al., 2011, p. 56). On the other hand, summative assessment "…is used to confirm what students know and can do, to demonstrate whether they have achieved the curriculum outcomes, and occasionally to show how they are placed in relation to others" (Nusche et al., 2011, p. 56). These definitions illustrate that formative and summative assessments have very different purposes, and Gipps (1994, p. 14) argues that any attempt to use formative assessment for summative purposes will impair its formative function.

In Norway, the Regulations to the Education Act provides a distinction between formative and summative assessment. It states in §3-11, that the purpose of formative assessment, or assessment for learning, is to use continuous assessment to help a student improve his or her competence in a subject. The Act states in §3-17, that the purpose of summative assessment, or assessment of learning, is to provide information on the competence of a student at the end

of the instruction in a subject. These definitions clearly state that formative and summative assessment have different purposes in the classroom.

However, while the distinction might seem clear in terms of purpose, in terms of practice the two will necessarily have to be related in order to ensure validity and reliability (see sections 2.5 and 2.6). In a 2009 revision of the Regulations to the Education Act, a new paragraph (§3-16) on the relationship between formative and summative assessment was introduced. Here it states that formative assessment is not only to be understood as continuous classroom assessment to guide instruction and learning, but it is also to be understood as having the function of informing the teacher of a student's competence when deciding on a student's overall achievement grade. The overall achievement grade is characterized as summative assessment in §3-17, but §3-16 in the revised version of the Regulations to the Education Act state that:

> Assessment for learning shall promote learning and give the student the opportunity to improve his or her competence during the subjects teaching period. The competence the student has shown during the teaching period is part of the assessment foundation when determining the subject course grade. (Regulations to the Education Act, § 3-16. Own translation).

The relationship between the two purposes of assessment is complex and has been an ongoing concern. By implementing the national initiative *Assessment for learning*, it is evident that Norwegian educational authorities have seen formative assessment as an important tool to enhance student learning and as a tool to ensure more just and reliable assessment practices in Norwegian schools. However, after introducing §3-16 in the Regulation to the Education Act, which state that formative assessment is part of what the teachers are to base overall achievement grades on at the end of the school year, the line between the two purposes of assessment became blurred. This lack of clarity between the purposes of assessment has also been viewed as a concern by the OECD review team (Nusche et al., 2011, p. 58) who in 2011 advised Norwegian educational authorities and national policy makers to further strengthen the coherence and clarity about the purposes and uses of different assessments in Norwegian classrooms.

However, the importance attributed the summative function of classroom assessment was further emphasized in a 2016 White Paper to the Ministry of Education and Research. The Ministry compared overall achievement grades to examinations, and stated that:

> …it is important to emphasize that the exam- and overall achievement grades are different expressions of competence. An exam grade is based on a more limited assessment foundation than an overall achievement grade …. Persisting discrepancies should be an indicator to school owners and school leaders that it is necessary to change current assessment practices". (Ministry of Education and Research, 2016, p. 63. Own translation).

This situation emphasizes the difficult reality teachers face when they are to assess in the classroom. Dysthe (2009, p. 33) underscores that the overall aim of education is learning, but the national initiative *Assessment for Learning* makes it difficult to separate learning situations from assessment situations. When the Ministry of Education and Research (2016, p. 63) emphasizes that the local level should change their assessment practices if achievement grades do not correlate with examination grades, the summative function of classroom assessment becomes fundamental. The responsibility put on local school teachers to change their practices makes it reasonable to question the autonomy teachers feel like they have in what to teach and assess in the classroom.

However, there have been attempts to provide perspectives on the relationship between the purposes of assessment in classrooms, and David Boud (2000, p. 160) presents a notion of a complimentary relationship between formative and summative assessment purposes. He states that every assessment act that teachers' implement has more than one purpose, and argues that assessment practices should be developed to perform what he refers to as "double duty". This entails focusing both on the performance of the task at hand simultaneously with the long term educational goals. This view of assessment as performing "a double duty" is relevant to classroom assessment in Norway, where assessment is both a tool to enhance learning and a tool to decide on a student's overall achievement. It is thus important that there is a shared understanding of what to assess so that teachers can assess it consistently and thus help students achieve their educational goals. However, having such goals entails having a standard for what constitutes different levels of performance, but there are no such standards in the Norwegian educational context. As students' overall achievement is to be based on their achievement of the competence aims in the English subject curriculum, it is therefore relevant to this study to discuss this main document of reference and identify what students are to be taught and assessed on.

1.5 The English subject curriculum

The English subject curriculum is part of what is known as the Knowledge Promotion (LK06). It is the current national curriculum in Norway, and is the main document teachers must base their teaching and assessment on (Norwegian Directorate for Education and Training, 2013). The national curriculum was introduced in 2006, revised in 2013[3], and covers 10 years of compulsory schooling as well as upper secondary education and training. The curriculum is output-orientated, and students are to be assessed on their attainment of competence aims, which make up the curriculum. In addition, the subject curriculum defines five basic skills, which are common to all subjects in school. 'Oral skills' is defined as of these basic skills (Norwegian Directorate for Education and Training, 2013, pp. 4-5), which emphasizes the importance educational authorities attribute to the development of oral competence. I elaborate on how the definition of 'Oral skills' is relevant to this thesis in section 1.5.2.

It should here be noted that the English subject has a somewhat undefined status in Norwegian education. In 2014, the English subject was established as a national priority and was further defined as one of three core subjects, along with Norwegian and mathematics (Ministry of Education and Research, 2014, p. 11). Bøhn (2016, p. 5) argues that Norwegian educational authorities have given the English subject a special status by no longer subsuming the subject under the label "foreign languages". However, the English subject is not explicitly referred to as a second language either. While English has traditionally been regarded as a foreign language in Norway (Simensen, 2014, p. 1), claims have been made that Norway, among other countries, are in the process of shifting towards an ESL status, due to the increased usage of English (Graddol, 1997, p. 11). Rindal (2014, pp. 1-2) argues that the conflicting beliefs about language will often lead to conflicting opinions about which aspects of language are important to teach and assess. The lack of a clear definition of whether English is to be considered a second or foreign language might affect teachers' understandings of what to be assessed when assessing oral English. I return to in a discussion in section 5.1.

---

[3] The national curriculum is currently under revision. The new curriculum is expected to be implemented in 2020 (Norwegian Directorate of Education and Research, 2017).

1.5.1 Relevant competence aims

The competence aims in the English subject curriculum are structured under four main areas: (i) *Language learning;* (ii) *Oral communication*; (iii) *Written communication*; and (iv) *Culture, society and literature* (Norwegian Directorate for Education and Training, 2013). Out of these four main subject areas, I identify the competence aims under the three main areas: *Language learning*, *Oral communication* and *Culture, society and literature* as relevant to oral assessment. 18 competence aims are identified under these main areas, and it must be acknowledged that it is unreasonable to assume that teachers assess all of these competence aims in every oral assessment situation in the classroom. Rather than presenting all of these aims, I discuss the main areas identified as relevant to oral assessment. The list of relevant competence aims is included in Appendix 1.

The main area *Language Learning* focuses on the development of skills to help students self-assess their own use of language. Students should be able to identify what they need to improve and select appropriate strategies and working methods to further develop their competence in English. In order to improve, metacognition and reflection become central skills students should master (Norwegian Directorate for Education and Training, 2013, p. 10). For example, one of the competence aims state that the aim of the teaching is to enable the students to "evaluate own progress in learning English» (Norwegian Directorate for Education and Training, 2013, p. 10). The importance attributed student self-assessment may be seen as having affinities to the principles of quality assessment in the *Assessment for Learning* program (Norwegian Directorate for Education and training, 2015, pp. 1-2). However, one may question how student achievement in self-assessment is to be assessed. I return to this in section 2.2.

The main subject area *Oral Communication* is a language-specific main area. It is detailed and has a wide scope. The competence aims that make up this area focus on both productive and receptive skills, and students are expected to be able to listen, speak, converse and apply suitable communication strategies to use and understand English (Norwegian Directorate for Education and Training, 2013, p. 10). This involves developing a vocabulary and using idiomatic structures and grammatical patterns in oral communication, as well as being able to speak fluently and coherently suited to the purpose and situation. In addition, it is stated that students should be able to "use patterns for pronunciation, intonation, word inflection and various types of sentences in communication" (Norwegian Directorate for Education and

Training, 2013, p. 10). However, the curriculum does not specify a target variety of English for students to obtain or what constitutes mastery, which might have implications for what teachers understand as salient to assess.

The last main subject area identified as relevant to oral assessment is the main area *Culture, society and literature*. It is made up competence aims relating to subject matter-content, and focuses on cultural understanding in a broad sense (Norwegian Directorate for Education and Training, 2013, p. 11). It covers topics connected to social issues, literature and other cultural expressions, and students are expected to be able to both discuss and elaborate on content matter. For example, one competence aim states that students should be able to "discuss and elaborate on different types of English language literary texts from different parts of the world" (Norwegian Directorate for Education and Training, 2013, p. 11). However, specifically what students are to discuss and elaborate on is not stated. The lack of specific content matter in the subject curriculum is not necessarily problematic, as this allows the local level autonomy in what to teach. However, the general nature of the competence aims opens for a wide range of operationalizations which might make for very different understandings for how these aims are to be operationalized and assessed.

1.5.2 Oral skills as basic skill

As this thesis explores what teachers understand as important to assess in the oral assessment of English, it is purposeful to present and discuss how the subject curriculum defines the basic skill 'Oral skills'. Having oral skills 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 (Norwegian Directorate for Education and Training, 2013, pp. 4-5).

This definition indicates that the purpose of developing 'oral skills' relates to more than being able to produce speech, as students are to develop suitable strategies in order to achieve successful oral communication in a variety of settings and for a wide range of purposes. In other words, students must develop a wide set of skills and abilities, and the combination of

these make up their competence in oral communication. While there is consensus in the field of applied linguistics that 'communicative competence' is an adequate reference point in the assessment of oral language proficiency (see sections 1.6.1 and 2.3), there does not seem to be a shared collective understanding of how one can best assess this complex competence (see section 2.3). One issue relates to how one can define 'competence' in a manner so that it can be assessed:

In the Knowledge Promotion (LK06), 'competence' is defined as: "[T]he ability to solve tasks and master complex challenges. The students show competence in concrete situations by employing knowledge and skills to solve tasks" (Norwegian Directorate for Education and Training, 2016, para. 1). The Directorate further explains that 'competence' cannot always be taught as is, but must be operationalized into smaller properties[4]. The Norwegian Directorate for Education and Training (2016, para. 3) states that:

> …sometimes parts of a competence aim or particular skills or abilities must be taught separately. It might then be appropriate to develop local learning goals. At the same time, the different parts must be reassembled into the competence the student is to develop, and the competence aims must not be divided in such a way that loses perspective of the competence. It is the competence aims the student shall work towards and be assessed in" (2016, para. 3, own translation).

This points to a paradox in the Knowledge Promotion. On the one hand, competence aims are formulated and implemented on a national level, while on the other hand awaits control and responsibility on the local level. While the aims must be formulated in order to be assessed, they must be sufficiently general for local teachers to take ownership through local operationalization (Backmann & Sivesind, 2012, p. 251). This may be experienced as a paradox if the system is understood as based on aims that must unite conflicting interests: they are to be general and possible to assess at the same time. This relates to an issue raised by Hartberg, Dobson and Gran (2012, p. 24), which can be referred to as "the urge to operationalize". If teachers spend too much time breaking the relevant competence aims into smaller fractions with the intent of making it clearer to students what is expected of them, they might lose focus on how to progress to achieve the holistic goal which is the development of

---

[4]The Directorate for Education and Training are currently developing 'core elements' in all subjects as a part of the national curriculum revision. The core elements are to encompass the most important content as well as what students need to learn to master and use the different subjects. This includes subject-specific knowledge areas, methods, terminology, ways of thinking and ways of expression. The core elements are expected to be completed by June 2018 (Norwegian Directorate of Education and Research, 2017).

their communicative competence, which again has implications for what teachers understand as important to assess.

1.6 The Common European Framwork of Reference

The Common European Framework for Reference (Council of Europe, 2001) is not a nationally administered document for reference in Norway. However, it has been influential in Norwegian curriculum development (Ludvigsen et al., 2014, p. 79). While one can identify both language and content-related constructs in the English subject curriculum that are to be assessed (see section 1.5.1), there are no national rating scales or assessment criteria for teachers to employ in their assessment of these. For this reason, I choose to employ the CEFR. The framework provides a detailed description of what constitutes 'communicative language competence', as well as a range of reference level descriptors that describe levels of achievement[5]. As Bøhn (2016, p. 22) does in his doctoral thesis, I employ the reference level descriptors of oral and spoken production (Council of Europe, 2001, pp. 58-74). These are holistic reference descriptors and work well with the many components that make up 'communicative competence', and will further provide a reference against which teachers' understandings of what to assess will be discussed in Chapter 5.

The Common European Framework describes what learners need to know about and do with language in order to use it for successful communication, and has been referred to as "something of an encyclopedia of language learning and use" (Alderson, 2007, p. 661). The framework was introduced in 2001 and is based on theoretical models of communicative competence. These models I return to in section 2.3. The CEFR is most widely known for its language proficiency scales, which consist of descriptors of what learners can do at each of six reference levels (Snow & Katz, 2014, p. 233). The six reference levels (A1, A2, B1, B2, C1, and C2), both with and without their descriptors, have become commonplace in European education at all levels to describe learner language proficiency, and terms like 'beginner' or 'intermediate' have become synonymous, if not replaced, with the reference level labels such as A1 or B1 (Figueras, 2012, p. 479). Interestingly, while Norway is ranked as "very highly proficient" in EF's annual Proficiency Index (Education First, 2017a), and one might assume

---

[5] The Council of Europe introduced new descriptors in 2018 (Council of Europe, 2018).

that this is close to mastery, "very high proficiency" only corresponds to a CEFR level B2 (Education First, 2017b).
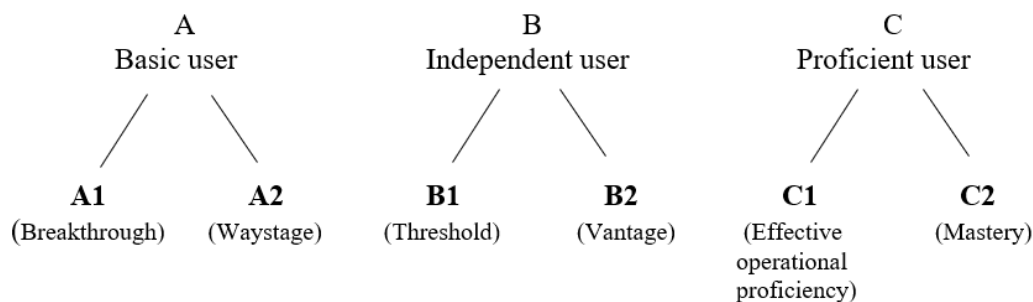
Figure 2: The CEFR's reference levels

1.6.1 Communicative Language Competence

The CEFR lists three basic components and a range of sub-components to describe what constitutes 'communicative language competence'. The three main components of communicative language competence are: (i) linguistic competences, (ii) sociolinguistic competences and (iii) pragmatic competences (Council of Europe, 2001, p. 13). Figure 2 is a modified version of Bøhn's (2016, p. 21) illustration of the many components and sub-components of communicative language competence presented in the CEFR. In addition, the CEFR (Council of Europe, 2001) also list 'fluency' and 'propositional precision' as generic factors which "determine the functional success of the learner/user" (p. 128), and these have been added to the figure to illustrate their relationships to the other components. 'Fluency' is defined in the CEFR (Council of Europe, 2001) as "the ability to articulate, to keep going, and to cope when one lands in a dead end" (p. 128), and 'propositional precision' as "the ability to formulate thoughts and propositions so as to make one's meaning clear" (p. 128).

Beyond these, the CEFR also list additional competences that contribute to a learner's communicative abilities, and refer to these as "knowledge of the world" (p. 101), "sociocultural knowledge" (p. 102) and "intercultural awareness" (p. 103) (Council of Europe, 2001) which one may argue have affinities to the main subject area 'Culture, society and literature' in the English subject curriculum (Norwegian Directorate for Education and Training, 2013, p. 11). In addition to these, the CEFR lists non-verbal communication as something users of the framework need to either recognize or use, and emphasizes that body language as something that "differs from practical actions accompanied by language in that it

carries conventionalized meanings, which may well differ from one culture to another"
(Council of Europe, 2001, pp. 88-90). Such explicit references to the importance of non-verbal
communication are not identified in the competence aims in the English subject curriculum.

COMMUNICATIVE LANGUAGE COMPETENCES

LINGUISTIC COMPETENCES

- *Lexical competence*
- *Grammatical competence*
- *Semantic competence*
- *Phonological competence*
- *Orthographic competence*
- *Orthoepic competence*

SOCIO-LINGUISTIC COMPETENCES

- *Linguistic markers of social conventions*
- *Politeness conventions*
- *Expressions of folk wisdom*
- *Register differences*
- *Dialects and accents*

PRAGMATIC COMPETENCE

- *Discourse competence* (i.e. ability to produce coherent stretches of language in terms of thematic organization, cohesion etc.)
- *Functional competence* (e.g. imparting and seeking information, expressing, and finding out attitudes socializing, communication repair).

FLUENCY AND PROPOSITIONAL PRESICION

Figure 3: The CEFR's model of communicative language competence

1.6.2 Language proficiency scales

Listed below in Table 1 and Table 2 are extracts from the CEFR's reference level scales
describing overall oral production and oral interaction. It should be noted that only the B2 and
C1 levels are included in this thesis. The empirical research findings suggest that teachers pay
more attention to the narrow linguistic performance aspects of higher level students (see
section 4.6), which is why the higher-level descriptors are included in this thesis, as these are
more detailed in this respect. Presenting two reference levels also allows for comparison and
discussions of teachers' understandings of what constitutes different levels of achievement.

The reference level descriptors indicate that a range of features are involved in oral production
and interaction, and the listed descriptors make references to fluency, coherence, idiomaticity,
grammatical control, situational use and level of formality, which point to the linguistic, socio-

linguistic and pragmatic components that constitute the CEFR's 'communicative language competence' (Bøhn, 2016, p. 23; Council of Europe, 2001, p. 13). This corresponds to a fair degree to what has been identified in the English subject curriculum as relevant for oral assessment, and since there are no nationally administered reference tools in the Norwegian context, these descriptor scales are considered suitable to the discussion of what teachers understand as important to assess.

	OVERALL ORAL PRODUCTION (Council of Europe, 2001, p. 58).
C1	Can give clear, detailed descriptions and presentations on complex subjects, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion. Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail.
B2	Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples.

