

Overall achievement- and exam grades in English - two of a kind?

A study of assessment practices in Norwegian lower secondary school





The Faculty of Arts and Education

MASTER'S THESIS

| | |
|--|--|
| Study programme: Advanced teacher education for levels 8-13 | Spring term, 2020 Open |
| Author: Jane Hodne | |
| Supervisor: Kjetil Vikhamar Thengs | |
| Title of thesis: Overall achievement- and exam grades in English - two of a kind? A study of assessment practices in Norwegian lower secondary school | |
| Keywords: summative assessment, overall achievement grades, exam grades, teachers' assessment, grading practices | Pages: 82 + attachment/other: 55 Stavanger, 11.05.2020 |

Abstract

This study concerns the deviations between overall achievement- and exam grades in the English subject at lower secondary schools in Norway. The aim of the study is to describe and compare deviations between overall achievement- and exam grades in light of teachers' conceptions and the Norwegian school system. As a result, the study is based on both qualitative and quantitative data. Materials from six semi-structured interviews with teachers from lower secondary schools in Norway and their assessment of an exam paper written in schoolyear 2017/2018 were used to gain a perspective of teachers' grading practices and how they perceive deviations between overall achievement- and exam grades. Public statistics of the average overall achievement- and exam grades from the platform *Skoleporten* (2020) have been collected and analysed in order to see whether there are systematic deviations between the two forms of assessment. However, in order to see whether these systematic deviations were true for the majority of pupils, statistics of individual pupils and their overall achievement- and exam grades in English were also analysed. Two lower secondary schools therefore provided the present study with individual statistics in both written and oral English.

Assessment is a challenging topic for teachers, and many enlighten the need for clearer guidelines in the field. The present study found that the national average grade score was higher in oral examinations compared to overall achievements. The national average grade scores shown in written English were lower for examinations than overall achievements. Looking into particular schools show contradictory findings. A written exam paper graded by six teachers managed to receive three different grades from the grade scale, even though there are guidelines in written examinations seeing as it is centrally given, as opposed to oral examinations which are locally given. The overall goal from these materials is to shed light on grading practices in lower secondary schools in Norway and to add valuable inquiry and research into the field of applied linguistics.

Keywords: summative assessment, overall achievement grades, exam grades, teachers' assessment, grading practices

TABLE OF CONTENTS

| | | |
|------------|---|-----------|
| 1 | INTRODUCTION | 1 |
| 2 | THEORETICAL BACKGROUND | 4 |
| 2.1 | Assessment | 4 |
| 2.1.1 | Formative and summative assessment | 4 |
| 2.1.2 | International studies of grading and assessment practices | 6 |
| 2.2 | Assessment practices..... | 9 |
| 2.2.1 | Ipsative assessment..... | 9 |
| 2.2.2 | Norm-referenced assessment | 10 |
| 2.2.3 | Criterion-referenced assessment | 10 |
| 2.2.4 | Informal and formal assessment..... | 11 |
| 2.3 | The Norwegian context | 12 |
| 2.3.1 | The English subject curriculum..... | 12 |
| 2.3.2 | Overall achievements in the English subject..... | 13 |
| 2.3.3 | Examinations in the English subject | 15 |
| 2.3.4 | White papers | 18 |
| 2.3.5 | Utdanningspeilet..... | 21 |
| 2.4 | Norwegian studies of deviations between overall achievement- and exam grades | 24 |
| 3 | METHODOLOGY AND MATERIALS | 27 |
| 3.1 | Qualitative data | 27 |
| 3.1.1 | Participants | 30 |
| 3.2 | Quantitative data | 31 |
| 4 | PRESENTATION OF FINDINGS..... | 33 |
| 4.1 | Interviews..... | 33 |
| 4.1.1 | Teacher experience | 34 |
| 4.1.2 | Grading practice | 36 |
| 4.1.3 | Teachers' support when grading..... | 37 |
| 4.1.4 | Overall achievement grades | 38 |
| 4.1.5 | The relationship between overall achievement- and exam grades | 42 |
| 4.2 | Exam assessment | 47 |
| 4.2.1 | Participant 1..... | 47 |
| 4.2.2 | Participant 2..... | 48 |
| 4.2.3 | Participant 3..... | 49 |
| 4.2.4 | Participant 4..... | 50 |
| 4.2.5 | Participant 5..... | 50 |
| 4.2.6 | Participant 6..... | 52 |
| 4.3 | Statistics from <i>Skoleporten</i> | 53 |
| 4.3.1 | Oral English | 53 |
| 4.3.2 | Written English | 56 |
| 4.4 | Individual statistics..... | 59 |
| 4.4.1 | Oral English | 60 |

| | | |
|----------|--|-----------|
| 4.4.2 | Written English | 61 |
| 5 | DISCUSSION..... | 65 |
| 5.1 | Deviations between overall achievement- and exam grades in the English subject | 65 |
| 5.2 | The possible factors which might influence a deviation in grading | 67 |
| 5.3 | The importance teachers attach to differences between overall achievements and exams | 72 |
| 5.4 | Teachers' approach with grading in the English subject | 74 |
| 5.5 | Grading practices in light of recommendations and regulations | 78 |
| 6 | CONCLUSION | 80 |
| | REFERENCES..... | 83 |
| | APPENDIX 1: Interview guide – English version | 89 |
| | APPENDIX 2: Interview guide – Norwegian version..... | 91 |
| | APPENDIX 3: Interview summary – Participant 1..... | 93 |
| | APPENDIX 4: Interview summary – Participant 2..... | 98 |
| | APPENDIX 5: Interview summary – Participant 3..... | 103 |
| | APPENDIX 6: Interview summary – Participant 4..... | 110 |
| | APPENDIX 7: Interview summary – Participant 5..... | 117 |
| | APPENDIX 8: Interview summary – Participant 6..... | 123 |
| | APPENDIX 9: Consent form | 129 |
| | APPENDIX 10: Exam paper | 131 |