Table 1: The CEFR's production scale for overall oral production

	OVERALL SPOKEN INTERACTION (Council of Europe, 2001, p. 74)
C1	Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language. Can use the language fluently, accurately and effectively on a wide range of general, academic, vocational or leisure topics, marking clearly the relationships between ideas. Can communicate spontaneously with good grammatical control without much sign of having to restrict what he/she wants to say, adopting a level of formality appropriate to the circumstances.
B2	Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with native speakers quite possible without imposing strain on either party. Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments. Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thoughts on more abstract, cultural topics such as films, books, music etc.

Table 2: The CEFR's production scale for overall spoken interaction

These level descriptors are intended to illustrate different levels of communicative language competence. However, they have received criticism. Claims have been made that the reference scales rely on alternating qualifiers like 'some', 'many', 'most' and 'all' to describe different levels of competence, which may reduce differences in the different scale levels to mere

semantic variations (North, 2014, p. 26). This concern may be illustrated by using the level descriptor for Overall Oral Production. The reference level B2 states that a learner "can give clear, detailed descriptions and presentations on a *wide range* [emphasis added] of subjects" (Council of Europe, 2001, p. 58), while a more proficient learner at level C1 "can give clear, detailed descriptions and presentations on *complex* [emphasis added] subjects" (Council of Europe, 2001, p. 58). How one is to understand the relative difference between 'wide range' and 'complex' is not specified by the CEFR, and this is possibly problematic, as it relies on an assumption that teachers have a shared and internalized notion of the level concerned, around which they just norm-reference, and does not take into account that one teacher's "some" may be another's "many".

Another issue is the problematic nature of the difficulty level of the elements put on the same level on the scale. Hulstijn (2007, p. 663) explains that the level descriptors in the CEFR rest on the two closely intertwined pillars: 'quantity' and 'quality'. Quantity refers to what the learner is able to do, while quality refers to how well the learner is able to do it (Hulstijn, 2007, p. 663). This is possibly problematic. One may argue that some students might have a narrow range in terms of quantity, but great depth in terms of linguistic quality; others may have a broad quantity range, but little linguistic quality; and some students' quantity range might match their performance quality.

While it is emphasized that the CEFR is intended as a compendium to profile across categories, not just to give a holistic result, this raises the question of how teachers are to practically employ the framework in oral assessment situations (North, 2007, p. 658). McNamara (2000, p. 20) argues that employing the level descriptor scales for narrower linguistic features would make the assessment more analytical by testing students' competence in discrete points, and not test their actual communicative competence. Another issue is the lack of transparency in how to understand the relative importance of the features that make up the notion of 'communicative language competence'. It is evident that the CEFR provides a framework that is flexible enough to adapt to local contexts, but paradoxically, a very possible issue seems to be that the framework is too flexible and consists of so many features that it might be difficult for teachers to conceptualize what constitutes 'communicative competence'. I return to this in section 5.2., where I discuss different understandings of how to operationalize communicative competence.

1.7 Literature review

In this literature review, I situate this thesis in the oral assessment research context. The choice of literature has been guided by: (i) the focus on what teachers identify as important to assess; (ii) the fact that English subject curriculum specifies that both features of language and subject content is to be assessed; (iii) the lack of a common rating scale and assessment criteria; and (iv) relevant research conducted in the Norwegian educational context. Before reviewing relevant research, four limitations to this literature review must be explained and clarified. First, most of the included studies relate to assessment in speaking proficiency tests, rather than curriculum-based achievement assessments, which is the case in Norway. Second, the studies have involved different age groups and different levels of proficiency. Third, few of the studies on L2 speaking assessment have been conducted in contexts without a common rating scale, and fourth, none of the reviewed studies, including the two conducted in the Norwegian context (Bøhn, 2016; Yildiz, 2011), have researched understandings of what to assess specifically related to the classroom setting, where formative assessment has a central purpose. While these studies are more in line with high-stakes summative assessments, I consider these studies to be highly relevant. Since this thesis explores what teachers understand as important to assess in the classroom setting where both formative and summative assessment purposes are important, I consider research from high-stakes assessment to also be applicable to this thesis. The lack of similar studies of what to assess in oral assessment in a dual-purpose context further emphasize the need for this project to be carried out.

Research suggests that teachers working from the same scoring rubric may very well arrive at similar grades for quite different reasons. David Douglas (1994,) conducted a study where he compared the test scores of six Czech graduate students who had received similar grades, with the transcripts of their oral performances. He analyzed these with various aspects of performance, such as 'precision of fluency', 'content', 'vocabulary', and 'organization', and found that there was little correlation between the test scores and the language produced by the students (Douglas, 1994, p. 134). He speculated that a possible reason for the discrepancy between score and language could be that the teachers conducting the assessment were influenced by other aspects not included in the scoring rubrics.

This view of teachers paying attention to non-relevant constructs in oral assessment is supported by Jenkins and Parra (2003). They investigated the role of non-verbal behavior in

oral proficiency testing, and their participants were four Spanish-speaking and four Chinese-speaking international teaching assistants. Jenkins and Parra (2003, p. 90) found that non-verbal and paralinguistic behavior influenced the raters' perception of the test-taker's oral proficiency. The non-verbal features included among other things: eye contact and positive body language, such as the use of gestures and head nods. Paralinguistic features are referred to as vocal qualities not necessarily associated with linguistic features, such as pitch range, rhythm, articulation, and speed (Jenkins & Parra, 2003, p. 92). They found that test-takers who employed non-verbal behavior considered appropriate to the situation, were rated higher than those who did not. Jenkins and Parra (2003, p. 102) argue that active non-verbal behavior creates involvement between those evaluating and those participating, making the raters' impression of the test-taker's interactive competence affect the rating of verbal features in oral assessments. They further argue that impact of these features in the assessment of communicative competence may have ramifications for students, and that classroom instructors of ESL and EFL should be aware of and trained to understand the role of non-verbal cues in assessments of communicative competence (Jenkins & Parra, 2003, p. 103).

The understanding of the relative contribution of particular features to overall oral proficiency is also emphasized in a large-scale project on speaking proficiency in English as a second language (Iwashita, Brown, McNamara & O'Hagan, 2008). In this project, researchers studied the relationship between detailed features of spoken language produced by students and holistic performance scores awarded by teachers (Iwashita et al., 2008, p. 24). They found that a set of linguistic features seemed to have an impact on the overall assigned score. These were 'vocabulary', 'fluency', 'grammatical accuracy', and 'pronunciation'. Out of these features, the teachers reported that vocabulary and fluency were considered particularly important to assess. However, results also showed that if a test-taker's performance was lacking in one aspect of language, such as vocabulary, then that did not necessarily affect the overall rating of the student's proficiency, indicating that it is a combination of linguistic aspects that determine the assessment of a student's overall proficiency. Findings also revealed that teachers linked pronunciation with intelligibility. If the listener is unable to make out the words, then they are not in a position to evaluate other aspects, such as syntax and content; making pronunciation act as a "first level hurdle" in oral assessment (Iwashita et al., 2008, p. 44).

Research on what teachers identify as important to assess in the Norwegian context has also been identified. In her master's thesis, Lill Mari Yildiz (2011) interviewed 16 teachers at the

upper secondary level, and examined what kinds of performance features teachers pay attention in English oral exams. Yildiz (2011, p. 64) found that although teachers pay attention to a wide range of performance features, she could identify five overarching categories of what teachers regard as important to assess. These categories were: (i) "Language competence"; (ii) "Communicative competence"; (iii) "Subject competence"; (iv) "Ability to reflect and discuss independently"; and (v) "Ability to speak freely and independent of manuscript". In addition, Yildiz (2011, p. 64) found that teachers pay attention to features not directly relevant only to the English subject, such as the ability to show interdisciplinary initiative. There was also great variation in how teachers used and weighed criteria, and Yildiz (2011, p. 95) questioned if these variations threaten the validity and reliability of the oral English exam.

Similar findings are reported in the doctoral thesis by Henrik Bøhn (2016), who investigated what kind of performance aspects teachers' pay attention to in an oral English examination in Norwegian upper secondary school. Bøhn (2016, p. 59) found that his informants mainly focused on two constructs in their assessment of oral English, namely 'Communication' and 'Content'. Bøhn organized the performance aspects hierarchically, and reported that the two main constructs comprised of a number of sub-categories, where 'Linguistic Competence' (belonging to Communication), and 'Application, analysis, reflection' (belonging to Content) turned out to be the most important. Bøhn (2016, p. 59) also found that Linguistic competence consisted of three large sub-categories, namely, 'Grammar', 'Vocabulary' and 'Phonology', and the latter was the most substantial of the three. In general, the teachers included in the study had a common understanding of the constructs and criteria to be tested, but the findings suggest that there was some variation in how the teachers valued the relative importance of these. In particular, Bøhn (2016, p. 59) found that teachers weighed the construct 'Content' differently, and that teachers disagreed on more narrow performance features, specifically those related to 'Phonology'. The results indicated that teachers hold very differing views on the importance of native speaker pronunciation, but that there is a strong sense of agreement that intelligibility is important for a student performance to receive a high score (Bøhn, 2016, p. 62). As for intonation, the teachers presented views of either finding it irrelevant to assess or not as important as other features (Bøhn, 2016, p. 62). Bøhn (2016, p. 59) also found that the teachers were more likely to pay attention to linguistic features in the lower levels of proficiency, but had a stronger focus on content at the higher levels. To sum up, the findings in Bøhn's thesis suggest that teachers generally hold similar conceptions, but that there are variations in how different constructs and criteria are understood. In order to ensure that there

is a shared assessment culture, Bøhn (2016, p. 71) advices Norwegian educational authorities to consider introducing common rating scale guidelines on the national level to strengthen the validity and reliability of oral examinations in the English subject.

1.8 Summary of chapter

This chapter has aimed to situate the assessment of oral English in the Norwegian educational context. First, it is evident that both formative and summative assessment purposes are considered important. However, while the distinction between the two might seem clear in terms of purpose in the Regulations to the Education Act, it is unclear how this distinction is to be understood in practice. Second, the English subject curriculum is the document teachers are to base their teaching and assessment on, and I have identified both content and language-related constructs relevant to oral assessment. Since there are no national assessment criteria or rating scales for teachers to employ when deciding on students' achievement of the competence aims, I have presented and discussed the CEFR's description of 'communicative language competence' and its language proficiency scales for oral spoken interaction and oral production. Employing these as reference tools will allow for a thorough discussion of what teachers understand as important to assess. Lastly, the studies in the literature review indicate that a wide range of constructs and criteria are considered important to assess in oral English assessment, and both the Norwegian studies (Bøhn, 2016; Yildiz, 2011) question how validity and reliability are to be ensured when their findings suggest that there does not seem to be a unified understanding among teachers of what to assess in oral English examinations.

2.   Theoretical framework

In this chapter I discuss the theoretical framework of the thesis. The aim of this thesis is to explore what teachers understand as important in the classroom assessment of oral English. Classroom assessment encompasses conflicting interests which might have implications for what teachers understand as important to assess. This is also why I have deemed it necessary to take on a reflexive approach to theory to provide a satisfactory framework for the discussion of empirical research findings. A reflexive approach to theory entails the involvement of various paradigms and several theoretical ideas for how to understand the object under study, which in this study is what teachers understand as important to assess, rather than employing a "…definite theoretical formulation and a privileged vocabulary for grasping it" (Alvesson, 2003, p. 25).

This chapter may be viewed as divided into three parts. First, I present and discuss the tension between two concurring assessment paradigms and prominent views of learning. How 'learning' is reflected in official documents may be relevant to what teachers understand as important to assess in oral assessment. Second, I discuss relevant theoretical perspectives on the concepts of 'language' and 'content', as these represent the overarching constructs identified in the English subject curriculum as relevant for oral assessment. Lastly, I discuss how the concepts of validity and reliability are of relevance. In a context where assessment has both formative and summative purposes, these concepts are central to just assessment practices, and must as such be elaborated on.

2.1 Assessment paradigms

As stated in section 1.4, summative and formative assessment practices may be viewed as belonging to different paradigms, and I follow Caroline Gipps (1994, p. 5) in referring to these as the *psychometric paradigm* and the *educational assessment paradigm*. Gipps further (1994) explains that "[t]he paradigm within which we work determines what we look for, the way in which we construe what we observe, and how we solve emerging problems" (p. 1). When two paradigms exist side by side, tensions may occur when the purposes of assessment collide or intertwine. Classroom assessment has the purpose of aiding learning, but it also has the purpose of certification through an overall achievement grade awarded each student at the end of the year. It is therefore relevant to present features of these two paradigms and discuss how they are reflected in classroom assessment practices.

The science of psychometrics was developed from work on intelligence testing and had the underlying notion that a student's abilities were fixed and could be measured. Assessment practices in the psychometric paradigm are primarily concerned with test outcomes, not the processes of learning, and it is claimed (Baird, Hopfenbeck, Stobart & Steen-Utheim, 2014, p. 16) that the psychometric tradition has nothing to do with learning at all. In the psychometric paradigm, knowledge has been thought to exist separate from the learner, and is something for the student to obtain, not construct (Serafini, 2000-2001, p. 385). Thus, the main purpose of assessment is to monitor learning. Tests are usually norm-referenced and are designed to produce familiar proportions of low, medium and high scores (Gipps, 1994, p. 5). This means that a student's performance is graded in relation to her peers, making the test performance relative, rather than absolute. This also entails that since a student cannot control the performance of her peers, she cannot control her own grades. In the psychometric tradition, a central purpose of testing is to compare individuals, and thus standardization and reliability of tests are of central importance.

The educational assessment paradigm, where assessment for learning is central, was developed in the late 20th century, and may be seen as a reaction against the psychometric paradigm (Gipps, 1994, p. 1; Throndsen, Hopfenbeck, Lie & Dale, 2009, p. 31). The educational assessment paradigm is different from psychometrics in its views on the interaction between assessment and learning. In this paradigm, assessment is viewed as a contextually specific interpretive activity, and knowledge is believed to be constructed by the individual within a social context (Serafini, 2000-2001, p. 387). There is a preference for

criterion-referenced assessment, which Glaser (1994) states "…provide information to the degree of competence attained by a particular student which is independent of reference to the performance of others" (p. 520). Criterion-referenced assessment is also defined as "performance in relation to established standards or criteria" (Nusche et al., 2011, p. 53). In the Norwegian elementary and secondary educational system, these established standards are the competence aims in the different subject curricula. However, as stated in section 1.4, this is problematic. The competence aims state what the students should be able to do, but they do not provide an absolute standard for what it means to have achieved a level of competence. Also central to assessment in the educational assessment paradigm is its clear formative function, where the aim is not to only indicate a student's current level of achievement, but to support the student's further learning (Gipps, 1994, p. 3).

2.2 Assessment and theories of learning

The national initiative *Assessment for Learning* has been of considerable influence in Norway in the past decade, and as discussed in section 1.4.1, effective assessment practices are considered integral to successful teaching and learning (Black & Wiliam, 1998b, p. 41; Norwegian Directorate for Education and training, 2015, pp. 1-2). In a context where one of the purposes of assessment is to be used as a tool to further support learning, it is reasonable to assume that there should be a degree of alignment between assessment purposes and current theoretical understandings of learning.

However, theories of learning and assessment do not share the same theoretical foundation, and developments and tensions in the field of learning theory have coexisted with theoretical advances in the psychometric paradigm and the educational assessment paradigm, where assessment have fundamentally different purposes (de Corte, 2010, pp. 36-44). I see it relevant to further discuss how prominent learning theories relate to classroom assessment in the Norwegian context. I follow James (2006) in categorizing theories of learning into behaviorist, cognitive-constructivist and sociocultural. As the aim of this thesis is to explore what teachers understand as important to assess in oral classroom assessment, providing a clear understanding of the context within they assess is relevant, as one might assume that how views of learning are reflected in the Norwegian educational context might affect both how and what teachers they consider important to assess.

According to Dysthe (2008, p. 17), classroom assessment practices in Norwegian schools have traditionally been in line with the behaviorist theory of learning, which emerged in the 1930s and had precedence into the 1970s. In behaviorist theory, learning is viewed as being the conditioned response to external stimuli (James, 2006, p. 7), and it is believed that all complex skills can be divided into small fractions, and that learning occurs by accumulating these atomized bits of knowledge. Skinner (1954) explains that that:

> The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step. This solution to the problem of creating a complex repertoire of behavior also solves the problem of maintaining the behavior in strength .... By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the possibly aversive consequences of being wrong are reduced to a minimum (p. 94).

This view of learning is mirrored in how the Knowledge Promotion (LK06) defines 'competence' (see section 1.5.2), and in how the CEFR's reference level descriptors are hierarchically organized (see section 1.6.2). A basic tenet of behaviorist theory is that learning is seen as linear and sequential, and complex understandings can only occur when elemental prerequisite objectives are mastered (Gipps, 1994, p. 19). Thus, analytical discrete-point test formats work well with assessment, which entails the testing of items separately, such as phonology, grammar, or vocabulary without extensive context (Ajideh, 2009, p. 165; Borger, 2014, p. 22). Dysthe (2008, p. 17) argues, the main purpose of assessment in the behaviorist view of learning has generally been to check a student's attainment of factual knowledge in a subject, and to measure *how much* of something a student knows, and the mental states of students have not been of much interest in behaviorist theories to explain learning (Baird et al., 2014, p. 23). The focus on measurement and the deemphasizing of the processes of learning makes behaviorism fit well with summative purposes of assessment that has worked well within the psychometric paradigm (section 2.1).

While behaviorist theories of learning are explicitly uninterested in mental processes to explain learning (Baird et al., 2014, p. 5), cognitive-constructivist theories are not. In this theory of learning, which began its growth in the 1960's (James, 2006, p. 8), knowledge is seen as actively constructed, and it is believed that students make sense of the world by organizing and structuring concepts in relation to previous knowledge and experiences. This makes formative assessment an integral part of cognitive-constructivist theories of learning, and the role of the teacher is thus to facilitate learning by guiding students in their attempt to

assimilate new knowledge with their prior knowledge (James, 2006, p. 9). In order for students to assimilate new knowledge with prior knowledge, they must have access to their own metacognitive processes. This makes self-assessment and reflection integral parts of how to improve learning (Gipps, 1994, p. 28). It is evident that AfL's principles for quality assessment (Norwegian Directorate for Education and training, 2015, pp. 1-2) and the focus on self-assessment and learning strategies in the subject curriculum (Norwegian Directorate for Education and Training, 2013, p. 10) are in keeping with cognitive-constructivist perspectives of learning. While one can hardly question that these are important skills to learn, one can question how these skills can be assessed in valid and reliable manners. How is the teacher to assess a student's achievement in the ability to self-assess?