TABLE OF FIGURES

| | |
|--|----|
| Figure 1: "The normal distribution in norm-referenced assessment" (Bøhn 2018: 233). | 10 |
| Figure 2: Characteristics of goal achievement after year 10 in oral English (UDIR 2017, my translation). | 14 |
| Figure 3: Characteristics of goal achievement after year 10 in written English (UDIR 2017, my translation). | 14 |
| Figure 4: Characteristics of goal achievement for ENG0012/ENG0025 at centrally given exam (UDIR 2020, my translation). | 17 |
| Figure 5: Categories and questions presented from the interviews. | 34 |
| Figure 6: The national average overall achievement- and exam grades in oral English from schoolyears 2014/2015-2018/2019. | 54 |
| Figure 7: Oral English overall achievement- and exam grades at School A from schoolyears 2014/2015-2018/2019. | 54 |
| Figure 8: Oral English overall achievement- and exam grades at School B from schoolyears 2014/2015-2018/2019. | 55 |
| Figure 9: Oral English overall achievement- and exam grades at School C from schoolyears 2014/2015-2018/2019. | 56 |
| Figure 10: The national average overall achievement- and exam grades in written English from schoolyears 2014/2015-2018/2019. | 57 |
| Figure 11: Written English overall achievement- and exam grades at School A from schoolyears 2014/2015-2018/2019. | 58 |
| Figure 12: Written English overall achievement- and exam grades at School B from schoolyears 2014/2015-2018/2019. | 58 |
| Figure 13: Written English overall achievement- and exam grades at School C from schoolyears 2014/2015-2018/2019. | 59 |
| Figure 14: Overall achievement- and exam grades in oral English 2018/2019 at School A. | 60 |
| Figure 15: Overall achievement- and exam grades in oral English 2018/2019 at School C. | 61 |
| Figure 16: Overall achievement- and exam grades in written English 2018/2019 at School A. | 62 |
| Figure 17: Overall achievement- and exam grades in written English 2018/2019 at School C. | 64 |
| Figure 18: The distribution of grades from Schoolyear 2018/2019 in written English at School A. | 69 |
| Figure 19: The distribution of grades from schoolyear 2018/2019 in written English at School C. | 69 |
| Figure 20: Participants 1-6 assessment of an exam paper from School C in schoolyear 2017/2018. | 76 |

1 INTRODUCTION

This thesis is a mixed method study of deviations between overall achievement- and exam grades in the English subject in Norwegian lower secondary school. The aim of the study is to describe and compare deviations between overall achievement- and exam grades in light of teachers' conceptions and the Norwegian school system. In order to achieve this, the study is based on both qualitative and quantitative data. Materials from six semi-structured interviews with teachers from lower secondary schools in Norway and their assessment of an exam paper written in schoolyear 2017/2018 were used to gain a perspective of teachers' grading practices and how they perceive deviations between overall achievement- and exam grades.

In addition, public statistics of the average of overall achievement- and exam grades from the platform *Skoleporten*¹ (2020) have been collected and analysed in order to see whether there are systematic deviations between the two types of assessments. However, in order to see whether these systematic deviations were true for the majority of pupils, the researcher felt it was necessary to look into statistics of individual pupils and their overall achievement- and exam grades in English. Two lower secondary schools have therefore provided the present study with individual statistics in both written and oral English. The overall goal from these materials is to shed light on grading practices in lower secondary schools in Norway and to add valuable inquiry and research into the field of applied linguistics.

One of the most central purposes of a country's grading practice is selection for further education and career. It is therefore important that teachers' grading practices are of high legitimacy in order to properly fulfil this task. Pupils' achievements should correspond with their grade. A significant part of the grades pupils receive on their diplomas are overall achievement grades. For lower secondary schools, these constitute a large part of the calculation basis for compulsory school credits, which is the basis for admission to upper secondary schools. The overall achievement grades are set by the teacher of the given subject. However, an exam which is based on a single performance, either written or oral, also remains on the pupils' diplomas. There is a vast media attention towards exam grades and it usually deals with the gap between overall achievement- and exam grades. The attention is

¹ <https://skoleporten.udir.no/> Accessed 12 February 2020.

linked to what could be seen as natural discrepancy between the two grades in a subject. What does it mean if the deviations are large? Is it teachers who cannot do their job? There are several studies regarding deviations of overall achievement- and exam grades (Hægeland et al. 2005; Gravaas et al. 2008; Galloway et al. 2011). However, there are few studies in Norway concerning teachers' own perception of these deviations (Prøitz and Borgen 2010; Hovdhaugen et al. 2018).

Research shows that the average grades for a written exam are generally somewhat lower than the overall achievement grades in the same subject (Gravaas et al. 2008). In itself, it is not a problem that there are differences between the pupils' overall achievement- and exam grades. This could be explained, for instance, by random circumstances, or that the school exam partly measures something different from what the overall achievement does (Gravaas et al. 2008). Examples of this may be the ability to concentrate and master pressure. However, this is a problem if the differences are systematic between schools. If we cannot trust that the overall achievement grades measure the pupils' academic level, this will have an impact on how pupils achieve a lower or a higher grade for a written exam. The grade difference could then indicate systematic differences in schools' assessment practices.