Sociocultural views of learning are also based on constructivist views of learning, and are heavily influenced by the works of Vygotsky (1978). According to this perspective, learning is not thought of as acquired by a student from a teacher, but constructed in the interaction between the learner and their social environment (Baird et al., 2014, p. 26). Learning is thus seen as a social and collaborative activity. According to this theory, students can create new knowledge and therefore assessment needs to capture this by utilizing a variety of techniques that are closely tied to the learning situation. Sociocultural learning theories have been recognized as more closely related to *Assessment for Learning* than the other learning theories (Dobson & Engh, 2010, p. 38). However, James (2006) claims that assessment practices are weakly conceptualized within this view of learning. Baird et al. (2014) argue that since sociocultural theories of learning understand learning as socially constructed, it opens up for the negotiation of assessment criteria between students and teachers. However, this raises the question of validity and reliability. Are students and teachers to collectively decide on achievement grades? Both James (2006) and Baird et al. (2014) argue that sociocultural theories of learning are currently not sufficient to justify specific assessment practices with summative purposes, and agree that further research is needed to develop valid and reliable approaches to assessment suitable to this view of learning.

Due to the lack of coherence between the individual learning theories and the two assessment purposes, James (2006, p. 12) has suggested that there are possibilities for synthesis, where blending elements from the different theories may provide a more complete and inclusive learning theory for guiding practice of teaching and assessment. The aim of this thesis is not to provide such a theory. However, the problematic relationship between the three learning theories and the purposes of assessment is not easily identified in literature on classroom

assessment, and deserves more attention. The lack of direct correlation between specific learning theories and the different assessment purposes have implications for validity and reliability in classroom assessment situations, as it can be assumed that different views of what 'learning' is and how it should be understood will influence both what teachers assert as important to assess and how they choose to assess in the classroom. I return to this in sections 5.2. and 5.3 in the discussion.

2.3 Theoretical perspectives on communicative competence

Today, the notion of *communicative competence* (see also 1.6.1) is widely accepted as a reference point in the assessment of oral language proficiency. Communicative competence was introduced into applied linguistics as a reaction to grammar-focused theories of language (Luoma, 2004, p. 97), and may be seen as a realization that having perfect spoken linguistic competence does not adequately account for the complex forms of how language is used in social situations.

While there is consensus in both LK06 and the CEFR that 'competence' is to be understood as more than a skill or knowledge (see 1.6.2 and 1.7.1), Bagarić and Djigunović (2007, p. 94) claim that *competence* is one of the most controversial terms in the field of general and applied linguistics. Its definition is often associated with the classical distinction between *competence* and *performance* drawn by Noam Chomsky (1965), where the former is concerned with a monolingual speaker's knowledge of language, and the latter is associated with the use of actual language in concrete social situations (Chomsky, 1965, p. 4).

This definition of competence as something purely linguistic has since been rendered inadequate, and Hymes (1972) developed a broader definition of communicative competence by incorporating a sociolinguistic perspective. Hymes (1972) argues that social knowledge is an equally important aspect of competence, as the speaker must know "…when to speak, when not, and as to what to talk about with whom, when, where, in what manner" (p. 277). Communicative competence is thus to be understood as more than just being able to produce speech. Canale and Swain (1980) explain communicative competence as a merger between grammatical, sociolinguistic and strategic competences. Grammatical competence is understood to "…include knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology" (Canale & Swain, 1980, p. 29). Sociolinguistic

competence is concerned with using language to fulfill communicative functions in social contexts, and strategic competence is understood as the ability to combine the previous two competences in situations of communicative breakdowns (Canale & Swain, 1980, p. 30). Savignon (1972) puts greater emphasis on ability in social contexts in her understanding of communicative competence, and defines it as "the ability to function in a truly communicative setting – that is, in a dynamic exchange in which linguistic competence must adapt itself to the total informational input, both linguistic and paralinguistic, of one or more interlocutors" (p. 8), and equates communicative competence with language proficiency (Bagarić & Djigunović, 2007, p. 96). Bachman (1990) proposes "communicative language ability" as a more appropriate term for communicative competence, which he explains as consisting of "…both knowledge, or competence, and the capacity for implementing, or executing that competence in appropriate, contextualized communicative language use (Bachman, 1990, p. 84). He further develops the notion of 'strategic competence' introduced by Canale and Swain (1980), suggesting that it should not be limited as a compensation strategy, but should be seen as an integral part of communicative language use, performing a function in determining the most effective means of achieving a communicative goal in situational contexts of communication (Bachman, 1990, pp. 100-107), which one must assume is the main aim of communicative language teaching and assessment.

There is an evident development in the understanding of communicative competence, and there seems to be consensus that a competent user of language should not only possess knowledge of a language, but also possess the ability to use this knowledge to communicate and negotiate meaning by interacting meaningfully and accurately in communicative situations. However, despite being widely used in language testing, there are challenges to these theoretical models of communicative competence. Given the complexity of the various models, it has been questioned how teachers can make practical use of them in assessment situations. For instance, McNamara (2000, p. 20) addresses the issue of models of communicative competences being too complex to practically employ, and states that they are not adequate to account for the way the different aspects act upon each other in actual communication. McNamara argues (2000) that paradoxically, "…as models of communicative competence become more analytic, so they take us back to the problems of discrete point testing usually associated with testing the form alone" (p. 20). Harding (2014) is also critical of how to practically employ these models of communicative competence for purposes of assessment, and argues that the solution has been for language developers to rely on

"…frameworks which have been designed to "unpack" existing models of communicative language ability. The Common European Framework of Reference (CEFR) is currently playing this role across many contexts as an accessible de facto theory of communicative language ability" (Harding, 2014, p. 191). However, as argued in the discussion of the CEFR in section 1.6, the issue seems to be that 'communicative competence' consists of so many components that it is difficult to understand how teachers can practically employ the framework in a manner that is fit for assessment.

2.4 Theoretical perspectives on 'content'

'Content' is one of the identified constructs in the English subject curriculum relevant to oral assessment, and may be found in the main subject areas *Language Learning* and *Culture, Society and Literature* (Norwegian Directorate for Education and Training, 2013). The former is concerned with the processes of language learning and emphasizes that metacognition and reflection are abilities students should master, while the latter main area focuses on cultural understanding in a broad sense.

The integration of language and content instruction has been a growing phenomenon in the language field since the 1980s (Met, 1999, para. 1), and different approaches to this form of instruction have been developed in different contexts, but all share the feature that students engage with content using a language other than their native language. Met (1999, para. 6) developed a continuum of language education with language-driven approaches on one extreme and content-driven approaches on the other. In language-driven approaches, content is a useful tool for furthering learning objectives of a language curriculum, but teachers and students are not to be held accountable for content learning outcomes. In content-driven approaches, content learning outcomes are the main purpose of instruction, and acquisition of language plays a secondary role. The English subject curriculum for GSP1/VSP2 may be placed somewhere in the middle of the two extremes, as both language and content constructs are to be taught and assessed (Bøhn, 2016, p. 24) (see also section 1.5). The figure below illustrates that the English subject curriculum for GSP1/VSP2 may be placed in the center of Met's (1999) continuum of language education.
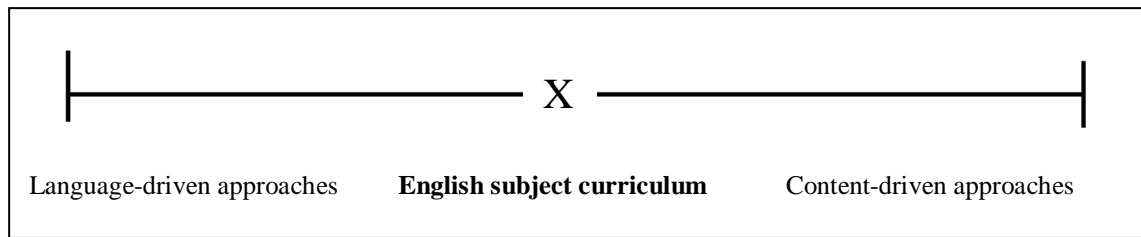
Figure 4: Met's continuum (1999)

As Bøhn (2016, p. 24) also remarks, there is very limited theoretical support to be found for the analysis of content in oral assessment. Literature on language assessment has primarily centered on the use of language assessment for making inferences of students' communicative language abilities, such as language knowledge and strategic competence, rather than content knowledge (Snow & Katz, 2014, p. 231). In the models of communicative competence, content is given little attention. Bachman and Palmer (1996) refer to 'content' as "topical knowledge" (p. 65), and claim that it "…needs to be considered in a description of language use because it provides the information that enables them to use language with reference to the world in which they live, and hence is involved in all language use" (p. 65). However, Bachman and Palmer (1996) do not elaborate on how topical knowledge is to be understood, or how it is to be operationalized to be assessed.

As discussed in 1.6.1, the CEFR (Council of Europe, 2001) describes 'content' as "sociocultural knowledge" (p. 102), "intercultural awareness" (p. 103), and "knowledge of the world" (p. 101), but does not elaborate on how these are to be understood. An explanation for the lack of specific content definition can be found with North (2004), who states that the CEFR "…doesn't try to define what should be taught (content specifications), let alone state how it should be taught (methodology). Content specifications differ according to the target language and the context of the learning; methodology varies with pedagogic culture" (p. 1). Despite this, the lack of a content-specific definition in the CEFR makes it problematic to conceptualize 'content' in teaching and assessment in the Norwegian context. Like the CEFR, the English subject curriculum is made to be adaptable to local contexts, allowing teachers relative autonomy in what specific content-matter to teach and assess (Eurydice, 2008, p. 31).

However, theoretical support for the assessment of 'content' as identified in the English subject curriculum may be found in Anderson and Krahtwohl's (2001) revision of Bloom's Taxonomy (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956), and Bøhn (2016) also employs

this framework in his doctoral thesis. Their revised Taxonomy is a framework for classifying educational objectives, and was developed with the intention of improving educational practices by influencing teachers to ensure that instruction and assessment are aligned with educational objectives (Anderson & Krathwohl, 2001, p. xxii). According to Anderson and Krathwohl (2001, p. 4), a statement of a learning objective contains a noun and a verb, where the noun generally describes the knowledge students are expected to learn, while the verb generally describes the intended cognitive process. Based on this way of constructing learning objectives Anderson and Krathwohl (2001) developed a two-dimensional framework, where one dimension relates to *Knowledge* and the other to *Cognitive processes*. The Knowledge dimension relates to subject matter content and is divided into four general types of knowledge: factual-, conceptual-, procedural-, and metacognitive knowledge (Anderson & Krathwohl, 2001, p. 27). This dimension may be related to the main subject area *Culture, Society and Literature* in the subject curriculum (see section 1.6). The Cognitive process dimension is divided hierarchically from simple to complex in the following six categories: remember, understand, apply, analyze, evaluate and create (Anderson & Krathwohl, 2001, p. 30), and may be related to the main subject area *Language Learning* (see section 1.6).

As the framework is two-dimensional, and educational objectives thus are represented in two dimensions, Anderson and Krathwohl (2001, p. 27) constructed a taxonomy table, where the table's vertical axis forms the knowledge dimension and the horizontal forms the cognitive processes dimension. The taxonomy assumes that learning at higher levels is dependent on having attained prerequisite knowledge and skills at the lower level in both axes (Murtonen, Gruber & Lehtinen, 2017, p. 116), and the intersections of the dimensions form the cells where educational objects are to be classified. The notion that learning at higher levels is dependent on the attainment of prerequisite knowledge mirrors the behavioristic view of learning which I discussed in section 2.2., and I will return to this in 5.3.

The construction of learning objectives with a noun and a verb applies to the Norwegian context, where competence aims are formulated in such a manner (see section 1.5.1). This may be illustrated with the following competence aim from the main area *Culture, society and literature*, where it is stated that a student should be able to "*evaluate* and *use* suitable listening and speaking strategies adapted for the purpose and the situation» (Norwegian Directorate for Education and Training, 2013, p. 11). By employing the taxonomy table in Table 3 to illustrate, one may place the verb 'evaluate' in grid 5 with the same name, while the verb 'use' may be placed in grid 3 under "Apply". The Knowledge dimension of the competence aim refers to

'listening and speaking strategies adapted for the purpose and situation', and may be placed under "D: Metacognitive Knowledge". I further discuss this way of understanding 'content' in oral assessment in section 5.1.3.

The Knowledge Dimension	The Cognitive Process Dimension					
	1. Remember	2. Understand	3. Apply	4. Analyze	5. Evaluate	6. Create
A. Factual Knowledge						
B. Conceptual Knowledge						
C. Procedural Knowledge						
D. Metacognitive Knowledge			X		X	

Table 3: Placement of aim in Anderson and Krathwohl's (2001) taxonomy table.

2.5 Validity in classroom assessment

Validity is regarded as a central concern in language assessment, and commonly refers to the quality or 'soundness' of an assessment procedure (Luoma, 2004, p. 184). Ruch (1924) is credited for presenting the most influential definition of 'validity', and stated that: "By *validity* is meant the degree to which a test or examination measures what it purports to measure" (p. 13). According to this classic definition, 'validity' is viewed as a property of the test itself. However, this definition has since been regarded inadequate, and is rejected by prominent researchers within the field of educational measurement (Fulcher & Davidson, 2007; Moss, 2003; Newton, 2012). The argument for rejecting this definition, as reported by Newton (2012, p. 3), is that no matter how well-developed a test is, the measurement quality would differ if the test is administered incorrectly or applied to a context for which it was unintended.

The current consensus definition of validity is presented within the *Standards for Educational and Psychological Testing*: "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests … It is the interpretations of test scores for proposed uses that are evaluated, not the test itself" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education) [AERA, APA, & NCME], 2014). This view advocates an

understanding of validity as not a property of a test, but inferences that are made from assessment results (Bøhn, 2016, p. 14). Despite this consensus definition being widely recognized, the applicability of validity theory to the classroom setting has been questioned (Moss, 2003; Newton, 2012), as it does not take into account that assessments also have a formative purpose.

The notion of score meaning presented in the *Standards for Educational and Psychological Testing* (2014) is problematic in the Norwegian context. As stated in section 1.4.1, the Regulations to the Education Act (§3-4) state that the grade 4 means "good degree of competence in the subject" (own translation). However, one may ask: "in relation to what?" Gordon Stobart (2012, p. 237) argues that one of the threats to the validity of an assessment situation is the lack of a clear understanding of what reaching the intended standard will involve. The main issue in the Norwegian context is that there is no standard administered for assessment other than the competence aims in the subject curriculum and the general descriptions of level achievement in the Regulations to the Education Act (§3-4). How these are to be operationalized, and thus be understood, is the responsibility of local teachers, which I discussed in detail in section 1.5. Sandvik (2013) contextualize the concept of validity to classroom assessment in the Norwegian context and explains this as:

> Validity in assessment is thus not only about the validity of test results; it is also about the chain of interpretations and possible misinterpretations that may occur when the competence aims in the curriculum are to be operationalized for teaching and learning purposes. (p. 41. Own translation).

For oral English classroom assessment to be of high validity, there needs to be a common understanding of what the required standard both for what constitutes specific grades and how to understand the competence aims in the subject curriculum. Any failure to achieve such a clarity will be a threat to valid assessments of both formative and summative purposes. Sadler (2010, p. 546) argues that in order to be able to evaluate competence we need to know what the required standard is. It is therefore crucial for sound assessment practices that there is a common understanding of what is to be assessed in oral English assessment, and such common understanding also ensures reliability.

2.6 Reliability in classroom assessment

Reliability is also a central concern in language assessment, and can be defined as the consistency of measurement, where the idea is that a test taker should receive the same score if a test is taken several times within a reasonable period of time (Fulcher & Davidson, 2007, p. 375). However, Sari Luoma (2004, p. ix) argues that speaking is the most difficult language skill for teachers to assess reliably. A student's speaking ability is usually assessed in real-time face-to-face interaction, and the teacher must make instantaneous judgements of both what is being said and how it is being said, while it is being said. A student might also perform very differently in an oral presentation in front of an audience than when speaking in a two-party conversation, and might also produce better spoken language if the topic being spoken about is one that the student is familiar and comfortable with.

It is generally agreed upon that no test can be considered perfectly reliable, especially those regarding the assessment of oral skills (Luoma, 2004, p. x). Thus, the most important for a test is then for variations in scores to be caused by differences that are relevant to what is being tested, not factors that are irrelevant, "…such as who did the scoring, the particular selection of items used for the test, and whether the student was having a 'good' or 'bad' day" (Black and Wiliam, 2012, p. 244). This notion of reliability is often used in relation to assessments that only have summative purposes (Black and William, 2012, p. 260), and the issue of reliability has not generally been considered an area of concern in classroom assessment, which also has formative purposes, as it is problematic conceptually.

The *Assessment for Learning* approach, which has been very prominent in the Norwegian school system, is mainly concerned with promoting student learning, and has not been intended to be used for accountability or certification purposes (Black, Harrison, Marshall, Wiliam & Lee, 2003, p. 2). Black and Wiliam (2012) argue that formative assessment is "reliable to the extent that the assessment process being used generate evidence that consistently lead to better, or better founded decisions" (p. 260). Central to the assessment for learning approach is that teachers *want* their students to score differently in their next assessment situation. Classroom assessment is to be used as a tool to enhance instruction and learning, but it is also used to inform the teacher of a student's accomplishments when deciding on a course grade, which illustrates the complicated nature of reliability in classroom assessment.

In Norwegian upper secondary education, there is a strong reliance on grades as a mechanism for making decisions for higher education admissions. It is the sum of course grades in different subjects that make up the majority of grades used for admission to higher education institutions, where admission generally is norm-referenced (The Norwegian Universities and Colleges Admission Service, 2013). When classroom assessment results are used for deciding on course grades that are used for purposes of selection, then one can assume that such assessment results are critical to students. This makes reliability a crucial aspect of classroom assessment, and should be given serious attention. Reliability in classroom assessment in the Norwegian context might thus benefit from a reconceptualization. Stiggins (1997) suggests that a course grade should "sample student performance in a representative manner with sufficient depth to permit confident conclusions about proficiency" (p. 23), and Smith (2003) argues that reliability for classroom assessment would benefit from being renamed to "sufficiency of information" (p. 30). The purpose of reliability in oral classroom assessment is thus for the teacher to decide if enough information has been collected on oral competence so that a reasonable decision of the student's abilities in the tested domain can be made (Smith, 2003, p. 30).

2.7 Summary of chapter

This chapter has aimed to provide a framework for the discussion of the empirical research findings. I have discussed how the tension between the formative and summative assessment purposes stem from conflicting paradigms, and I further argue that there is no direct correlation between specific learning theories and current assessment practices. This might make teachers draw from several perspectives in oral assessment procedures, which is not necessarily problematic, but I argue that how learning is reflected in official documents might influence teachers in their assessment practices and consequently in what they understand as important to assess. I further discuss this in section 5.3.

I have further discussed the development of 'communicative competence', and while there today is consensus that oral language use is a complex activity, the same complexity might also lead to the fragmenting of the complex competences to allow for more analytical assessment practices. The presented theoretical perspectives on 'content' suggest that one may place the English subject curriculum in the middle of a continuum of language-driven and content-driven language learning approaches. One may also divide 'content' as presented in

the English subject curriculum into a 'knowledge dimension' and 'cognitive processes', where the former relates more to the subject area *Communication, society and literature,* and the latter to the subject area *Language Learning*. I discuss this in detail in section 5.1.