There are guidelines and criteria for what kind of achievements are required to achieve a given grade, but they are not very detailed. They may appear as open for interpretation by teachers. The reasons for this are that it could be difficult to give as clear guidelines that they alone ensure a uniform grading. For the individual teacher in a subject there is a considerable degree of freedom when overall achievement grades are to be set. There may therefore be a risk of deviations developing between schools in terms of what kind of achievement is required to get a given grade.

The thesis therefore aims to answer the following research questions:

- Are there deviations between overall achievement grades and exam grades in the English subject and what possible factors might influence this deviation in grading?
- What importance do teachers attach to the difference between the two forms of assessment?
- How do teachers approach grading in the English subject at Norwegian lower secondary schools?

- How do teachers' grading practices in the English subject correspond to the recommendations and regulations for grading?

The first research question will be answered by exploring public statistics from *Skoleporten* (2020) of average grade scores for both overall achievement- and examination assessments. The possible factor that might influence a deviation in grading will be addressed through theory on the subject and interviews with six English teachers. The interviews, in addition to theoretical background, will also be the main source in order to answer the second research question. The following elements presented by Harlen (2005) will be examined in order to answer the third research question: How do teachers collect evidence? What is considered to be relevant evidence in final grading and how do teachers interpret this evidence? (2005: 207). The last research question will be answered by exploring the legal requirements and recommendations from the Norwegian Directorate for Education and Training through white papers and reports.

This thesis assumes that there might be too much freedom in school practices when it comes to assessment in Norway. From experience as a student attending a teacher programme at a university, there is not enough focus on how to assess pupils and there might develop differences in how teachers assess pupils. The intent is to problematize the need for guidance when grading pupils so that every pupil is graded and assessed fairly according to the competence aims in LK06 and the new subject curriculum that will be implemented by fall 2020. The hope is therefore to shed light on the matter to make a difference for teachers and pupils in the future.

At first, Chapter 2 gives an overview of the theoretical background of assessment and grading practices which is also seen in context of the Norwegian school system. The methodology selected for this study is presented in Chapter 3, together with materials and data collection. A presentation of the materials collected is found in Chapter 4, divided into the three different materials collected; Interviews, exam assessment and statistics. Chapter 5 presents a discussion of the findings, which is divided according to the research questions devised for the present thesis. At last, concluding remarks are given in Chapter 6, followed by a reference list and relevant appendices.

2 THEORETICAL BACKGROUND

First, the theoretical background consists of an overview of general explanations of assessment before moving on to present previous international studies of teachers grading practices and relevant findings. Furthermore, there will be an overview of different assessment practices and their relations with the Norwegian educational system. To view this in context of the Norwegian education system with its laws and requirements there is a following section regarding the English subject curriculum, white papers and yearly reports from the Norwegian Directorate for Education and Training (which hereafter is referred to with its Norwegian acronym UDIR). Finally, there will be an overview of previous research conducted in Norway of deviations between overall achievement- and exam grades and teachers' grading practices.

2.1 Assessment

Assessment and evaluation are of high importance for both teachers and pupils. There is a huge responsibility relying on teachers to assess pupils' work. Brown and Abeywickrama (2010) explain that the term 'assessment' is sometimes misunderstood. Many tend to think of "*assessing* and *testing* as synonymous terms, but they are not" (Brown and Abeywickrama 2010: 3). The distinction between them becomes apparent as assessment is "appraising or estimating the level of magnitude of some attribute of a person" (Mousavi 2009: 36) and tests are "a subset, a genre of assessment techniques" (Brown and Abeywickrama 2010: 3).

Measurement and evaluation are terms which often lie somewhere in-between the terms of assessment and testing. Because of that, they are often used as synonyms of one or the other term. Bachman (1990) states that "It is only when the results of tests are used as a basis for making a decision that evaluation is involved" (1990: 22-23). One could argue from these explanations that to evaluate and to assess a pupil can be understood as synonymous.

However, a test score is an example of measurement, whereas evaluation is to value the result of those test scores.

2.1.1 Formative and summative assessment

It is common to distinguish between formative and summative in descriptions of assessment and evaluation. The terms formative and summative evaluation dates back to when Michael Scriven (1967) introduced the terms, more than 50 years ago. At the time, Scriven presented

formative evaluation as a tool to evaluate curriculums and the effectiveness of learning strategies. However, Bloom et al. (1971) were the first who extended the usage of these terms to its generally accepted current meaning in their *Handbook of formative and summative evaluation of student learning*. The term ‘evaluation’ was used in the title of their handbook. Nonetheless, their focus was primarily on student assessment. Black and William (2003) commented on the work of Bloom et al. (1971) and emphasized that “From their earliest use it was clear that the terms ‘formative’ and ‘summative’ applied not to the assessment themselves, but to the functions they served” (2003: 623).