The concepts of validity and reliability have also been presented and discussed, and it is suggested that the conflation of formative and summative assessment purposes, as well as the lack of a common clear understanding of what constitutes different levels of performance in assessment in Norwegian classrooms might affect the validity and reliability of oral assessments in the English subject, if there is not a common understanding of what to assess. I will return to this in a discussion in section 5.4.

3. Method

3.1 Research focus

The studies in the literature review in section 1.7 indicate that teachers understand a wide range of constructs and criteria as relevant in the assessment of oral English. While two of these studies were conducted in the Norwegian context (Bøhn, 2016; Yildiz, 2011), neither of these explored what teachers understand as important to assess in the classroom setting. This emphasizes the importance of this study being carried out.

In this chapter I describe the chosen research design and explain how it contributed to answering the research questions (see section 1.1). I also outline the phases of the research process, and I explain how the collected material was analyzed. In addition, ethical aspects and possible limitations to the research process are discussed in relation to the study's research focus.

3.2 The phases of the research process

Several research designs and methodological approaches to collecting data on what teachers understand as important to assess in oral assessment were considered in the planning phase of this project. Explaining the different phases is beneficial for understanding the choice of how I decided to carry out the data collection, and thus deserves further explanation. Initially a mixed-methods approach, employing both interviews and questionnaires as research instruments was considered most suitable, as it was thought that employing only one method could lead to unnecessary fragmentation in the approach to answering the research questions. The questionnaire was thought to collect data on the research question on *what* teachers regard as important to assess, and the interview was intended to collect data to help answer the research questions on teachers' *understandings* of what they deem important to assess.

It should here be noted that I am not employing the terms 'qualitative' and 'quantitative' as opposing binaries when presenting the logic of the research design of this thesis and the tested methodological approaches. With the full knowledge that questionnaire as a research instrument is often regarded as a 'quantitative' instrument, and interview is often considered a 'qualitative' instrument, I will refer to them simply as 'questionnaire' and 'interview'. By using the terms 'quantitative' and 'qualitative', "paradigmatic characteristics" of the research

approaches might be implied (Kvale, 2007, p. 46), which is not my intention. Educational research is a complex field, and in the core of all research lies personal judgements about research questions, samples and methods of analysis - regardless of how the data is to be collected. Based on this view, the research process of this thesis has followed in the footsteps of those advocating the use and trial of whatever methodological tools required to answer the research questions under study (Gorard, 2012).

3.2.1 Pilot phase

As the initial research strategy involved a mixed methods design, a pilot questionnaire was conducted with seven L2 teachers to test the suitability of this research instrument as a means to answering the research question on what teachers understand as important to assess. The pilot was modeled on a questionnaire from Bøhn's doctoral thesis on oral English examinations (2016, p. 98). However, several changes had to be made. Bøhn's questionnaire was based on teachers' rating practices in oral examinations, and as this project's questionnaire aimed to collect data on what teachers understand as important to assess in the classroom setting, some questions were altered, and some were added to fit with this thesis' research focus. This is because the exam format in Bøhn's study (2016) differs from classroom assessment, as the exam format in his study consisted of a preplanned student presentation followed by a discussion between the student and the examiners.

However, in classroom assessment, teachers are free to employ whatever practices they want, and it may be assumed that this might affect what they consider important to assess. Thus, the questionnaire had to be altered to include different assessment practices that might be employed in classrooms. In addition to writing a list of questions with the purpose of eliciting teacher orientations on constructs and criteria they regard as important to assess, space was created for teachers to explain how they would describe the different constructs and criteria included in the questionnaire. Space was also added for the respondents to provide feedback on the relevance of the included items, and they were further encouraged to add items they thought should be included. The questionnaire was written in Norwegian, as it was assumed that the majority of the teachers speak Norwegian as their L1, and would be more inclined to answer without hesitation if the questions are formulated in their first language.

The main finding from the pilot questionnaire concerned the usefulness of this as a research instrument. The yielded data in the questionnaire provided information on teachers'

'evaluations' of constructs and criteria to be assessed, rather than what they themselves regard as important to assess and what their understandings of constructs and criteria were. Despite allocating space for the respondents to provide their reflections, little information on reflections were collected. Another important finding concerned the validity of the collected data from the questionnaire. It must be acknowledged that the included questions were developed by the researcher, and it must thus be assumed that they to some extent reflect my understanding and personal experiences with how oral assessment may be carried out in the classroom, as well as assumptions of what teachers might evaluate and understand as important constructs. This implies that my personal understanding and bias might be too leading when eliciting what teachers identify as important to assess, which would compromise the validity of the collected data. Thus, employing a questionnaire as a research instrument was therefore considered unsuitable as a means to collecting data on what teachers understand as important to assess. However, the yielded data and experiences with potential researcher bias in this pilot proved fruitful when writing questions for an interview guide, and is thus considered an important phase in the project's research process, and one which deserves attention.

Based on the responses in the questionnaire, I developed an interview guide to test whether interview as a research instrument could yield responses relevant to the research focus. As researcher bias proved to be a concern in the pilot questionnaire, I decided to conduct individual pilot interviews with two teachers to test the suitability of this instrument. I characterize the pilot interview guide as semi-structured with an introspective focus. This type of interview allows the respondents to express their understandings by answering open-ended questions, but additional questions may be asked by the researcher when probing is deemed necessary to elicit information useful for the research focus (Kvale, 2007, p. 65). The collected data from the two pilot interviews elicited information on both what teachers identify as important to assess as well as their understandings of specific constructs and criteria. Thus, I concluded that employing a semi-structured interview guide as the main research instrument would be appropriate for data collection. The participants in the pilot provided feedback on the phrasing of the questions in the interview guide, which was taken into consideration, and in the following section is an outline of the revised interview guide employed in the main phase of collecting data. The interview guide in Norwegian may be found in Appendix 2.

3.2.2 Interview guide as research instrument

The interview guide was written with an outline of topics with subsequent questions to be covered. The guide was written in Norwegian. The initial questions related to the participants' background, such as their age, their L1, an outline of their education, and their teaching experience. In addition to being relevant data to collect, starting an interview with questions that are relatively easy to answer might help make the interview situation more comfortable for the participants. Kvale (2007, p. 57) argues that interview questions must be devoid of academic language, which one may claim that 'construct' is. Thus, the term 'construct' was not employed in the interview guide and was avoided in the conducted interviews. Words like 'criteria', 'skills', 'competences', and 'abilities' were also avoided, as it was assumed that these might carry different meanings to the different individuals taking part in the interviews, and was thought to potentially be threatening to the validity of the results. Thus, the interviewees were asked 'what' they consider important to assess in English oral assessment.

Kvale (2007, p. 58) further argues that the *what* questions should be asked and answered before the *how* questions. In this research project, the *what*-questions relate to *which* constructs and criteria teachers understand as important to assess, and the *how*-questions relate to their understandings of the constructs and criteria. As it became evident in the pilot phase that researcher assumptions of what to assess might be too leading, the initial questions were broad and open. These were followed by sub-questions meant to guide the researcher if the participants did not touch on these topics themselves. This allowed the participants to answer without much researcher interference, but simultaneously allowing the researcher to probe if deemed necessary.

All three teachers were asked the general 'what' they consider important to assess, before being asked more specific questions, unless they touched on these themselves. All three teachers were asked to answer what they consider important to assess when assessing 'language', 'content', and the relationship between these. Based on their responses to these questions, I asked more analytical questions. As one of the research questions is on how teachers understand specific linguistic criteria, probing was considered necessary unless the teachers elaborated on these themselves. I return to this as a possible threat to validity in section 3.4. Below is an example of how questions were asked based on the structure in the interview guide. The questions are translated from Norwegian to English:

- *What* do you consider important to assess when you are assessing a student in an oral assessment situation?

  (One teacher answered 'language' and 'content', and thus, the following question was:)

  o *What* do you assess when you assess **language?**

     (One teacher mentioned, among other things, that he looks for 'fluency'. Thus, the following question was:)

     • *How* do you understand **fluency?**

     • *How* important is **fluency** when you assess oral English*?*

3.2.3 Participants and procedures

After completing the interview guide, participants had to be recruited. In order to establish contact with teachers to participate in interviews, I sent an email with information on the project and what participating would entail to the heads of language departments at several upper secondary schools in eastern Norway. Their contact information was identified on the schools' websites. When contact with schools was established, the email was forwarded by the heads of the language departments to teachers who teach English on the GSP1 program, who then were encouraged to get in contact if they would like to participate as interview objects. The project information included in the email may be found in Appendix 3.

For reasons of validity, I only invited formally qualified teachers with at least 60 credits in English and formal teacher training (PPU) to participate in this study. Research in Norway (Langseth, 2013, p. 119) points to teachers' formal competence in English being central to how they interpret and operationalize the aims in the subject curriculum. Teachers operationalize the competence aims based on their education and personal experiences, and Langseth (2013, p. 119) reported that in 2012 58% of English teachers in upper secondary had formal qualifications for teaching English.

Three formally qualified English teachers for upper secondary education from three different schools in two different counties in eastern Norway reached out and were recruited as

interview participants. The following are brief introductions of the three interviewees. Pseudonyms are employed for anonymity purposes, and other details that might threaten the anonymity of the informants have not been included.

Name	**Hedvig**
L1	Norwegian
Formal qualifications	60 credits in English
Teaching experience	20+ years at the same upper secondary school

Name	**Nora**
L1	Norwegian
Formal qualifications	Bachelor's degree in English Master's degree in literature PPU[6]
Teaching experience	Less than five years as an upper secondary teacher

Name	**Brand**
L1	Norwegian
Formal qualifications	Bachelor's degree in English. PPU
Teaching experience	Lower secondary teacher Less than five as upper secondary teacher

Figure 4: Presentation of participants

The three teacher interviews were conducted at locations that were convenient to the participants, which for two of the interviews entailed traveling to the schools where the teachers are employed. The teachers were given the choice of either being interviewed in Norwegian or English, but all three stated that they preferred the interviews to be conducted in Norwegian. Before starting the interviews, the teachers were presented with a formalized

[6] 1 year post graduate teacher training

document with information on the project, and were asked if they would agree to signing a consent form at the end of the interview. The teachers were also informed that the project's research focus and interview guide had been approved by The Norwegian Centre for Research Data (NSD) (see Appendix 4). All three teachers agreed to signing the consent form and also complied to having the interviews audio-recorded for transcription purposes. The interviews were scheduled to last approximately 30 minutes based on experiences from the pilot interviews. However, the three interviews ended up being considerably different in length, the shortest lasting for 22 minutes and the longest lasting over 50 minutes. I discuss this as a possible limitation in section 3.4.

3.3 Framework for data analysis

After completing the interviews, the collected interview data had to be analyzed. Alvesson (2003) advocates a reflexive pragmatist view on interviews, and like Bøhn (2016), I found this approach to be appropriate. Reflexive pragmatism calls for epistemological awareness, acknowledging the uncertainty of empirical material and knowledge claims (Alvesson, 2003, p. 25). This approach is favorable for two main reasons: (1) it avoids the belief that data simply reveals reality, and (2) it appreciates the potential richness of meaning in the empirical material (Alvesson, 2003, p. 14). Regarding interviewing as a technique for tapping interview participants on their knowledge about their experiences neglects the social encounter between the interviewer and interviewee as both a socially and linguistically complex situation. Taking on a reflexive pragmatist approach "…means challenging and reconsidering assumptions and beliefs of what data is all about" (Alvesson, 2003, p. 26). The aim of this research project is not to uncover "true" knowledge of teachers' understandings, but knowledge that is "useful" (Brinkmann & Kvale, 2015, p. 65) for both identification of and further discussions of teachers' understandings of the constructs and criteria to assess. This may lead to research results that are multiple in character, which is not to be understood as problematic, as I consider that a more rigorous approach might have meant that I would have to sacrifice relevant material.

3.3.1 Transcription

The audio-recorded data from the interviews was transcribed within a week after completing the individual interviews, and notes on each interview encounter were written immediately

after each interview. Since the interviews were conducted in Norwegian, they were also transcribed in Norwegian. Excerpts relevant to producing the final report were translated to English. I recognize the interview situations as social encounters, and the notes were taken to supplement the transcriptions, which inevitably lose data from the original encounters. Contextual factors, such as the visual and non-verbal aspects of the interview are not visible in written transcriptions, and the notes were taken to inform on the immediate impression and thoughts of each interview situation to reduce the overall data loss.

I chose to not employ a standard transcription convention when transcribing the interviews, as it was assumed that applying such a convention could affect the data negatively. I acknowledge that transcribed data is already interpreted data, and thus I chose an approach which was intuitive and mostly verbatim, including as many details as possible. The transcriptions also attempted to capture the informal style of the participant responses. Roberts (1997) explains that: "…all transcription is representation, and there is no natural or objective way in which talk can be written" (p. 168). As it was the same researcher conducting, transcribing, and analyzing the interview data, it was considered most appropriate to take on an approach that made intuitive sense to the researcher, and thus retaining more data for analysis than if following a more rigorous convention for transcription. In order to balance accuracy and readability, making choices of what to include and what to exclude from the audio-recordings in the transcriptions was not considered at this stage, as I could not know all of what might turn out to be important information for analysis in this phase of the research process. A sample from the transcriptions may be found in Appendix 5.

3.3.2 Coding of data

The approach to coding the transcribed interview data may be characterized as *inductive thematic analysis* (Braun & Clarke, 2006, p. 83). This type of analysis is not wedded to any pre-existing theoretical framework (Braun & Clarke, 2006, p. 81), and thus works well with the reflexive pragmatist approach. Inductive thematic analysis is the process of coding data without predetermined themes, and may therefore be considered a data-driven approach. However, possible problematic conceptions of this approach must be addressed. I recognize that allowing themes to develop from the teachers' statements in the interviews "…can be misinterpreted to mean that themes 'reside' in the data, and if we just look hard enough they will 'emerge' like Venus on the half shell" (Ely, Vinz, Downing and Anzul, 1997, p. 208). Such a view of data analysis would deprive the researcher of being seen as an active participant

in the identification and selection of themes, and I acknowledge that "if themes 'reside' anywhere, they reside in our heads from our thinking about our data and creating links as we understand them" (Ely et al., 1997, p. 208). Thus, researcher judgement of what a theme is and how it should be understood, is necessary. Braun and Clarke (2006) explain that "…the 'keyness' of a theme is not necessarily dependent on quantifiable measures – but rather on whether it captures something important in relation to the overall research question" (p. 82). The research aim of this thesis is to explore what teachers understand as important to assess in the assessment of oral English. To answer this, I chose to employ Braun and Clarke's (2006, pp. 86-93) six-step guide of doing thematic analysis, and the following is an overview of the six phases. It must be emphasized that the process of analyzing the interview data was not as linear as this six-step guide suggest, but rather recursive, with movement back and forth throughout the different phases of analyzing whenever deemed necessary. In Appendix 6 I have included an excerpt of how the transcribed data were coded, specifically steps 2 through 5.

Phase	Description of the process
1. Familiarizing yourself with your data:	Transcribing data, reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Table 4: Phases of thematic analysis (Braun & Clarke, 2006)

3.4 Possible limitations and ethical considerations

The empirical research is based on the oral responses from three teachers, and the choice of only involving three informants in this study deserves further explanation. It is not the intention of this study to prove that something is 'true', or to present results that are to be representative to the wider population of teachers. The interview participants in this study only represent themselves. However, it should also here be noted that the analysis of the empirical research findings from the three interviews indicate that even though the participants hold many similar conceptions of what to assess, findings also suggest that there are variations in how they understand specific criteria and the relationship between language and content. Based on these findings, three participant interviews were considered enough to support a thorough discussion of the importance of a common shared understanding of what to assess.

I did not reach out to specific teachers when attempting to recruit participants, which I consider advantageous. Allowing teachers to get in contact if they wanted to participate, was thought to minimize the issue of lack of informed consent. One might also assume that teachers choosing to participate have a genuine interest in the research focus. Two of the teachers expressed that they decided to participate in the study because they saw it as an opportunity for both reflecting on and developing their assessment practices. This is not considered problematic, but it speaks against generalization, and therefore deserves mention.

A possible limitation to the study is the difference in how the teachers expressed themselves in the interview situations. One teacher was very economical in her use of language, only giving fairly short answers relevant to the questions asked, while the two other teachers gave lengthy responses, sometimes getting off track of the asked questions and referring back to previous questions. When writing up the results chapter, it proved difficult to extract relevant information without being too selective in what to include or exclude. This was specifically problematic when selecting examples from the lengthy interview responses, as comprising the responses to present the gist of the given answer might give an imprecise impression of what the teacher was trying to convey.

Another possible limitation to the collected data concerns the probing of respondents when asking them what they understand as important to assess. This is especially relevant to one of the teachers, who answered in short sentences, and did not elaborate unless asked to do so. This concern also regards the questions asked about how the teachers understand the linguistic

features *fluency, grammar, vocabulary, pronunciation* and *intonation*. All three teachers mentioned some of these when asked what they assess when they assess 'language', but none of the teachers reported to assess all of these without probing. However, when asked how they understood the different linguistic features and how important they considered these features to be, all three teachers had an opinion of how to understand and the relative importance of all five features. To illustrate, none of the teachers reported to assess 'intonation', without probing, but when asked how they understood this linguistic feature, all three teachers reported that they understand 'intonation' as related to 'accent', and that this was something they might assess in some cases. I present this as a possible limitation because the teachers might have answered that they consider these linguistic features when they assess oral English because they thought it was 'right' to do so. However, I do not consider asking the teachers how they understand the different linguistic features to threaten the validity of the results. It is included as a research question, and all of these features are mentioned under the subject area *Oral communication* in the English subject curriculum (Norwegian Directorate for Education and Training, 2013, p. 10). Even though the teachers might not assess all of these linguistic criteria in every oral assessment situation, they are to assess their students in these at some point in the school year, which seems to suggest that the teachers should hold opinions of how these criteria are to be understood.

4.   Empirical research findings

This chapter presents the empirical research findings of this study. The chapter is structured after the interview guide, and the constructs and criteria teachers identify as important to assess in oral assessment and how they understand these are presented hierarchically and thematically. The teachers' understandings of the specific linguistic criteria are presented last. The interview guide and samples of the transcripts of the three interviews are included in Appendices 2 and 5, but stretches of text in the transcriptions that might identify the informants and other individuals have been excluded for anonymity purposes.

4.1 What teachers identify as important to assess

The teachers were encouraged to speak freely and elaborate on their views when asked the open initial question of 'what' they consider important to assess in oral English classroom assessment. The three teachers presented similar views on what to assess, but in varying detail. Hedvig provides the most holistic approach to the overall question of oral assessment, while Nora and Brand report assessment practices of greater analytical character. Hedvig considers 'communicative competence' as the most essential construct to assess, and gave the following concise explanation in the interview of what she expects from her students in oral assessment situations:

> That they communicate and that they are able to express what they want to express. That is what I consider to be the most important. That they can express themselves; that they can communicate. So that I can understand them.

Nora presents a similar initial view of what to assess, but states that the overall impression is based on many different factors, which are both language and content-related. Nora highlights confidence in topical knowledge as important to the overall impression, and explains that it is usually an indicator that students are not very confident if they read a lot from their scripts in oral presentation situations. In terms of linguistic features and communicative competence, Nora reported that:

> I look at pronunciation, grammar, and vocabulary, of those things that are purely linguistic. And I also look at those things that are relevant in all languages, or generally for communication, such as not speaking too fast and speaking loud and clear. So yes, there are many factors involved in creating an overall impression.

Brand presents the most detailed initial account of what he assesses in oral English classroom assessment, and reports to assess features related to language and content, but also features relevant specifically to oral presentations. Brand also reports communication to be essential in oral assessment, but explains that the 'what' he assesses is often specific to what they have been working with in class, and he makes assessment criteria specific to that.

> So, if we've been working with grammar, such as subject-verb agreement, then that is something I include in the criteria. The assessment criteria have up until now been pretty general, including things like "few grammar mistakes", but with a little lenience in what is considered a little or a lot of mistakes, but yes, grammar mistakes that don't disrupt communication.

In terms of linguistic criteria, Brand reports grammar, fluency, pronunciation, and vocabulary. He also reports that he assesses paralinguistic features, and lists presentation skills, body language, and eye contact. Brand explains that when he assesses content, he does not only assess holistically in relation to the English subject curriculum, but holistically in relation to other school subjects as well. Brand explains that:

> I try not to micro-manage too much, and those who choose to implement things from other subjects get credit for that, because then they sort of show that there is a relationship between subjects, that they think cross-curricular, and that is positive.

These findings indicate that while the teachers consider 'communicative competence' to be the most salient construct to assess, there is evidence indicating that they operationalize this construct to varying detail, as well as having different understandings of how broad the construct of 'communicative competence' is to be understood.

4.1.1 Formal and informal assessment

When asked 'what' they consider important to assess, all three teachers make it clear that they consider there to be a distinction between formal and informal oral assessment in the classroom, and this must be further elaborated on. All three teachers agree that individual oral presentations in front of the class is what mainly constitutes oral English assessment, and that this is how they mainly assess oral English. In such formal assessment situations, all three teachers report that they employ assessment criteria, and that the students are made familiar with these ahead of the assessment situation. The three teachers also report that they sometimes assess oral English informally in the classroom, but report different opinions of what to assess in such situations. All three teachers report that informal oral assessment mainly has a

formative purpose, but Brand reports that he considers oral participation in class to affect the end-of-term grading of the individual students. These findings point to teachers mainly assessing their students' oral competence in a summative manner. Nora explains her distinction between formal and informal oral classroom assessment as:

> There are formal assessment situations where you have the students prepare a presentation on a subject …. But then there's also oral activity in class, which is very positive because it is a way for the student to show competence.

Nora reports that she does not assess linguistic criteria, such as pronunciation, in informal oral classroom assessment, and emphasizes that the main objective in these situations is to have the students participate orally in class. Attending to different features in informal and formal assessment invites validity issues, which is further discussed in section 5.3. Nora further explains that in informal oral assessment, such as asking students questions, she sees this as an opportunity to assess the students' listening skills. Nora states that if students participate orally in class, then they show that they have understood what she was asking them about.

Hedvig has a similar view of the situational distinction between formal and informal classroom assessment, but reports that she might assess linguistic criteria in informal assessment situations if these affect the student's ability to communicate clearly. Hedvig reports that

> I assess in assessment situations, like presentations and such. But also when I walk around in the classroom listening to them reading in pairs, for example. I might correct them while they're reading aloud, but I make sure not to disrupt too much if they're making themselves understood. If I teach smaller groups, I might have them read aloud to improve their reading skills, and then I will ask if they want me to correct them while they read or afterwards.

Brand also makes a distinction between formal and informal assessment, and reports that oral presentations in class is what he considers to constitute formal oral assessment. In informal classroom assessment, Brand reports that he takes notes on answers and general oral activity in class to create what he refers to as a "linguistic impression" of the individual students. However, he explains that the distinction between formal and informal assessment is not clear cut, and that a student's oral participation in informal class situations might influence his expectations of how the student will perform in a formal assessment situation where the student is orally presenting in front of class.

> I might have an impression of students' activity in the classroom, based on their reflective abilities, linguistic skills, and confidence when their participating orally in class, but then they might fall short

when holding a presentation …. I feel like oral proficiency is a little hard to assess, if I'm being honest. Especially in those cases where I might expect more of one student and less of another, and then they surprise me both positively and negatively. Being able to be objective in these cases is demanding, and sometimes I think it might be beneficial to audiotape presentations, to hear what they *actually* presented.

Brand finds it challenging to balance being both the trainer and the judge at the same time when assessing oral English. The former needs to draw on his knowledge of individual students to help them progress, while the latter must, as Brand puts it: "…be more objective" and decide on a student's current level. Brand questioning the trustworthiness of his assessment practices relates to the central issue of this thesis, and complicated relationship between formative and summative assessment is further discussed in section 5.2.

4.1.2 Assessment criteria and grading

The three teachers all reported that they employ assessment criteria when they assess oral English, and this deserves further attention. All three teachers report that they make assessment criteria themselves, and both Hedvig and Nora report that their students are involved in the process of making the criteria. Hedvig makes assessment criteria before each formal assessment situation, while Nora reports that she usually employs the same criteria throughout the year. Brand reports that he makes the criteria without the students, but that the criteria are presented to the students before formal oral assessment situations. The three teachers also report that they employ different types of assessment criteria. Hedvig employs a list of criteria to assess, while Brand employs a very detailed scoring rubric. Nora may employ either a list or a scoring rubric, but this depends on the level of the students and when the assessment takes place in the school year.

Hedvig explains that she uses assessment criteria when assessing oral English, and that she does not grade presentations as often as she used to, because she considers grading to affect the formative purpose of assessments. She also explains that she does no longer employ a scoring rubric, and explains that this is because:

> We used to work with low-, medium-, and high achievement, but they just identify that as 1-2, 3-4 and 5-6.

Hedvig reports that they now work with what she refers to as "checks and upward-pointed arrows" when she is assessing oral English. Checks for what the students do well, and arrows for what they can work to improve. Hedvig explains the reason for this as: "it's so that the

students won't identify with a grade, but rather focus on the feedback". She now conducts four to five assessments per year, where two are without grades, but with written feedback. Hedvig explains that the assessment criteria-lists with "checks and arrows" are developed with the students and that they chose competence aims from the subject curriculum together in class and make oral assessment criteria based on these.

Nora explains that she employs assessment criteria for the assessment of oral English, but she does not make a scoring rubric with low-, medium, - and high achievement in what she refers to as "lower level groups". In such groups she discusses relevant assessment criteria with her students and they are encouraged to come up with criteria for oral presentations and they make a list together. Nora explains that they usually include language features, speaking with adequate volume, and not being too dependent on written notes. Nora explains that the assessment criteria are usually the same throughout the school year, but at the end of the spring term, she usually makes a "more advanced scoring rubric". Nora usually grades her students' formal oral presentations.

Brand explains that he employs assessment criteria in the form of a scoring rubric, and states that he aims to make his assessment criteria specific to the tasks given in class, in order for there to be correlation between the teaching and the assessment situations. Brand expresses that making oral assessment criteria is challenging, and refers to his colleagues who have expressed a wish for there to be national assessment criteria to make this task easier on the individual teachers. Brand reports that he disagrees with this, and explains that he thinks each group is different, and that it is important to take into account that students are at very different levels, making it up to each teacher to make assessment criteria specifically for the level of each class. Brand reports that he considers student self-assessment an important part of assessing oral English, and has his students assess themselves after formal oral situations. He does this by conversing with the students individually, where both parties explain their impression of how the student did in the oral assessment situation. However, Brand reports to have the final say in the grading of the performance.

4.2 'Language' as construct

When asked the question of how they assess 'language' in oral assessment situations, the teachers all present views that they consider students' overall communicative competence to be more important to assess than more narrow linguistic performance aspects. However, all

three teachers reported that a student's mastery of specific linguistic criteria will affect how they grade the students, specifically those aiming at grades 5 and 6. Hedvig and Nora also report that if a student communicates with a simple vocabulary, or makes many grammar mistakes, then that is likely to affect the student's performance grade. Hedvig reported that:

> It might be that the student uses a fairly simple language, but I still manage to get a good grasp of what the student is trying to convey. But that might make the grade a little poorer, or a lower level of achievement …. If they conjugate a verb incorrectly, I wouldn't consider that to be a major discrepancy, for example …. They might very well make a mistake once or twice, and that's not a problem, but if they repeatedly make the same mistake, then that's an indicator that that's something we have to work on.

Nora presents a similar view, but reports that out of the three linguistic criteria she considers important to assess, vocabulary is most important. Like Hedvig, Nora considers the range of linguistic mistakes to affect her assessment of a student's oral performance.

> I think I consider 'vocabulary' to be most important. In terms of 'pronunciation', I think it's okay to have a few mistakes, and when it comes to 'grammar', I think it's important that it doesn't affect communication. Of course, if your vocabulary is limited, then that might affect communication, but if you make a few mistakes with "is" and "are", then that doesn't really affect communication. But of course, if there are a lot of mistakes, then that might have a negative effect, both in terms of pronunciation and grammar.

Brand reports that he assesses five linguistic criteria in oral presentations, and lists grammar, vocabulary, fluency, pronunciation, and volume. Like Hedvig and Nora, Brand emphasizes that communicative competence is what he considers important to assess when he assesses language, but that students' competence in narrower linguistic criteria will affect how he assesses, or grades, their oral performances.

> I have one bullet point on grammar mistakes, but I try not to put too much emphasis on that. But if you want to achieve a 6, I feel like you should be making relatively few mistakes when you're speaking, or be able to self-correct.

These findings suggest that all three teachers operationalize 'language' into linguistic criteria, but these findings do not indicate that there is a shared unified understanding of how many criteria the construct language consists of, or which of these are most salient to assess.

4.3 'Content' as construct

The three teachers gave relatively different accounts of how important they consider the construct of 'content' to be in oral assessment situations. This might indicate that the teachers have different understandings of the relationship between 'language' and 'content' in the English subject. All three teachers report that they assess their students' ability to reflect, discuss and argue in oral assessment situations. Hedvig states that:

> One thing is being able to express yourself, but another thing is what you express. The students might include examples they consider important to present so that can argue their opinions … and they should also be able to discuss and argue for what they choose to present.

Nora reports that she considers it most important that her students make the content material their own in oral presentation situations. She explains that many of the students copy and paste information from the internet, and that it is better if they use simpler words than what they just copied from the source material. Nora further explains that the ability to reflect on subject matter content is an ability she expects her students to develop throughout the school year.

> If you are working towards a 4, 5, or 6, then the content, when the task allows for it, should include reflections and analysis. Of course, in the first presentation my class had on making a travel route in the US, there is not much room for reflection, but in later presentations they're expected to have a more advanced content, and this should not only be repetitions of source material they found on the internet, but something they must have thought out themselves.

Brand explains 'content' as more important than 'language' in oral assessment, and emphasizes that he views reflection as the most salient criteria of the content-construct. Brand reports that he assesses his students in their ability reflect and argue their opinion when they discuss in the classroom, but in formal oral presentations, he is more concerned with the students being able to explain why they have chosen to include certain information in the presentation, and assesses his students' ability to reflect on this.

> For me it's a lot about reflection. If we have worked with a topic, then that's the content relevant to the assessment. If they are able to reflect on that and maybe other things related to that, then that is just "thumbs up".

These findings indicate that the teachers understand 'content' as the ability to reflect to a larger degree than the presentation of subject-matter content. This will be further discussed in section 5.1.3.

4.4 Relationship between language and content

The teachers were asked how they understand the relationship between 'language' and 'content' in oral assessment, and whether they assess these constructs separately. All three teachers report that they assess language and content separately, but that they grade holistically. Hedvig reports that she chooses to assess language and content separately to make it easier for the students to understand what they need to work with.

> I might write "structure", "language" and "content" as assessment criteria and comment that I think they had relevant points, but that the language wasn't quite relevant to what they wanted to express. Content is what you want to say, but language is how you say it. A student might have good intentions about what he wants to say, but his language is too simple. The vocabulary might be too simple to express something in a way to get high achievement.

Nora explains that she does not grade language and content separately, but that she might write feedback on a student performance where she explains that a student's pronunciation and vocabulary is at 5 level, but that the content needs to be improved. Nora further reports that most of the time it is the overall impression of language and content together that is the basis for the oral assessment, and that the goal is always to find the competence that the students have, and not look for the competence the students do not have.

Brand explains that he assesses language and content separately in formal oral assessment situations, and this is because they belong in different rubrics in his scoring rubric. He emphasizes that he considers 'content' to be the most important to assess in oral assessment. Brand has not had any oral assessment situations where he has only assessed linguistic criteria, and reports that he does not think he will in the future either, as he does not see the relevance of assessing language separately from content. In informal oral assessment in the classroom, Brand sees language and content as intertwining, but expresses that his focus is on the content, or reflective abilities, of the students.

4.5 Attending to non-relevant constructs

The three teachers assess, but do not necessarily grade, performance aspects that are not relevant to the competence aims in the English subject curriculum. The teachers all report that they assess other features, such body language, or like Hedvig explains:

> No, I do not assess it, but I comment on it. For the benefit of the student. If they constantly turn away from the audience …then people will not catch what they are saying … which is exactly what I consider to be important.

This notion of body language as relevant for successful communication is supported by Nora who reports that she will not give her students any less credit because of their body language, unless it disrupts communication. Brand also expressed the importance of body language, and especially eye contact, as essential for successful communication. He emphasized that making judgements about body language only was relevant for oral presentations, and that he has included "body language" as an assessment criterion under the overarching heading 'language' in the scoring rubric he employs. Brand explains that he considers body language to affect other performance aspects in assessment situations. He states that:

> If your notes are imprecise, then you depend too much on looking at your script …. Which will affect the overall impression of the oral presentation.

It is evident that the teachers consider non-verbal communication to be important in oral assessment, and body language appropriate to the situation is seen as an essential component of the students' communicative competence. This is further discussed in section 5.1.5.

4.6 Teachers' understandings of linguistic criteria

The teachers reported in varying detail when asked what they consider important to assess in oral English assessment. After presenting their initial accounts of what they assess, the three teachers were asked how they understand different linguistic criteria: *fluency, grammar, vocabulary, pronunciation,* and *intonation*.

4.6.1 Fluency

When asked how they would describe 'fluency', the three teachers reported that they consider it to be a very important assessment criterion, but describe it differently and relate it to several performance aspects. The teachers do not consider 'fluency' only to be a linguistic criterion, but as a feature that also relates to coherence in content and presentation skills in oral assessment.

As a linguistic criterion, Hedvig explains that 'fluency' in assessment situations means that there are "no stuttering" and "that there are no stops". Brand reports that he links fluency to

vocabulary, and that many students do not yet have a vocabulary that allows them to speak fluently. He also describes "pace" as a sub-criterion of fluency, and explains that is important that students speak with a pace that is natural for them in assessment situations. Both Nora and Hedvig report fluency in oral assessment as also being able to circumvent if students lack specific vocabulary.

All three teachers also report that they understand fluency as 'flow'. Hedvig links fluency to 'transitions' in oral presentations, suggesting that fluency is not considered only to be a linguistic criterion, but one that relates to content and structure in presentations as well. Nora links 'fluency' to a student's confidence, and describes fluency as being able to present content material in oral assessment situations without much hesitation. Brand describes fluency as an overarching feature in assessment situations, and explains that fluency also relates to content structure in oral presentations. Brand views 'fluency' as being a linking and overarching feature in oral assessment, and has a major influence on the overall impression of a student's communicative competence.

> I mean, fluency is part of a complex combination of features, which makes it have a pretty big impact on the overall impression, in my opinion.

While all three teachers emphasize that they consider 'fluency' to important criterion in oral assessment, they hold very different conceptions of how it is to be understood. It is evident that the teachers consider 'fluency' to be an elusive notion, and I return to this in section 5.1.2.

4.6.2 Grammar

When asked how they would describe the assessment of 'grammar' in oral assessment, all three teachers state that they do not consider this linguistic criterion to be of major importance for the overall assessment. They emphasize that communicative competence is the main construct they assess, but that grammatical mistakes will affect their grading of a student's oral English. This suggests that 'grammatical accuracy' is feature of some salience. Hedvig explains the following of the assessment of grammar in oral assessment:

> As long as they don't make too many mistakes, I don't consider that to be the worst thing. As long as the students are able to express what they want to communicate, then that indicates that their level of proficiency is adequate for using English not only in school, but also later in life. Of course, in order to achieve the top grades there's a limit to how many grammar mistakes you can make, and also which types of grammar mistakes these are.

Nora explains that she puts less emphasis on grammar in oral assessment situations than in written assessment, and reports that grammar in oral assessment only matters if it negatively affects communication. Frequency and types of mistakes is also a relevant factor, and Nora states that:

> Typical mistakes are "is" and "are", and if the student makes those mistakes a few times, but everything else is great, then I think that that shouldn't stop you from getting a 5 or a 6.

Brand also reports that he does not consider perfect grammatically correct language to be of major importance in oral assessment. He states that he often makes grammatical mistakes in English, and questions his right to expect something of his students that he himself is unable to do. In terms of the grading of oral English, Brand explains that to achieve a grade 6, a student should make relatively few grammatical mistakes in oral presentations, but it does not matter as much to the overall picture of a student's communicative competence if they are graded at level 2, 3 or 4.

4.6.3 Vocabulary

The three teachers consider vocabulary to be the most salient linguistic criterion to assess in oral English assessment, and report that the vocabulary the students use must be suitable to the communicative situation. Hedvig explains that the students' vocabulary is often what separates the top grades in assessment situations. She does not necessarily expect her students to employ a very advanced vocabulary, but that they adapt their vocabulary to the assessment situation, emphasizing that the language the students use should sound "natural".

Nora also reports that using a language that is suited to the situation is important in oral assessment, and further reports a correlation between vocabulary and grading of oral English.

> If I think of different levels of achievement, then I think that in order to get a 2, you at least have to employ a vocabulary that communicates. Some students might use a vocabulary that on the surface sounds very advanced, but then that is too advanced and complicated because they use words that are completely wrong for how the word is supposed to be used, like you have many different synonyms for 'to get', by not all are suitable for all situations.