Formative assessment is often seen as «assessment for learning», whereas summative assessment is seen as «assessment of learning» (Harlen and Gardner 2010: 27). Formative assessment is most frequently used for classroom assessment according to Brown and Abeywickrama (2010: 7). Brown and Abeywickrama (2010) describes formative assessment as “evaluating students in the process of “forming” their competencies and skills with the goal of helping them to continue that growth process.” (Brown and Abeywickrama 2010: 7) However, grading an exam or an overall achievement can be seen as a summative assessment seeing as it:

aims to measure, or summarize, what a student has grasped and typically occurs at the end of a course or unit of instruction. A summation of what a student has learned implies looking back and taking stock of how well that student has accomplished objectives, but it does not necessarily point the way to future progress. Final exams in a course and general proficiency exams are examples of summative assessment. Summative assessment often, but not always involves evaluation (decision making).

(Brown and Abeywickrama 2010: 7)

Summative assessment therefore often involves evaluation because teachers values the pupils’ test scores either for an overall achievement or an exam. Bøhn (2018) explains that “Such evaluation is often regarded as “high-stakes”, since it may have serious consequences for learners.” (Bøhn 2018: 235). Pupils might not be accepted into their preferred upper secondary school because they were one grade point short. It is therefore important that final assessments are as fair and dependable as possible.

2.1.2 International studies of grading and assessment practices

Internationally, it has long been pointed out that teachers set grades in varying degrees to what can be seen as recommended practice (Stiggins and Conclin 1992). Research shows that teachers set grades based on what they think are the consequences of grades, and that this sometimes happens at the expense of what a grade actually means. The difference between recommended and actual grading thus becomes a problem for the validity of the grades (Brookhart 1991). Brown and Abeywickrama (2010) states that “grades must be the most talked about topic in anyone’s school years” (2010: 318). Grading is explained as “quasi-measurement (since grades do not possess the characteristics of true measures) and is affected by a number of socially-driven factors whose influence is generally not well understood” (Yorke 2011: 251). Similarly, Brown and Abeywickrama (2010) made “some principles and guidelines for grading and evaluation” (2010: 337):

- Grading is not necessarily based on a universally accepted scale
- Grading is sometimes subjective and context-dependent
- Grading of tests is often done on a “curve.
- Grades reflect a teacher’s philosophy of grading
- Grades reflect an institutional philosophy of grading
- Cross-cultural variation in grading philosophies needs to be understood
- Grades often conform, by design, to a teacher’s expected distribution of students across a continuum

(Brown and Abeywickrama 2010: 337)

These principles show that grading is sometimes subjective. Likewise, Yorke (2011) believes that grading becomes more judgemental and less a matter of measurement “if broad categories are used as the basis of grading” (2011: 251), such as an overall achievement or an exam. He also argues that “grading tends to be treated inappropriately” (2011: 251). In Yorke’s (2011) study he puts emphasis on the need for “a sustained developmental effort at sectoral and institutional levels” (2011: 270).

In teacher education it is most commonly recommended to look exclusively at pupils’ achievement when grading (Dyrness and Dyrness 2008; McMillan 2008). Dyrness and Dyrness (2008) argue that grades should not be based on a pupil’s behaviour, instead they “should represent a measure of students’ knowledge of the subject and not be used to coerce a

certain behavior” (2008: 118). However, other studies show that factors such as work habits, effort and progress is considered alongside with the pupils’ achievement when deciding grades (Guskey, 2011; Yesbeck 2011). Guskey (2011) argues that teachers “combine aspects of students’ achievement, attitude, responsibility, effort, and behavior into a single grade that’s recorded on a report card – and no one questions it” (2011: 19).

Several researchers have pointed out that there may be a lack of theoretical coincidence between what teachers are recommended to do and what is done in school (Stiggins and Conclin 1992; Arasian and Jones 1993; Brookhart 1991, 1994, 2003). The hardest aspects to distinguish between are effort and achievement in subjects. When grades are to be awarded, it is recommended practice by school authorities to emphasize achievement and not effort. In the Norwegian context, we find this in Regulations to the Education Act, where it is emphasized that assessment in subjects must be done in relation to the overall competence goals in the curricula for the subjects (Regulations to the Education Act, § 3-3). However, many teachers find that separating effort and achievement is not a relevant recommendation for the classroom situation (Arasian and Jones 1993; Brookhart 1994, 2003).

Prøitz and Borgen (2010) show that teachers find it easier to assess fairly in the subject mathematics, which is characterized by scoring and test results. These characterizations leave less room for discretion (2010: 89). It is also documented internationally by Duncan and Noonan (2007) that teachers base their assessment in mathematics to a greater extent on cognitive characteristics such as measurable, professional factors compared to other subjects such as English, where effort and motivation can be included in the assessment. Teachers do not use score points in the Norwegian or English subject, but measures performance against the competence aims in the subject. The opportunity to assess in collaboration with other teachers is therefore important according to Prøitz and Borgen (2010: 89). This corresponds with Haugstveit et al. (2006), which emphasize the importance of collaboration when assessing. Nonetheless, Prøitz and Borgen (2010) show that teachers still describe the Norwegian subject as more difficult to assess a pupil fairly. It could be natural to imagine that these trends of assessment in the Norwegian subject can be transferred to the assessment in the subject English as well, seeing as they are constructed similarly but in a second language.