Brand refers to a competence aim in the English subject curriculum when describing his understanding of vocabulary in oral assessment. The curriculum states that students should be able to employ a "wide general vocabulary and an academic vocabulary related to her education program" (Norwegian Directorate for Education and Training, 2013, p. 10). Brand

reports that he makes a distinction between high frequency and low frequency vocabulary, where he understands high frequency words as relating to a wide general vocabulary, and low frequency words as relating to academically related vocabulary. In oral assessment situations, Brand reports that he assesses his students' use of both low- and high frequency words in oral assessment.

4.6.4 Pronunciation

The teachers report similar understandings of 'pronunciation' in oral assessment. All three teachers present the understanding of 'pronunciation' as both 'accent' and 'intelligibility', but all three teachers report that they do not consider it very important for the students to have the pronunciation of a native speaker of English in oral assessment situations. However, both Hedvig and Brand admit that native-speaker pronunciation will likely affects their overall impression of a student's communicative competence in oral assessment. Hedvig emphasizes that not all non-native speakers of English will be able to acquire the pronunciation of a native-speaker, and Nora reports that it is not an aim of hers that the students acquire a pronunciation that sounds American or British. Nora explain that there are many different accents in her classroom, where some students have thick local eastern Norwegian accents, others have more general Norwegian-sounding accents, and some have foreign accents. She explains that:

> We have speakers of foreign languages who can have really thick accents from their first languages - that are not Norwegian, right, so I really try to not emphasize that.

Nora and Brand both report that even though the will not assess accents, they will assess specific sub-criteria of pronunciation in oral assessment situations. Nora refers to the sub-criteria she assesses as "common mistakes Norwegian students often make". She lists the th-sound /θ/, and "how they stress words", such as "it is *vege*tables not vege*tables*". Brand assesses pronunciation if it affects other linguistic criteria, and mentions fluency and vocabulary as features pronunciation might interfere with. Brand emphasizes that:

> It is my personal philosophy that it should be possible to obtain a 6 without a perfect British or American accent.

The relatively little importance attributed 'pronunciation', and especially native-speaker-like pronunciation may be related to the undefined status of English as a school subject in Norway, which I return to in section 5.1.4.

4.6.5 Intonation

The teachers explain intonation as a linguistic criterion they might assess in some cases, especially if a student is aiming for a high grade. Hedvig explains that she will tell her students to "flatten their English", rather than speaking with the intonation of their Norwegian dialect, which she states many might do. Hedvig states that:

> It is purposeful for the students to understand that they can't raise their voice at the end of every sentence, because then everything will sound like a question.

Nora states that 'intonation' is not a criterion she considers to be very important, because most students will be able to communicate well with an intonation "that isn't quite right". However, Nora reports she might consider intonation a criterion to assess of those students aiming for the grades 5 and 6.

Brand reports that his understanding of intonation is "how you emphasize the sounds in the different words", but that your choice of pronunciation pattern affects the intonation you use. He further states that intonation is not something he considers to be very important, as long as it does not affect 'fluency' or flow in language, such as changing patterns of pronunciation while speaking. Brand explains that being consistent is key.

4.7 Summary of empirical research findings

To summarize the empirical research findings, I consider there to be three main findings that should be further discussed in the following chapter. First, the findings indicate that the teachers hold relatively similar conceptions of what to assess in oral English assessment, but that there are variations in how they understand the relative importance and relationship between specific constructs and criteria. The teachers all report to understand 'communicative competence' to be most important to assess, but report to operationalize this overarching construct differently. Second, the findings suggest that the teachers tend to introduce other and more criteria when assessing the higher-level students than the lower-level ones. Finally, the findings suggest that the assessments the teachers report to conduct, mainly in the form of oral presentations accompanied by pre-set assessment criteria, may be characterized as mini-summative with formative intentions which further has implications for the validity and reliability of assessment. These findings will be further discussed in the following chapter.

5. Discussion

As stated in section 1.1, this thesis is about the complex nature of assessment in Norway and the consequences for oral classroom assessment. To shed light on this I have interviewed three upper secondary teachers to collect data on what they understand as important to assess in oral English assessment in the classroom setting at the upper secondary level. The teachers were asked to present their general understandings of what to assess, how they understand the relative importance of what they assess, and they were also asked to give their reflections on how to understand specific linguistic criteria. As stated in section 1.2, little research has been conducted on the assessment of oral English in the Norwegian context, especially in the classroom setting, and the purpose this thesis is to contribute to the field by providing empirical findings on these aspects.

In this final chapter I discuss the main findings. As summarized in the previous section (section 4.7), the teachers present fairly similar views of what to assess, but they understand the relative importance and relationship between specific constructs and criteria differently, as well as reporting to attend to constructs not identified in the subject curriculum. I will in the present chapter discuss the most central implications of this. For the purposes of a tidy discussion, I divide this chapter into two parts. First, I synthesize and discuss the empirical findings with official and other relevant documents, previous research, and the presented theoretical perspectives on language and content. Second, I discuss larger concepts, and attend to issues connected to the operationalizing 'communicative competence', the possibly problematic nature of the presence of several theories of learning in assessment practices, and lastly, I address issues of validity and reliability of attending to different criteria when assessing different students in a setting where formative and summative purposes are conflated.

5.1 Teachers' understandings of what to assess

5.1.1 Language-related understandings

Like reported in Bøhn's doctoral thesis (2016), there seems to be consensus among the teachers in this study that 'language' and 'content' are the most important constructs to assess in oral English assessment, and that these two constructs make up their students' 'communicative competence'. 'Language', as reported by the teachers, comprises a range of criteria (see section 1.3.1 for terminology), and findings show that 'vocabulary' and 'fluency'

are considered to be especially salient, which corresponds with results in the studies by Iwashita et al. (2008) and Bøhn (2016). The teachers further reported to attending to the linguistic criterion 'pronunciation', and present understandings of this as both 'accent' and 'intelligibility', where the latter was regarded as more important to assess than the former. This finding is also supported by the results reported by Iwashita (2008) and Bøhn (2016) in their respective studies.

Interestingly, this study's informants acknowledged that native speaker-like accents might affect their grading of oral presentations, and further reported to mainly attending to narrower linguistic criteria when assessing higher-level students. This does not correspond with results presented by Bøhn (2016), where it is suggested that teachers are more likely to pay attention to linguistic criteria in the lower levels of proficiency. The teachers in this study reported that they do not find it particularly important to attend to linguistic criteria unless these linguistic features are perceived to disrupt a student's ability to communicate, but they were unable to articulate to which degree lacking competence is considered to disrupt communication. However, as reported on 'intonation', the findings also clearly indicate that oral performance grading is influenced by the students' mastery of linguistic criteria, but none of the teachers presented views suggesting that they hold conceptions or have clear descriptions of what constitutes 'mastery' either. These findings correspond with the OECD review team's report that there does not seem to be a shared understanding of what constitutes adequate, good and excellent performance in Norwegian education (Nusche et al., 2011, p. 129). The subject curriculum competence aims describe the competences students are to obtain, but do not define what constitutes different levels of achieved competence, making the responsibility of defining what constitutes competence the responsibility of local teachers.

The teachers reported to attend more of the narrower linguistic criteria when assessing higher-level students. This indicates that teachers do not only expect the higher-level students to perform better, but they are also expected to be able to do more than the lower-level students. Support for this understanding can be found in how the reference level descriptors in the CEFR are structured. As discussed in section 1.7.1, Hulstijn (2007, p. 663) explains that the reference level descriptors in the CEFR rest on the two closely intertwined pillars: 'quantity' and 'quality'. Quantity refers to what the learner is able to do, while quality refers to how well the learner is able to do it, and in the CEFR's reference level descriptors these are interwoven. I further discus the assumption that student mastery of the different criteria correlate to the same level of achievement in section 5.2.1.

5.1.2 Fluency – an elusive notion

While the teachers reported to consider mastery of linguistic criteria to be important, special attention should be paid to how the teachers reported to understand 'fluency'. The findings show that the teachers hold very different conceptions of whether it primarily relates to 'language' or 'content', and there was no definite consensus in how 'fluency' is to be understood. While all three teachers view 'fluency' as a linguistic criterion, two of the teachers relate it to non-linguistic criteria as well. One teacher identified 'fluency' as purely linguistic, another understood it as also relating to confidence in the presentation of content, and the third reports an understanding of 'fluency' as a criterion that intertwines all the other constructs and criteria.

It is evident that fluency is an elusive notion, and support for the lack of a common understanding among the teachers may be found in previous research (Schmidt, 1992; Guillot, 1999; Simensen, 2008), which has observed that not only is the term fluency difficult to define, it also has a wide range of definitions associated with it. 'Fluency' may be understood in a broad sense as a global ability, not very different from the term proficiency, or it may be understood in a narrower sense, where fluency in speech is understood as "…automatic, not requiring much attention or effort" (Schmidt, 1992, p. 358). 'Fluency' in the English subject curriculum is included in an aim placed among the competence aims under the main area *Oral Communication*, and students are expected to "express [themselves] fluently and coherently in a detailed and precise manner suited to the purpose and situation" (Norwegian Directorate for Education and Training, 2013, p. 10). In the CEFR (Council of Europe, 2001, p. 128) 'fluency' is not listed as a linguistic, sociocultural, or pragmatic competence, but as a generic factor and is defined as "the ability to articulate, to keep going, and to cope when one lands in a dead end" (Council of Europe, 2001, p. 128). The notion of 'fluency' in the subject curriculum presents a narrower linguistic understanding of 'fluency' than the definition presented in the CEFR, where 'fluency' is understood as determining a speaker's functional success. The lack of a common understanding of how to understand fluency has implications for validity and reliability, which I further discuss in section (5.2)

5.1.3 Content-related understandings

As for the content-construct, findings indicate that the teachers have relatively similar understandings, which correspond with findings reported by Bøhn (2016). The teachers report

that they assess their students' abilities to discuss and reflect to a larger degree than their attainment or presentation of subject matter knowledge. This way of conceptualizing 'content' fits with the revised taxonomy by Anderson and Krathwohl (2001, p. 27), as consisting of one dimension relating to cognitive processes and the other dimension to knowledge (see section 2.4). The teachers presented views of being more concerned with the cognitive process dimension than the knowledge dimension of the taxonomy, and I identify two interrelated reasons for this.

One possible reason is the general nature of the curriculum main area *Culture, Society and Literature* (Norwegian Directorate for Education and Training, 2013, p. 11) related to subject matter content, as it opens up for a wide range of topics, making the specificity of *what* being talked about less important than *how* it is talked about. This links to the other possible reason for the importance teachers attributed to reflection as a cognitive ability, which relates to the hierarchical structuring of verbs in the cognitive process dimension of the taxonomy. It must here be reiterated that the taxonomy assumes that learning at higher levels is dependent on having attained prerequisite knowledge and skills at the lower level (Murtonen, Gruber & Lehtinen, 2017, p. 116).

The competence aims under the main area *Culture, Society and Literature* are phrased with the verbs 'discuss' and 'elaborate', while the verb 'evaluate' is used with the competence aims in the main area *Language Learning* (Norwegian Directorate for Education and Training, 2013, p. 10). 'Discuss' and 'elaborate' may be placed on the second lowest level in the cognitive process dimension of the taxonomy, while 'evaluate' is placed on the second highest level. In this sense the use of verbs for the competence aims relating to *Language Learning* in the curriculum may be considered by teachers to be more advanced than the aims under *Culture, Society and Literature*. This is a possible explanation for the emphasis teachers place on the ability to reflect, and especially those aiming to receive the higher grades, because reflection is seen as a more advanced cognitive skill. However, there is the danger that such narrow understandings of the verbs in the curriculum assumes that teachers have a shared understanding of the hierarchical understanding presented in the taxonomy of cognitive abilities. An additional danger is that the importance attributed cognitive skills deemphasizes the importance of developing skills in how to convey topic knowledge.

5.1.4 Understandings of language-content relationship

Another notable finding that should be elaborated on is the teachers' presenting different views on how to understand the relationship between 'language' and 'content'. While two of the teachers reported views indicating that they consider language and content somewhat equally important to assess, the third clearly held a view of content to be the most salient construct to assess, and presented a view of language primarily as a tool for communicating content in assessment situations. The integration of language and content instruction has been a growing phenomenon in the field of second language teaching and assessment (Met, 1999), and I established in section 2.4 that the English subject may be placed somewhere in the middle of Met's (1999) continuum of language-driven and content-driven approaches, as both language and content are to be taught and assessed.

However, the findings suggest that out of the three teachers in this study, one would place the English subject on the very right on Met's (1999) continuum along with content-driven approaches. A possible reason for the lack of a common understanding of the relative importance between language and content is the undefined status of English as a school subject in Norway (see section 1.5). English is neither considered a foreign language nor a second language (Bøhn, 2016, p. 5), and Norway continue to rank as a country with very high English proficiency (Education First, 2017a). The somewhat special status of English and the lack of an official definition from educational authorities, some may consider English a language most students master to a fair degree, and for that reason there might be more importance attributed to content-knowledge in the English subject as oppose to language subjects with a foreign language status.

This finding also illustrates the importance of having a common understanding of not only the narrower linguistic sub-criteria, but also of the relationship between the larger constructs 'language' and 'content' in oral assessment. This lack of a common understanding might relate to how teachers operationalize 'communicative competence', which also invites validity and reliability issues. I attend to issues of operationalizing communicative competence in section 5.2 and issues related to validity and reliability in section 5.4.

5.1.5 Attending to non-relevant constructs

The lack of a common understanding of the constructs and criteria to assess is also present in that the teachers attend to constructs not directly relevant to the aims in the English subject curriculum when assessing oral performances, which correspond with findings by both Yildiz (2011) and Bøhn (2016). All three teachers reported to assess both body language and eye contact, as well as speech volume in oral presentation situations. While the findings in the study by Jenkins and Para (2003) suggest that situationally appropriate body language and volume positively affect the overall impression, the teachers in this study report to assess these paralinguistic features explicitly, and one even had body language listed as a linguistic criterion his scoring rubric.

Paralinguistic features are not part of the English subject curriculum, but there is mention of these features in the CEFR. In the CEFR, body language appropriate to the situation is emphasized as an important feature of non-verbal communication (Council of Europe, 2001, pp. 88-90). The importance attributed to body language may be seen in relation to the teachers reporting to mainly assessing oral English in performance situations, making presentation skills an integral part of what they assess in oral English, and thus also a relevant feature to assess when assessing their students' communicative competence. In addition, the relative teacher autonomy in what to teach (Eurydice, 2008, p. 31) gives teachers the opportunity to incorporate learning outcomes which may not be specified in the English subject curriculum. However, if teachers use these non-relevant constructs as part of the foundation for what constitutes their students' overall achievement, then this invites issues of validity and reliability, which further emphasizes the importance of having a shared understanding of what to assess.

5.2 Operationalizing Communicative Competence

The findings suggest that while teachers agree that 'communicative competence' is the overarching construct in oral assessment, but there is evidence in the empirical research findings that there are different understandings as to how it should be operationalized for assessment. In their initial accounts of what to assess, one teacher emphasized a holistic view that 'communicative competence' was what she considered most important to assess, while the two other teachers presented more detailed and analytical accounts. In addition, two of the

teachers presented understandings of language and content as equally important, while the third understood content as most important. Based on the discussion of the CEFR's description of 'Communicative language competence' (section 1.6.1), a relevant question to ask is whether it actually is possible to both identify and assess all the criteria that make up the construct 'communicative competence' and attend to these in the same oral assessment situation. While the Norwegian Directorate for Education and Training (2016, para.1) reports that it is sometimes beneficial to divide 'competence' aims into skills and abilities, the Directorate also emphasizes that these parts must be put back together, as it is the competence that teachers are to assess. However, as discussed in section 1.5.2, one may question if it is possible to assess communicative competence holistically without operationalizing this complex competence in an analytical discrete-point manner.

This points to a tension between analytical and holistic approaches to the assessment of communicative competence. The teachers report to assess a range of criteria in an analytical manner, but grade performances holistically. Analytical assessment is arguably essential in formative assessment in order to identify specific skills and abilities teachers should address to help students improve, as it is the combination of these skills and abilities that make up their competence. While the Norwegian Directorate for Education and Training (2016, para. 1) state that competence should be assessed holistically, a problem with assessing such complex constructs as 'communicative competence' in such a manner is the notion of 'aggregation' or the 'collapsing' a detailed performance profile into one single grade or level of achievement (Gipps, 1994, p. 85). Aggregating detailed assessment information into a single grade or level of achievement compromises the information offered by the assessment, and in this sense, a grade may obscure more than they clarify. This may be seen as especially problematic when oral assessments have formative purposes. The empirical research findings suggest that teachers have different understandings of the relative importance of the constructs and criteria that make up communicative competence, and conflating these into one overall achievement grade does not point specifically which constructs or criteria the students need to improve. The lack of a common understanding of the relative importance of what to assess also corresponds with the findings by Douglas (1994) and Bøhn (2016) in that teachers might arrive at the same grade, but for different reasons.

This is also a problem with the structure of the reference level descriptors such as the one found in the CEFR, as it rests on the assumption that student mastery of the different criteria correlate to the same level of achievement, but as the teachers report, some students may have

a narrow range in terms of quantity, but great depth in terms of linguistic quality; others may have a broad quantity range, but little linguistic quality; and some students' quantity range might match their performance quality. In other words, while the teachers report to understand that students may reach different levels of achievement in different constructs and criteria, reference level descriptors such as the ones found in the CEFR, do not allow for such understandings. The teachers recognize that the development of spoken language production is complicated and complex, but reference levels descriptors - such as the one found in the CEFR, are based on the assumption that development in learning happens in a linear and sequential fashion. The question is then how teachers are to attend to the assessment of communicative competence in a holistic manner to both collect information to use for formative purposes as well as making sure that the assigned grade is valid and reliable for summative purposes.

At the present time there is a new national curriculum being developed, and this deserves attention in this section, as central elements related to 'competence' are to be included in the new curriculum, and these based on recommendations in a Norwegian official report titled *The School of the Future* (Ludvigsen et al., 2015). This report emphasizes that basic skills and competences related to literacy and communication are to form the core school subjects also in future curricula, and will for this reason continue to be important aspects of what is to be assessed in the subject of English (Hasselgreen & Ørevik, 2018, p. 365). In addition, it is stated in the report that the term 'competence' is to be broadened to encompass more skills and abilities than it has in the Knowledge Promotion (LK06). The report states that competence aims in the future curricula are to encompass both subject-specific and overarching competences, such as social and emotional competences, as these are important skills to develop (Ludvigsen et al., 2015, p. 22). While one hardly can argue with the importance of developing such skills, it is reasonable to question how these are to be operationalized and assessed if they are integrated in the competence aims that make up both what to teach and what to assess. While learning is the main purpose of education (Dysthe, 2009, p. 33), it must again be emphasized that the conflation of assessment purposes in the classroom setting makes it difficult to separate learning from assessment, and this deserves more attention from those who develop curricula, as the inclusion of more competences in the competence aims may have further implications for validity and reliability.