There are also many studies on teachers’ different practices in summative assessment and various factors that influence their practices (Cheng and Sun 2015; Tierney 2015; McMillan

et al. 2002). These studies tend to conclude that there is a lack of internal consistency between different teachers' grades and that teachers need more competence in assessment (Black et al. 2010; Brookhart 2013). Central to these studies are questions about how the authorities' increased use of different types of national tests or examinations for management, control and accountability, affect teachers' practices for assessment in new ways. This is a relevant question in Norway when expectations of conformity between overall achievement grades and exam grades are used to a greater extent, for example in the school owner's management of the school's work with development (Aasen et al. 2012; Mausethagen et al. 2018). Furthermore, some research also indicates that different contexts require different approaches when grading, suggesting that teachers may not always need more competence regarding assessment. Instead, there may be a need for more support and guidance in teachers' work on grading (Brookhart 2013; Wyatt-Smith and Klenowski 2013). Nonetheless, research from Brookhart (1991) has shown that teachers who have participated in assessment training still struggle with grading based on achievement alone.

It has been pointed out that teachers in the grading process should ideally function as judges, but research has shown that teachers also take on the role of attorney for their own pupils in grading. The role of being a judge falls naturally when being an external examiner, while the role of being an attorney is largely used when assessing own pupils. The attorney role also seems to have a double standard when it comes to fair judgement. Brookhart (1994) points out that average pupils "get as deserved", while low proficiency pupils are "given more slack" if there is any way to justify it. Bishop (1992) and Wilson (1996) make the distinction between the dual roles of teachers in classroom assessment: judges and coaches. Teachers should base pupils' grades on achievement exclusively as a judge. Judges use their judgment, justice, fairness and objectiveness when making a decision. On the other hand, coaches consider what their pupils find beneficial in order to develop and learn. Coaches bring in many non-achievement factors such as, development, encouragement, effort and improvement through grading. Bishop (1992) stated that teachers cannot be both coaches and judges at the same time. His suggestion was that a teacher should only take on the role of a judge in external assessment, and to focus on being a coach and mentor for their own pupils in classroom assessment.

2.2 Assessment practices

In order to understand how assessment practices can vary from teacher to teacher, from school to school, or be influenced by the school's pupil composition, it is useful to know different forms of pupil assessment. The summative assessment is considered to be the most central in this study due to the focus of final assessments. Further in this thesis, assessment practice will refer to the relative aspect of assessment; how the perception of which academic level behind a grade can vary, and how this not only varies between teachers but also between schools.

2.2.1 Ipsative assessment

An ipsative assessment “means that students are compared against themselves when being assessed” according to Bøhn (2018: 232). This type of assessment is most common to measure a pupil's progress and to motivate and “spur them on” (Hughes 2014: 5). This is similar to what Haugstveit et al. (2006) explains of individual-related assessment, where the pupil's achievements and products are assessed in relation to each pupil's prerequisites. One might assume that Haugstveit et al. (2006) and Bøhn (2018) talks of the same assessment but have different concepts of it. Ipsative assessments includes the pupils' effort and progression. For instance, a low-achieving pupil could receive more positive feedback than a high-achieving pupil when effort and progression is of importance in addition to achievement.

In individual-related assessment there is an opportunity to assess the pupil from his or her background and prerequisites. This is documented in part by Black and William (1998), who argue that teachers' assessment strategies vary in correlation with their pupils' performance level. This is also described by Brookhart (1994) and Prøitz and Borgen (2011). Haugstveit et al. (2006) claim that there are signs that “teachers give more general, social assessment to weak pupils, while good pupils receive more professionally relevant assessment.” (2006: 58, my translation). This can be explained by the partially incompatible relationship between customized training and neutral assessment. Nevertheless, neither ipsative- nor individual-related assessment is permitted to use in Norway in final assessments such as for overall achievements and exams. However, it can be used in continuous assessment situations in order to promote learning.

2.2.2 Norm-referenced assessment

The norm-referenced assessment allows the pupil's work to be compared to another pupil's work (Bøhn 2018: 233). If a pupil is being graded based on the quality of his or her peers' performance, it is called a norm-referenced assessment. Bøhn (2018) explains that this assessment form is most commonly used "to see how students rank in relation to each other" (2018: 233). A theoretical normal distribution is often used when norm-referenced assessment is involved. This would imply that "if the number of students were sufficiently large, such as on a national level, their marks, or scores would be evenly distributed along a bell-shaped curve." (Bøhn 2018: 233). Assessing pupils' performance in such a way was common from 1939 and onwards in Norway. A typical bell-shaped curve was presented by Bøhn (2018) and can be viewed in Figure 1.

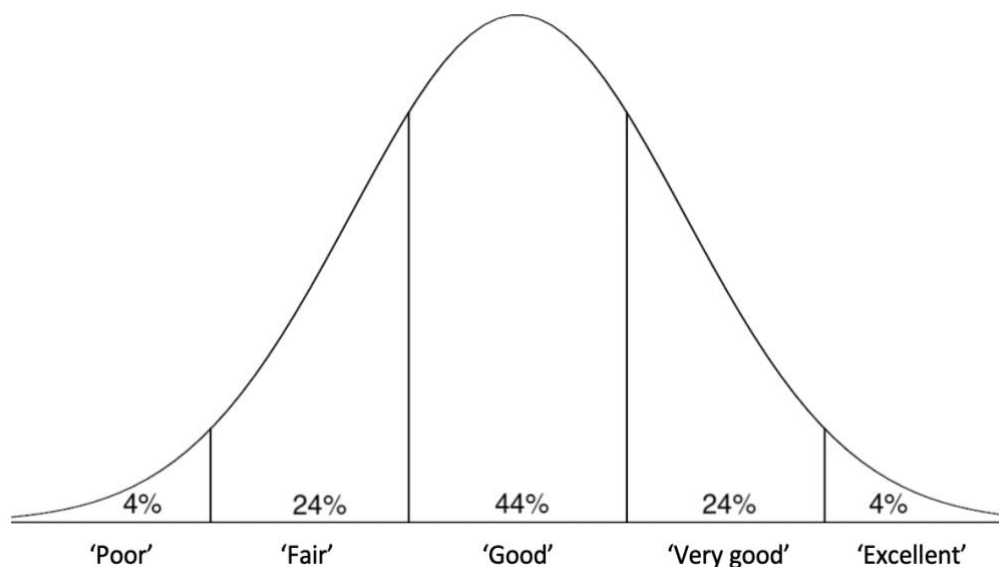


Figure 1: "The normal distribution in norm-referenced assessment" (Bøhn 2018: 233).