5.2.1 Assessment criteria as operationalization of competence

While the teachers reported to mostly grade oral performances holistically, this grade was reportedly based on the analytical assessment of constructs and criteria, and all three teachers reported to use assessment criteria for more formal assessment situations, such as student oral presentations, which they had developed either by themselves or in collaboration with their students. Sadler (2010, p. 548) argues that the use of rubrics and criteria standards might actually inhibit the understanding of a full-bodied concept of quality, which one may argue is the holistic view of communicative competence, because teachers might then tend to prioritize specific qualities, rather than quality as a global property. Luoma (2004, p. x), on the other hand, argues that employing assessment criteria in oral assessment situations is crucial for assessments to be of high validity and reliability.

However, one may identify a third element relevant for discussion. Two of the teachers reported that their students are involved in the making of assessment criteria and one reported that his students are to self-assess their performance after an oral assessment situation. Yet, none of the teachers reported views indicating that their students have any involvement in the achievement grading of performance, which makes the teachers understanding of the criteria they have co-constructed the basis for the assessment. Findings in this thesis suggest that the teachers have different understandings of how to understand certain linguistic criteria, and one might thus assume that students might also hold different conceptions of how the same linguistic criteria are to be understood.

I use 'fluency' as an example, which is viewed as an elusive notion in section 5.1.2. Two teachers might report that they consider a student to be a fluent speaker, but they could be referring to very different aspects of performance. As the findings suggest, one might be referring to pace in spoken production, while another to transitions in oral presentations. As students are reportedly involved in the making of assessment criteria, a student might also agree that 'fluency' should be a criterion to be assessed. However, if the teacher and student do not have a shared perception of the concept of 'fluency', then this has implications for assessment. For instance, it is difficult to understand how student progress is to take place if the teacher reports that a student's 'fluency' needs to improve, but the student does not share the teachers understanding of what 'fluency' is or how it is to be understood. This illustrates the importance of having a shared understanding of what to assess in oral assessment, and any

lack of such has implications for both validity and reliability. There needs to be shared understanding of what to assess so that we can assess it consistently.

5.3 Learning theories in classroom assessment

The purposes of assessment in the classroom setting are conflated. Assessment is both a tool to enhance learning and a tool to make judgements of students' overall achievement. Inbar-Lourie (2008) addresses this as an issue of teachers having to "function simultaneously within two non-compatible cultures" (p. 388), and explains that teachers must "encourage within their classrooms to pursue sociocultural pedagogy and assessment practices, while concurrently required by school leaders and educational authorities to abide by the rules of the measurement culture" (p. 388). This corresponds well with the educational situation in Norway. As discussed in section 1.4.2, the Regulations to the Education Act (§3-16) states that formative assessment is not only to be understood as continuous classroom assessment to guide instruction and learning, but it is also to be understood as having the function of informing the teacher of a student's competence when deciding on a student's overall achievement grade. Formative and summative assessment must therefore aim to have a complimentary relationship and perform what Boud (2000, p. 160) refers to as a "double duty".

The notion of assessment having to perform a double duty is seen in the national initiative *Assessment for learning,* which has made it difficult to separate learning situations from assessment situations. *Assessment for Learning* is based on principles for quality formative assessment, where the purpose has been to improve learning by making students be involved in assessment practices (Norwegian Directorate for Education and training, 2015, pp. 1-2). The findings in this study indicate that students are involved in assessment practices by taking part in deciding what should be the assessment criteria for oral English presentations. This may be seen as is in line with sociocultural theories of learning, where learning is seen as a social collaborative activity. However, as both James (2006) and Baird et al., (2014) argue, assessment practices are weakly conceptualized within sociocultural views of learning.

Sociocultural theories of learning are, however, not the only learning theories that may be identified in classroom assessment practices. The teachers reported to make assessment criteria to allow for more analytical assessment. The act of operationalizing, or dividing complex competences into skills and abilities - which is what one generally does in the making of assessment criteria, may be traced back to behaviorist theories of learning (Skinner, 1954,

p. 94). As discussed in section 2.2, this is also mirrored in how the Knowledge Promotion (LK06) explains that one may have to operationalize 'competence' (Norwegian Directorate for Education and Training, 2016, para.1). However, as Baird et al. state (2014, p. 5), it is difficult to establishes any current assessment practices as entirely behaviorist, as importance is often attributed to the development of higher order thinking skills. The empirical research findings suggest that the teachers attribute a lot of importance to students showing the ability to reflect, and such emphasis attributed cognitive skills may also be found in the competence aims in the curriculum, which are formulated with verbs that describe intended cognitive processes (Norwegian Directorate for Education and Training, 2013).

This section has aimed to highlight that the boundaries between the different theories of learning are indistinct, which may often make assessment practices implicitly draw from a combination of theories (Baird et al., 2014, p. 23). This does not necessarily have to be problematic, but it deserves attention, as one may question how teachers are to have a shared understanding what to assess when there is no unified understanding of how to understand learning in a climate where the two may be seen as interdependent.

5.4 Validity and reliability when assessing different criteria

Luoma (2004) argues that "the validity of speaking scores is grounded in the purpose that the scores are intended to serve" (p. 185). However, this thesis questions how the concept of validity fits within a setting where assessment has two purposes. The findings suggest that the assessments the teachers conduct may be characterized as mini-summative with formative intentions. Teachers make a judgement of what the students master at a certain point in time, but also address what needs to improve. However, these findings indicate that the lack of a common standard to use for reference makes it difficult for teachers to define what constitutes different levels of performance and thus consequently have difficulty defining what students need to do in order to improve their competence. While it must be recognized that teacher subjectivity will always be an element involved in every act of assessment, it is essential that teachers' hold similar conceptions of what constitute different levels of performance both when assessing formatively and summatively for assessments to be valid and reliable.

The findings suggest that teachers attend to more of the narrow linguistic criteria when assessing higher-achieving students. One may assume that attending to different criteria when assessing different students is not necessarily problematic if the purpose of the assessment is

formative. However, it might be. Conversely, it is generally agreed upon that different students have different learning needs, and individual attention to these are essential for progress. That being said, if teachers tend to focus on the narrower linguistic criteria, such as grammar, only when assessing the higher-level students, their grammar is more likely to improve than the lower-level students, who are then at a disadvantage. In such cases, higher-level students might actually learn more than the lower-level students, and such assessment practice might have serious consequences for both the formative and summative purposes of assessment in a context where the two are conflated.

If the higher-achieving students have been assessed in a more analytical manner throughout the school year, one may assume that they are more likely to achieve the higher-grades because they already master and continue to improve the features teachers consider important. When teachers recognize that mastery of linguistic criteria is essential for receiving the higher grades, but do not attend to those when assessing lower-level students, then these students are not given the same opportunities to improve their competence and work towards the higher-levels of achievement as the already higher-achieving students. As a result, such practices may contribute to creating a greater gap between the lower performing and the higher performing students both in terms of learning and in terms of grading. This makes not attending to the same constructs and criteria throughout the school year an issue of validity and reliability in both formative and summative assessment when the overall achievement grade is to be decided on. If assessment results are to be used for certification, then it needs an adequate level of reliability for comparability purposes. For overall achievement grades based on classroom assessment to be reliable, Stiggins (1997) suggests that teachers should "sample student performance in a representative manner with sufficient depth to permit confident conclusions about proficiency" (p. 23). In other words, for teachers to sample student performance with sufficient depth, they need to attend to same criteria when assessing different students in order to make valid and reliable conclusions about their communicative competence in oral assessment situations. This further illustrates the importance of teachers having a shared understanding in the assessment of oral English in a setting where the assessment purposes are conflated.

6. Concluding remarks

In this thesis I have discussed the complex nature of assessment in Norway and consequences for oral English classroom assessment. Classroom assessment in the Norwegian context encompasses both formative and summative purposes, and while the distinction between the two might seem clear in terms of purpose in the Regulations to the Education Act (§3), it has been argued that it is unclear how this distinction is to be understood in practice. As part of this thesis' exploration, interviews were conducted with three L2 teachers at the upper secondary level to explore what they understand as important to assess in the classroom assessment of oral English. I now return to the research questions that served to guide this thesis:

The empirical research findings indicate that the teachers hold relatively similar overall understandings, and reported that they considered their students' 'communicative competence' to be most important to assess. Despite this, the teachers reported some variations in how they understood the relative importance and relationship between the more specific constructs and criteria, which suggests that the teachers operationalize 'communicative competence' differently. In addition, the findings suggest that teachers pay attention to features not identified in the English subject curriculum, such as body language. This might indicate that the teachers consider communicative competence to encompass more skills and abilities than the subject curriculum.

The teachers further presented different views on how to understand the relationship between the constructs 'language' and 'content' in oral assessment. Two of the teachers reported views of these being equally important, while the third reported 'content' to be more salient to assess. Content, as reported by all three teachers, relates to students' abilities to discuss and reflect to a larger degree than being able to convey subject matter knowledge. As has been discussed, English as a school subject has a somewhat undefined status, and this might be an explanation for the teachers' different understandings of the relative importance between these two constructs.

The teachers reported relatively similar understandings of the narrower linguistic criteria, expect from 'fluency'. Fluency is understood as both related to 'language' and 'content' as well as being an overarching feature that may affect the overall impression of a student's oral

performance. The empirical research findings further indicate that the teachers attend to more of these narrow linguistic criteria when assessing higher-achieving students. In a setting where the purposes of assessment are conflated, it has been argued that it is important that teachers attend to the same criteria when assessing different students. When teachers recognize that mastery of linguistic criteria is important for receiving the higher grades, then all students should be given the same opportunities to improve their competence.

Based on the findings in this thesis, I support Bøhn (2016) in his suggestion that educational authorities may consider introducing common rating scale guidelines on the national level to foster a more shared assessment culture. While Bøhn (2016) refers to oral examinations, such guidelines could benefit oral classroom assessment as well. A common rating scale might contribute to increase transparency in expectations of what constitutes 'communicative competence' as well as what constitutes different levels of performance. This may contribute to more valid and reliable assessment practises which has been argued are equally important in classroom assessment as in high stakes assessment.

However, classroom assessment also encompasses formative purposes, and as the findings in this thesis suggest, teachers do not necessarily attribute equal importance or have the same understandings of the specific constructs and criteria they assess. This suggests that the use of a common rating scale might not necessarily produce more authentic judgements, at least not unless the teachers conducting the assessments have a shared understanding of what they are to form an opinion on. This indicates that transparency in the assessment of oral English needs to be further developed on the local level. In a setting where assessment is both a tool to enhance learning as well as a tool to decide on overall achievement grades, it is of utmost importance that there is transparency in understandings of what to assess, and this should be given serious attention.

Bibliography

Alderson, J. (2007), The CEFR and the Need for More Research. *The Modern Language Journal, 91*(4), 659-663. https://dx.doi.org/10.1111/j.1540-4781.2007.00627_4.x

Ajideh, P. (2009). A Close Look at the Relationship between Multiple Choice Vocabulary Test and Integrative Cloze Test of Lexical Words in Iranian Context. *English Language Teaching*, *2*(3), 163-170. Retrieved from https://files.eric.ed.gov/fulltext/EJ1083051.pdf

Alvesson, M. (2003). Beyond neopositivists, romantics, and localists: A reflexive approach to interviews in organizational research. *Academy of Management Review*, *28*(1), 13-33. Retrieved from https://journals.aom.org/doi/abs/10.5465/amr.2003.8925191?related-urls=yes&legid=amr%3B28%2F1%2F13#

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

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

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

Backmann, K., & Sivesind, K. (2012). Kunnskapsløftet som reformprogram: fra betingelser til forventninger. In T. Englund, E. Forsberg & D. Sundberg (Eds.), *Vad räknas som*

kunskap? Läroplansteoretiska utsikter ock inblickar i lärarutbildningen och skola, 240-260. Stockholm: Liber.

Bagarić, V., & Djigunović J.M. (2007). Defining communicative competence. *Metodika*, *8*(1), 94-103.

Baird, J-A., Hopfenbeck, T.N., Newton, P., Stobart, G., & Steen-Utheim, A.T. (2014). *Assessment and Learning: State of the Field Review.* Oslo: Knowledge Center for Education. Retrieved from https://www.forskningsradet.no/servlet/Satellite?c=Rapport&cid=1253996755700&lang=en&pagename=kunnskapssenter%2FHovedsidemal

Black, P., & Wiliam, D. (1998a). *Inside the Black Box: Raising standards through classroom assessment.* Retrieved from https://www.rdc.udel.edu/wp-content/uploads/2015/04/InsideBlackBox.pdf

Black, P., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in education, 5*(1), 7–74. https://dx.doi.org/10.1080/0969595980050102

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

Black, P., & Wiliam, D. (2012). The reliability of assessments. In J. Gardner (Ed.), *Assessment and Learning* (pp. 243-263). London: Sage.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *The Taxonomy of educational objectives, handbook I: The Cognitive domain*. New York: David McKay Co., Inc.

Borger, L. (2014). *Looking beyond scores: A study of rater orientations and ratings of speaking.* (Licentiate thesis), University of Gothenburg, Gothenburg, Sweden. Retrieved from: http://hdl.handle.net/2077/38158

Boud, D. (2000) Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151-167. Retrieved from https://www.tandfonline.com/doi/pdf/10.1080/713695728?needAccess=true

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. Retrieved from https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa

Brindley, G. (1991). Defining language ability: The criteria for criteria. In S. Anivan (Ed.), *Current developments in language testing: Anthology series 25.* Singapore: Southeast Asian Ministers of Education Organization (SEAMEO) Regional Language Centre. Retrieved from: https://eric.ed.gov/?id=ED365150

Brinkmann, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing.* Thousand Oaks: Sage.

Broadfoot, P. (2007). *An introduction to assessment.* NY: The Continuum International Publishing Group.

Bøhn, H. (2015). Assessing Spoken EFR Without a Common Rating Scale: Norwegian EFL Teachers' Conceptions of Construct. *Sage Open*. October-December, 1-12. Retrieved from https://dx.doi:10.1177/2158244015621956

Bøhn, H. (2016). *What is to be assessed? Teachers' understanding of constructs in an oral English examination in Norway.* (Doctoral Thesis). Retrieved from https://www.duo.uio.no/bitstream/handle/10852/53229/PhD-Boehn-DUO.pdf?sequence=1&isAllowed=y

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics,1*, 1-47. https://dx.doi.org/10.1093/applin/I.1.1

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Council of Europe (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Strasbourg: Council of Europe, Language Policy Unit. Retrieved from https://www.coe.int/en/web/portfolio/the-common-european-framework-of-reference-for-languages-learning-teaching-assessment-cefr-

Council of Europe (2018). Common European Framework of Reference for Languages: Learning, Teaching, Assessment - Companion volume with new descriptors.

Retrieved from https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

de Corte, E (2010). Historical developments in the understanding of learning. In: H Dumont, D. Istance & F. Benavides (Eds.), *The Nature of Learning. Using research to inspire practice.* OECD. Retrieved from http://www.oecd.org/education/ceri/50300814.pdf

Dobson, S. & Engh, R. (2010). *Vurdering for læring i fag*. Kristiansand: Høgskoleforlaget.

Douglas, D. (1994). Quantity and Quality in Speaking Test Performance. *Language Testing, 11*(2), 125-144.

Dysthe, O. (2008). Klasseromsvurdering og læring. *Bedre Skole, 4*, 16-23.

Dysthe, O. (2009). Læringssyn og vurderingspraksis. In J. Frost (Ed.), *Evaluering – I et dialogisk perspektiv* (pp. 33-51). Oslo: Cappelen Damm.

Education First (2017a) *EF English Proficiency Index*. Retrieved from https://www.ef.no/epi/

Education First (2017b). *EF English Proficiency Index: About the index*. Retrieved from https://www.ef.no/epi/about-epi/

Ely, M., Vinz, R., Downing, M. & Anzul, M. (1997). *On writing qualitative research: Living by words*. London: Falmer.

Eurydice. (2008). *Levels of Autonomy and Responsibilities of Teachers in Europe*. Retrieved from http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/094EN.pdf

Figueras, L. (2012). The impact of the CEFR. *ELT Journal 66*(4), 477-485. Retrieved from https://dx.doi.org/10.1093/elt/ccs037

Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. Oxford: Routledge.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer.

Glaser, R. & Klaus, D.J. (1962). Proficiency measurement: Assessing human performance. In R.M. Gagné (Ed.), *Psychological principles in system development* (pp. 419-474). NY: Holt, Ringhart and Winston.

Glaser, R. (1994), Instructional Technology and the Measurement of Learning Outcomes: Some Questions. *Educational Measurement: Issues and Practice, 13*, 519-521. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3992.1994.tb00561.x

Gorard, S. (2012). Mixed Methods Research in Education: Some Challenges and Possibilities. *Mixed Methods in Educational Research: Report from March seminar 2012*. The Research Council of Norway. Retrieved from http://www.uv.uio.no/ils/personer/vit/kirstik/publikasjoner-pdf-filer/klette.-mixed-methods.pdf

Graddol, D. (1997). *The Future of English? A guide to forecasting the popularity of the English language in the 21st century.* London: The British Council. Retrieved from http://www.uniurb.it/lingue/matdid/rossi/2012-13/Lin-AZ_I_anno/The_Future_of_English_-_David_Graddol.pdf

Guillot, M.-N. (1999). *Fluency and its teaching*. Clevedon, England: Multilingual Matters.

Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly, 11*(2), 186–197. https://doi:10.1080/15434303.2014.895829

Hartberg, E.W., Dobson, S. & Gran. L. (2012). *Feedback i skolen*. Oslo: Gyldendal Akademisk.

Hasselgreen, A., & Ørevik, S. (2018). Assessment and Testing. In A.-B. Fenner & A., S. Skulstad (Eds.), *Teaching English in the 21st Century: Central issues in English didactics* (pp. 361-384). Bergen: Fagbokforlaget.

Hopfenbeck, T. N. (2014). Testing times: Fra Pisa til nasjonale prøver. In J. H. Stray & L. Witteck (Eds.), *Pedagogikk: En grunnbok*. Oslo: Cappelen Damm Akademisk.

Hopfenbeck, T., Tolo, A., Florez, T. & El Masri, Y. (2013). *Balancing trust and accountability? The Assessment for Learning Programme in Norway.* OECD report. Retrieved from http://www.oecd.org/education/ceri/Norwegian%20GCES%20case%20study%20OECD.pdf

Hulstijn, J. (2007). The Shaky Ground beneath the CEFR: Quantitative and Qualitative Dimensions of Language Proficiency. *The Modern Language Journal, 91*(4), 663-667. Retrieved from http://www.jstor.org/stable/4626094

Hymes, D. H. (1972). On Communicative Competence. In J. B. Pride, & J. Holmes, (Eds.), *Sociolinguistics* (pp. 269-293). Baltimore, USA: Penguin Education, Penguin Books Ltd.