The 'Poor' was reserved for 4% of the pupils, 'Fair' for 24%, 'Good' for 44%, 'Very good' for 24%, and 'Excellent' for 4% pupils according to this system. From the 1970s onwards, this system was abandoned in the Norwegian educational context. A criterion-referenced assessment would take its place after the Education Act specified it in 2001.

2.2.3 Criterion-referenced assessment

Criterion-referenced assessment is different from ipsative- and norm-referenced assessment seeing as it does not allow pupils to be compared to each other, nor to themselves. In criterion-referenced assessment, the pupils' "performance is judged against some predefined criteria or standards" (Bøhn 2018: 234). This is clearly stated in Regulations to the Education

Act that pupils are to be assessed according to the competence aims of the subject curriculum (§ 3-3). The criterion-referenced assessment is therefore the recommended practice in the Norwegian educational system.

The criterion-referenced assessment is similar to what Haugstveit et al. (2006) explain about goal-related assessment where it relates to general goals a pupil should be able to do. Goal-related assessment is considered the most prevalent in lower secondary schools (Haugstveit et al. 2006: 227). However, it is not impossible that pupils who are perceived as disadvantaged, are assessed on the basis of their prerequisites also at the end of 10th grade. Nonetheless, studies show that teachers express uncertainty around assessment of pupils. Haugstveit et al. (2006) state that the relationship between individual- and goal-related assessment provides the basis for such uncertainty.

2.2.4 Informal and formal assessment

Dale and Wærness (2006) distinguish between the two forms of pupil assessment; formal and informal. The formal assessment reflects specific objectives. It focuses on the pupil's degree of goal achievement in the subject and is expressed in grade-based tests and exams, and with both overall achievement- and exam grades. The formal assessment is a rating-based assessment according to Dale and Wærness (2006). This rating implies that formal assessment is both goal-related in the sense that it is based on the pupil's insight and understanding, and group-related, because the grading gives room for comparison of pupils (2006: 61). Overall achievement grades can thus be set from a comparison with other pupils in the same context, such as a class or school.

On the other hand, informal assessment is assessment based on factors such as the pupil's individual prerequisites, background and relative level. Dale and Wærness (2006) argue that the two forms stand in contradiction to each other and makes teachers uncertain in their assessment of pupils (2006: 57-61). Teachers plan teaching and provide pupils with customized learning activities so that growth and achievement of learning goals can happen through guidance. Within this logic, achievement is jeopardized if the pupil does not participate in the activities. From this, Arasian and Jones (1993) argue that teachers' decisions regarding final assessment are part of a larger assessment context. Their research show that

informal assessments related to ranking pupils and teaching during the school year affect the more formal assessments teachers make.

2.3 The Norwegian context

A national standard for assessment of pupils' academic achievements has for a long time been non-existent in Norway. Only with the introduction of the Knowledge Promotion in 2006, there was a more detailed formulation of assessment guides included in the English subject curriculum. Before this, the teachers were to a far greater extent making their own framework for how grades should best reflect the pupils' achievements (Dale and Wærness 2006). A number of steering documents in the Norwegian school system, such as the subject curriculum, Regulations to the Education Act and reports from the government called white papers are intended to serve as important principles for teacher assessment of pupils and the performance of their role, and a reference for reasoning and reflection (UDIR 2007: 8). The present thesis has therefore chosen to have an overview of the English subject curriculum, white papers and yearly reports from UDIR called *Utdanningsspeilet* (the Mirror of education).

2.3.1 The English subject curriculum

A subject curriculum is a document which teachers should be well informed about seeing as it is in fact a regulation with a legal status in education. The English subject curriculum consists of an overview of the subject's purpose, main subject areas, teaching hours distributed per year, basic skills and competence aims. Teachers are required to base their teaching and assessment from all of these aspects. The current subject curriculum is valid from 01.08.2013-31.07.2021. The new English subject curriculum will be implemented by the start of fall 2020. Minister of Knowledge and Integration, Jan Tore Sanner stated that this will be the biggest change of the school's content since the Knowledge Promotion in 2006, according to press release Nr: 259-19 (Kunnskapsdepartementet 2019). UDIR has received more than 20,000 responses in total regarding the new curriculum. The new version of the English curriculum (ENG01-04) will be introduced gradually from 01.08.2020. 1-9th grade implement the new curriculum in schoolyear 2020/21, while the 10th grade implement it in 2021/22.

The current English subject curriculum has 30 competence aims, whereas the upcoming English subject curriculum consists of 19 competence aims. However, the biggest change

with the new curriculum is concerning the overall achievement grades. Currently, pupils get one overall achievement grade for written achievements and one for oral achievements in the English subject. The new curriculum states that the pupils shall receive one overall achievement grade instead of two. This might introduce some problems regarding how to plan assessments. Some teachers may favour oral assessments over written assessments due to the fact that correcting written work might be considered as more labour-intensive.

2.3.2 Overall achievements in the English subject

Overall achievement grades are given after year 10 in lower secondary schools in Norway. In UDIR’s report of overall achievement assessment, it is explained that the grades must be determined in accordance with the curriculum for the given subject (UDIR 2019). This requires, among other things, good understanding of the curriculum and means that teachers, along with colleagues, have to discuss and reflect on what competence in the subject is at different levels.

In Regulations to the Education Act (§ 3-3), it is explained that the basis for assessment in subjects are the competence aims in the curriculum. It is emphasized here that “the prerequisites of the individual, absence or conditions related to order and conduct of the pupil, apprentice or learning candidate should not be drawn into the assessment in subjects.” (Regulations to the Education Act, § 3-3, my translation). However, pupils must attend and participate actively in class in order to give the teacher a basis for assessing the pupil’s competence in the subject. This means that a “large absence or other special reasons may cause the teacher not to have sufficient grounds to give a half-year assessment with a grade or an overall achievement grade” (Regulations to the Education Act, § 3-3, my translation).

UDIR (2017) has made guiding national characteristics of goal achievement for overall achievements after year 10 in both oral and written English. Translations of them can be viewed in Figures 2 and 3.

| Oral English | | |
|--|---|--|
| Grade 2 | Grades 3 & 4 | Grades 5 & 6 |
| The pupil extracts content from texts and spoken English on certain topics and expresses own opinion on this | The pupil extracts content and details from different types of texts and spoken English on some topics and reflects on this | The pupil pulls extracts content and details from different types of texts and spoken English on different topics, discusses and reflects on the purpose of this |

| | | |
|--|---|---|
| The pupil participates in conversations and tells about academic and literary subjects to some extent adapted to communication situations | The pupil receives and gives input in conversation and disseminates content from academic and literary topics adapted to various communication situations | The pupil participates constructively in conversations and conveys academic literary topics in an independent way, adapted to various communication situations |
| The pupil participates in conversations with relevant content and own opinions | The pupil uses strategies to carry on conversations, explains and elaborates content | The pupil uses appropriate strategies to lead conversations further, elaborates and complements with new input |
| The pupil expresses himself/herself with a certain intonation, understandable pronunciation, a vocabulary that conveys content and some context on familiar topics in various communication situations | The pupil expresses himself/herself with clear intonation and pronunciation, a vocationally covering vocabulary, essentially good flow and context adapted to content, form and recipient in various communication situations | The pupil expresses himself/herself with good intonation and pronunciation, a general vocabulary, flow and context adapted content, form and recipient in various communication situations. |

Figure 2: Characteristics of goal achievement after year 10 in oral English (UDIR 2017, my translation).

| Written English | | |
|---|---|--|
| Grade 2 | Grades 3 & 4 | Grades 5 & 6 |
| The pupil finds information in texts of familiar topics and extracts content, lists some sources | The pupil finds information in various types of texts on familiar topics and extracts main content and details, lists the sources used | The pupil finds information in a wide variety of texts on various topics and extracts main content and essential details, refers to the sources in a verifiable way |
| The pupil reproduces cultural knowledge in own text production | The pupil uses cultural knowledge in own text production | The pupil uses cultural knowledge independently in own text production |
| The pupil writes, shapes and changes own texts according to input on text: content, structure and language | The pupil plans, writes, develops content and revises own texts based on input and some knowledge of and experience with text: content, structure and language | The pupil plans, writes, develops content and revises own texts based on input and good knowledge of and extensive experience with text: content, structure and language |
| The pupil expresses himself/herself with an understandable vocabulary, a certain structure and context on familiar topics to a certain extent adapted to purpose, recipient and digital form requirements | The pupil expresses himself/herself with a covering vocabulary, clear structure, text binding and context on various topics adapted to purpose, recipient and digital form requirements | The pupil expresses himself/herself with a general vocabulary consistent structure, varied text binding and clear context about a wide range of topics clearly adapted to purpose, recipient and digital form requirements |

Figure 3: Characteristics of goal achievement after year 10 in written English (UDIR 2017, my translation).

UDIR explains these characteristics as a description of the quality of competence in subjects:

The characteristics are based on the competence goal as described in the syllabuses. The characteristics are designed across main areas of the subject to express competence in the subject as a whole. The characteristics must therefore be seen in conjunction with the curricula.

(UDIR 2017, my translation)

These characteristics are meant to help teachers when setting overall achievement grades in the Norwegian school system. The characteristics are designed on three levels, grades 2, 3-4 and 5-6. As can be seen in Figures 2 and 3, there are small adjustments between the different levels. The lowest grade 1 is not presented which makes it harder to make a distinction whether a pupil has achieved the grade 1 or 2. Furthermore, it is problematic having the same description for two different grades.

Teachers are required to present these characteristics to their pupils early on in the teaching to make them understand what is required of them and what they are being assessed on. UDIR (2017) recommends using these guides throughout lower secondary school, and not just at the end of year 10. Collaboration between colleagues on competency goals and characteristics could contribute to a common understanding and common language about what pupils should learn, and what characterizes different levels of goal achievement according to UDIR (2017). Such an interpreting does not mean that all assessment must take place in a certain way, but that joint discussion and understanding of the basis of assessment may render fair assessment of the competence of each individual pupil.