Inbar-Lourie, O. (2008). Language assessment culture. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7, pp. 285-289). New York: Springer.

Iwashita, N., Brown, A., McNamara, T. & O'Hagan, S. (2008). Assessed Levels of Second Language Speaking Proficiency: How Distinct?, *Applied Linguistics*, *29*(1), 24–49. https://dx.doi.org/10.1093/applin/amm017

James, M. (2006). Assessment and learning. In: S. Swaffield (Ed.), *Unlocking Assessment. Understanding for reflection and application* (pp. 20-35). Abingdon, UK, Routledge.

Jenkins, S. & Parra, I. (2003). Multiple Layers of Meaning in an Oral Proficiency Test: The Complementary Roles of Nonverbal, Paralinguistic, and Verbal Behaviors in Assessment Decisions. *The Modern Language Journal, 87*(1), 90-107. https://doi.org/10.1111/1540-4781.00180

Kunnan, A. J. (2008). Large-scale language assessments. In E. Shohamy & N. Hornberger (Eds.), *Encyclopaedia of language and education* (2nd ed., Vol. 7, pp. 135-155). New York, NY Springer Science.

Kvale, S. (2007). *Doing Interviews*. London: Sage Publications Ltd.

Langseth, I. (2013). Vurderingspraksiser i engelskfaget. In Sandvik, L. V., Buland, T. (Eds.), *Vurdering i skolen. Operasjonaliseringer og praksiser.* Retrieved from http://www.udir.no/Upload/Rapporter/2013/FIVIS2.pdf?epslanguage=n

Ludvigsen et al,. (2014). *Elevenes læring i fremtidens skole: Et kunnskapsgrunnlag.* (NOU, 2014:7). Oslo: Ministry of Education and Research. Retrieved from https://www.regjeringen.no/contentassets/e22a715fa374474581a8c58288edc161/no/pdfs/nou201420140007000dddpdfs.pdf

Ludvigsen et al,. (2015). *The School of the Future: Renewal of Subjects and Competences.* (NOU 2015: 8). Oslo: Ministry of Education and Research. Retrieved from https://www.regjeringen.no/no/dokumenter/nou-2015-8/id2417001/

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

McNamara, T. (2000). *Language Testing.* Oxford University Press.

Met, M. (1999). *Content-based instruction: Defining terms, making decisions.* NLFC reports. Retrieved from http://carla.umn.edu/cobaltt/modules/principles/decisions.html

Ministry of Education and Research. (2004). *Kultur for læring.* (Meld. St. 31, 2003-2004). Retrieved from https://www.regjeringen.no/contentassets/988cdb018ac24eb0a0cf95943e6cdb61/no/pdfs/stm200320040030000dddpdfs.pdf

Ministry of Education and Research (2014). Lærerløftet: På lag for kunnskapsskolen. (Plan/Strategy). Retrieved from https://www.regjeringen.no/no/dokumenter/Larerloftet/id2001933/

Ministry of Education and Research. (2016). *Fag-Fordypning-Forståelse – En fornying av Kunnskapsløftet.* (Meld. St. 28, 2015-2016). Retrieved from https://www.regjeringen.no/no/dokumenter/meld.-st.-28-20152016/id2483955/

Moss, P. A. (2003). Reconceptualizing Validity for Classroom Assessment. *Educational Measurement: Issues and Practice*, *22*, 13–25. https://doi:10.1111/j.1745-3992.2003.tb00140.x

Murtonen, M., Gruber, H. & Lehtinen, E. (2017). The return of behaviourist epistemology: A review of learning outcome studies. *Educational Research Review, 22,* 114-128. https://doi.org/10.1016/j.edurev.2017.08.001

Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research and Perspectives*, *10*(1-2), 1-29. https://doi:10.1080/15366367.2012.669666

North, B. (2004, April 15). Europe's framework promotes language discussion, not directives. *The Guardian*. Retrieved from https://www.theguardian.com/education/2004/apr/15/tefl6

North, B. (2007). The CEFR Illustrative Descriptor Scales. *The Modern Language Journal, 91*(4), 665-669. https://doi:10.1111/j.1540-4781.2007.00627_3.x

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

Norwegian Directorate for Education and Training. (2013). *The English subject curriculum.* Retrieved from http://www.udir.no/kl06/ENG1-03

Norwegian Directorate for Education and training. (2015). *Fire prinsipper for god underveisvurdering.* Retrieved from https://www.udir.no/laring-og-trivsel/vurdering/om-vurdering/underveisvurdering/

Norwegian Directorate for Education and Training. (2016). *Å forstå kompetanse.* Retrieved from https://www.udir.no/laring-og-trivsel/lareplanverket/forsta-kompetanse/

Norwegian Directorate of Education and Research (2017). Hva er fagfornyelsen? Retrieved from https://www.udir.no/laring-og-trivsel/lareplanverket/fagfornyelsen/hva-er-fornyelse-av-fagene/

Nusche, D., Earl, L., Maxwell, W., & Shewbridge, C. (2011). *OECD Reviews of Evaluation and Assessment in Education: Norway*. Retrieved from https://www.oecd.org/norway/48632032.pdf

Regulations to the Education Act, FOR-2009-07-01-964. (2009). Retrieved from https://lovdata.no/dokument/SF/forskrift/2006-06-23-724

Rindal, U. (2014). What is English? *Acta Didactica Norge, 8*(2), 1-17.

Roberts, C. (1997). Transcribing talk: Issues of representation. *TESOL Quarterly 31*(1), 167-172. Retrieved from https://www.jstor.org/stable/3587983

Ruch, G. M. (1924). *The improvement of the written examination.* Chicago: IL: Scott: Foresman.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144. Retrieved from https://link.springer.com/article/10.1007/BF00117714

Sadler, D. R. (2010) Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*(5), 535-550. https://dx.doi:10.1080/02602930903541015

Sandvik, L.V. (2013) Perspektiver på individuell vurdering i skolen. In Sandvik, L. V., Buland, T. (Eds.), *Vurdering i skolen. Operasjonaliseringer og praksiser.* Retrieved from http://www.udir.no/Upload/Rapporter/2013/FIVIS2.pdf?epslanguage=n

Savignon, S. J. (1972). *Communicative Competence: An Experiment in Foreign Language Teaching.* Philadelphia: The Centre for Curriculum Development. https://dx.doi.org/10.1002/9781118784235.eelt0047

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14,* 357–385. https://dx.doi.org/10.1017/S0272263100011189

Serafini, F. (2000-2001). Three paradigms of assessment: Measurement, procedure, and inquiry. *The Reading Teacher*, *54*(4), 384-393. Retrieved from http://www.jstor.org/stable/20204924

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, *24,* 86-97.

Simensen, A. M. (2008). Fluency: an aim in teaching and a criterion in assessment. *Acta Didactica Norge, 4*(1), 1-13.

Simensen, A. M. (2014). Skolefaget engelsk: Fra britisk engelsk til mange slags "engelsker"– og veien videre. *Acta Didactica Norge, 8*(2), 1-18.

Smith, J. K. (2003). Reconsidering Reliability in Classroom Assessment and Grading. *Educational Measurement: Issues and Practice*, *22*, 26-33. http://dx.doi.org/10.1111/j.1745-3992.2003.tb00141.x

Snow, M. A., & Katz, A. M. (2014). Assessing language and content. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 1, pp. 230-247). Chichester, UK: Wiley Blackwell.

Stiggins, R. J. (1997). *Student-centered classroom assessment* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Stobart, G. (2012). Validity in Formative Assessment. In J. Gardner (Ed.), *Assessment and Learning* (pp. 233-242). Sage: Los Angeles.

Svenkerud, S., Klette, K., & Hertzberg, F. (2012). Opplæring i muntlige ferdigheter. *Nordic Studies in Education, 32*(1), 35-49. Retrieved from https://www.idunn.no/np/2012/01/opplaering_i_muntlige_ferdigheter

The Norwegian Universities and Colleges Admission Service (2013). *Ordinær kvote*. Retrieved from http://www.samordnaopptak.no/info/opptak/opptakskvoter/ordinerkvote.html

Throndsen, I., Hopfenbeck, T. N., Lie, S., & Dale, E. L. (2009). *Bedre vurdering for læring: Evaluering av modeller for kjennetegn på måloppnåelse i fag.* Report. Retrieved from https://www.udir.no/globalassets/upload/forskning/2015/evalueringsrapport-til-udir-2.pdf

Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principle, Policy & Practice, 21*(2), 221-237. https://doi.org/10.1080/0969594X.2013.830079

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, Ma: Harvard University Press.

Yildiz, L. M. (2011). *English VG1 level oral examinations: How are they designed, conducted and assessed?* (Master's thesis, University of Oslo). Retrieved from https://www.duo.uio.no/bitstream/handle/10852/32421/YildizMaster.pdf?sequence=2

Appendices

Appendix 1

Competence aims in the English subject curriculum identified as relevant to oral assessment:

Language learning:

- evaluate and use different situations, working methods and learning strategies to further develop one's English-language skills
- evaluate own progress in learning English
- evaluate different digital resources and other aids critically and independently, and use them in own language learning

Oral communication:

- 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 program
- understand the main content and details of different types of oral texts about general and academic topics related to one's education program
- 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 program
- use patterns for pronunciation, intonation, word inflection and various types of sentences in communication
- interpret and use technical and mathematical information in communication

Culture, society and literature:

- discuss and elaborate on culture and social conditions in several English-speaking countries
- present and discuss current news items from English language sources
- discuss and elaborate on the growth of English as a universal language
- discuss and elaborate on different types of English language literary texts from different parts of the world
- discuss and elaborate on English language films and other forms of cultural expressions from different media
- discuss and elaborate on texts by and about indigenous peoples in English-speaking countries
- select an in-depth study topic within one's education program and present this

(Norwegian Directorate for Education and Training, 2013, pp. 11-12)

Appendix 2

<u>Intervjuguide</u>

1.1 Alder:

1.2 Morsmål (L1):

1.3 Utdanning:

1.4 Erfaring som lærer:

Åpning:

I hvilke situasjoner vurderer du muntlig engelsk?

- Vurderer du muntlig aktivitet i vanlige skoletimer?

Hva ser du etter når du vurderer en elev i muntlig vurderingssituasjon? (Åpent, la dette føre neste spørsmål)

Språk:

- Hvordan vurderer du språk når du vurderer en elev i en muntlig vurderingssituasjon?
- Kan du utype hva du anser som viktig å vurdere innen språk?

Avhengig av hva de svarer, still dette som oppfølging:

- Hvordan forstår du «…»? Hvor viktig tenker du «…» er?

(Spør om flyt, grammatikk, vokabular, intonasjon og uttale om de ikke tar det opp selv)

Innhold:

- Er innhold viktig i muntlig vurdering?
- I så fall, hvordan vurderer du innhold når du vurderer en elev i en muntlig vurderingssituasjon?
- Kan du utdype hva du anser som viktig å vurdere når det gjelder innhold?

Vurderer du noen gang språk og innhold hver for seg?

Når vet du at du vurderer språk og når vet du at du vurderer innhold?

Er det andre ting du vil si er viktig i vurdering av muntlig engelsk?

Appendix 3

Forespørsel om deltakelse i forskingsprosjekt

Hvilke kompetanser lærere forstår som viktige å teste i vurdering av muntlig engelsk

Bakgrunn og formål

Dette prosjektet gjennomføres i forbindelse med min mastergrad i kultur og språkfagenes didaktikk ved Høgskolen i Innlandet. Jeg ønsker å undersøke læreres forståelse av hvilke kompetanser de vurderer som viktige i vurdering av muntlig engelsk.

Hva innebærer deltakelse i studien?

Datainnsamlingen foregår gjennom intervjuer med lærere på VG1. Intervjuet gjennomføres på om lag en halv time. Spørsmålene omhandler hva lærere forstår som viktig å vurdere i vurdering av muntlig engelsk.

Hva skjer med informasjonen om deg?

Alle personopplysninger vil bli behandlet konfidensielt, og jeg og mine veiledere er de eneste som vil ha tilgang til disse. Opplysningene vil lagres for å ivareta informantenes konfidensialitet. Deltakere vil ikke kunne gjenkjennes i publikasjonen.

Prosjektet skal etter planen avsluttes 15.05.2018. Innsamlet data vil bli slettet ved denne datoen om prosjektet leveres til normert tid.

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

Stine Lisa Johannessen

Telefon:

Epost:

Veiledere:

Christina Sandhaug   christina.sandhaug@inn.no

Heidi Silje Moen   heidi.moen@inn.no

Appendix 4

NSD

Christina Sandhaug

2418 ELVERUM

Vår dato: 11.12.2017   Vår ref: 57442 / 3 / AMS   Deres dato: Deres ref:

Forenklet vurdering fra NSD Personvernombudet for forskning

Vi viser til melding om behandling av personopplysninger, mottatt 27.11.2017.
Meldingen gjelder prosjektet:

57442	Hvilke kompetanser lærere forstår som viktige å teste i vurdering av muntlig engelsk
Behandlingsansvarlig	Høgskolen i Innlandet, ved institusjonens øverste leder
Daglig ansvarlig	Christina Sandhaug
Student	Stine Lisa Johannessen

Vurdering

Etter gjennomgang av opplysningene i meldeskjemaet med vedlegg, vurderer vi at prosjektet er omfattet av personopplysningsloven § 31. Personopplysningene som blir samlet inn er ikke sensitive, prosjektet er samtykkebasert og har lav personvernulempe. Prosjektet har derfor fått en forenklet vurdering. Du kan gå i gang med prosjektet. Du har selvstendig ansvar for å følge vilkårene under og sette deg inn i veiledningen i dette brevet.

Vilkår for vår vurdering

Vår anbefaling forutsetter at du gjennomfører prosjektet i tråd med:

- opplysningene gitt i meldeskjemaet
- krav til informert samtykke
- at du ikke innhenter sensitive opplysninger
- veiledning i dette brevet
- Høgskolen i Innlandet sine retningslinjer for datasikkerhet

Veiledning

Krav til informert samtykke

Utvalget skal få skriftlig og/eller muntlig informasjon om prosjektet og samtykke til deltakelse.
Informasjon må minst omfatte:

- at Høgskolen i Innlandet er behandlingsansvarlig institusjon for prosjektet
- daglig ansvarlig (eventuelt student og veileders) sine kontaktopplysninger
- prosjektets formål og hva opplysningene skal brukes til

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Appendix 5

Interview

ME: Hvordan vurderer du språk når du vurderer en elev i en muntlig vurderingssituasjon?

H: Ja som sagt er det jo viktig at eleven får frem det dem vi si da ... og det er klart at da ... så lenge.. det kan jo være at eleven bruker litt enkelt språk men likevel skjønner jeg godt hva eleven vil formidle.. men da blir jo kanskje karakteren litt dårligere eller litt lavere måloppnåelse ... når man har et snevrere språk.. men tenker du også grammatikk her eller tenker du mest ja.. for grammatikken er jo ... så lenge de kommuniserer så synes jeg det er det viktigste ... om de bøyer et verb feil så synes jeg ikke det er så stor … jeg synes ikke det er et så stort avvik for eksempel.. men det kommer an på rekkevidden av det.. både på forståelsen av det og om de gjør det gjennomgående så vi ser at det er et problem … da … det kan godt være de sier en feil et par ganger og det er ikke noe problem ... men hvis det er gjennomgående masse grammatikkfeil så er det klart at da det er en pekepinn på at man må ta litt tak i det

Interview

ME: Hvordan vurderer du språk når du vurderer muntlig når du vurderer en elev i en muntlig vurderingssituasjon.. hvis vi ser på språk for seg selv?

N: Det er jo på en måte de tre faktorene som jeg nevnte ordforråd uttale og grammatikk … jeg legger vel kanskje størst vekt på ordforråd ... uttale er det vel på en måte holdt jeg på å si *greit* å ha noen feil med det og også grammatikk det jeg er opptatt av er jo at det ikke forstyrrer kommunikasjonen ... det er klart mangler du på en måte ordforråd så kan det forstyrre kommunikasjonen på en annen måte enn hvis du har noen feil med «is» og «are» for eksempel som ikke forstyrrer kommunikasjonen så mye .. men det er klart hvis det er blir veldig mange av en feil … så kan jo det trekke ned selvfølgelig om det gjelder uttale eller … grammatikk

Interview

ME: Så hvordan vurderer du språk når du vurderer en elev i en muntlig vurderingssitasjon? … du har snakka litt om det men …

B: I forhold til språk så føler jeg at det går.. i forhold til det med muntlig vurdering og nå snakker jeg bare om presentasjoner da så kan vi ta det som med klasserommet etterpå.. bare for å tydeliggjøre det skillet.. siden jeg har rubrikken min som blir veldig fastsatt.. i forhold til presentasjoner så går det både på det med kroppsspråk og på en måte hvordan de utstråler i klasserommet.. som da er hvordan de presenterer generelt.. og da går det på tilstedeværelse på en måte.. at de er der ... for noe er det mer vanskelig for noen er mer innadvendte ... ja så det er noen som er innadvendte som har overrasket veldig positivt og det er vel egentlig ingen som har vært på den negative siden egentlig i det hele tatt så langt.. og så går det litt ... jeg har ett punkt som går på grammatiske feil men jeg prøver å ikke vektlegge det så mye.. det er på en måte hvis det du skal oppnå sekseren så føler jeg at du burde ha relativt lite feil når du snakker eventuelt rette deg selv.. så går det på det med faglig og generelt ordforråd..

Appendix 6

Hva ser du etter når du vurderer en elev i muntlig vurderingssituasjon? …..På en måte hvis det er noen grammatikkområder som vi har gått mer innpå som subject-verb agreement og generelt bøyningsformer så blir det ofte det som blir implementert … det har vært mye fokus på gjennomgang av stoff og det å lære de hvordan de skal presentere .. så det har vært ganske breie vurderingskriterier som da har gått på for eksempel sånn bare få grammatikkfeil… men litt sånn skjønn i forhold til hva som er mye og hva som er få og da på en måte neste nivå er noe grammatikkfeil som ikke hindrer kommunikasjon og så videre og så har det vært litt på flyt i språket…uttale .. bare det å stå foran og presentere.. kroppsspråk .. øyekontakt……

Constructs	Criteria + sub-criteria

Language	General communication
	Linguistic competence
	   - Grammar
	         o Subject-verb
	   - Fluency
	         o Pace
	   - Pronunciation
	         o Native speaker
	   - Intonation
	   - Vocabulary
	         o General
	         o Academic

Content	General knowledge Comprehension Reflection **Fluency**
Other	Presentation tool Body language Eye contact Volume

Stine Johannessen
grammar

Stine Johannessen
Content

Stine Johannessen
Grammar

Stine Johannessen
Grammar + communicative competence

Stine Johannessen
Pronunciation

Stine Johannessen
Presentation skills

Stine Johannessen
Body language

Stine Johannessen
Eye contact