2.3.3 Examinations in the English subject

UDIR (2013) describes exam grades and overall achievement grades as a final assessment in the English subject curriculum. After finishing year 10, pupils shall receive one overall achievement grade for oral performance and one for written work in the English subject.

The pupils may be selected for a written examination. The written examination is prepared and graded centrally. The pupils may also be selected for an oral examination. The oral examination is prepared and graded locally.

(UDIR 2013: 11)

However, UDIR has not decided if this will be the following practice in the upcoming curriculum. Whether or not there will be any changes in the new curriculum regarding the

exam arrangements in the English subject will be debated in a separate hearing and completed later by UDIR.

UDIR issues the written exam nationally in Norway. Two external examiners then assess the same written exam anonymously. The pupils who are exam candidates in the subject have one day of preparation prior to the exam. Candidates will then receive a booklet of preparation material with an overarching topic of the written exam. This booklet provides the pupils with texts and advice for preparation. The pupils may also bring textbooks, notes, handouts, previous exam papers and other relevant material with them for the written exam. Hasselgreen and Ørevik (2018) in *Teaching English in the 21st Century* explains that:

In line with the principles that assessment criteria should be transparent and known to the student, the directorate also issues an exam guide with national exams, explaining the perspectives from which exam papers will be assessed and specifying characteristics of achievement at different levels.

(Hasselgreen and Ørevik 2018: 367)

The examination guide is provided for teachers and pupils every year and usually consists of 10 pages with necessary information such as what is required from the pupils and how teachers are supposed to assess the written exams. There is a characteristics of goal achievement template that accompanies the examination guide, regarding the written exam in English. A written translation of the template can be viewed in Figure 4, with descriptions of characteristics of goal achievement to centrally given exams in English after year 10. It is meant to be used by external examiners when assessing examination papers. UDIR's examination guide (2020) emphasizes that the descriptions provided in the characteristics of goal achievement should contribute to a fair assessment and a common understanding of the requirements for an exam paper.

The grade 1 is not thoroughly described in the examination guide and should only be used in cases where the examination paper shows very low competence in the subject. UDIR (2020) explains that exam papers at this level are most often very short and relate to a small extent to the assignment instructions. They often show that pupils are lacking the English skills required to answer the exam tasks. If an exam paper turns out to be a copy or previously published material, it will also be graded 1 according to UDIR's examination guide (2020).

| | The grade 2 The examination paper | The grade 3 & 4 The examination paper | The grade 5 & 6 The examination paper | |
|--|---|--|---|--|
| First impression | - communicates in a simple way based on the tasks given | - communicates for the most part based on the tasks given | - communicates holistically based on the tasks given | |
| Main areas of the subject curriculum in English: Language learning written communication, Culture, society and literature | content | - matches in some areas with the requirements of the assignment - reproduces knowledge where the task requires it - gives some examples | - matches mostly with the requirements of the task - uses knowledge where the task asks for it - describes and gives examples | - corresponds holistically with task requirements - uses knowledge on an independent basis where the task asks for it - discusses and gives relevant examples |
| | structure | - has simple structure and some context - is to a certain extent adapted to the purpose and digital form requirements based on the task order - has simple text binding on sentence and/or text level | - has structure and context - is mostly adapted to the purpose and digital form requirements based on the task order - has text binding which mostly highlights the content at sentence and text level | - has a thorough structure and context - is customized for purpose and digital form requirements based on the task order - has varied text binding which highlights content at sentence and text level |
| | Language | - has simple and limited vocabulary - has certain subject terms where the task asks for it - the language is to a certain extent adapted to the purpose of the task - has simple sentence structure | - has a vocabulary that covers some topics - has subject terms where the task asks for it - the language is for the most part customized to the purpose of the task - has some variation in sentence structure | - has a varied vocabulary that covers different topics - uses subject terms in a relevant way where the task asks for it - the language is adapted to the purpose of the task - has variation in sentence structure |
| | Formal skills | - has spelling and word bending which makes the text understandable - has to some extent English sentence structure | - has spelling and word bending which despite errors, does not hinder communication - has mostly English sentence structure | - has central patterns for spelling and word inflection which despite some errors, give good flow in communication - has a thorough English sentence structure |
| | Use of sources | - uses sources to a limited extent as a basis for writing where the task asks for it - states sources, if any | - uses sources as a basis where the task asks for it - states used sources mostly in a verifiable way | - uses relevant sources as basis for writing in an independent way where the task asks for it - refers to sources used in a verifiable way |
| | Overall assessment | Grade 2 expresses that the pupil has low competence in the subject. | Grade 3 expresses that the pupil has fair competence in the subject. Grade 4 expresses that the pupil has good competence in the subject. | Grade 5 expresses that the pupil has very good competence in the subject. Grade 6 expresses that the pupil has exceptionally good competence in the subject. |
| Grade 1 expresses that the exam paper shows very low competence in the subject, lower than what is described above. | | | | |

Figure 4: Characteristics of goal achievement for ENG0012/ENG0025 at centrally given exam (UDIR 2020, my translation).

Bøhn (2015) calls these characteristics of goal achievement templates for rating scales, seeing as teachers use them to rate pupils accordingly. When looking into the construct of the oral exam, there are only rating scales developed at a local level. Bøhn (2015) states that these rating scales hold reliability and validity issues, seeing as there is no national common rating scale in oral English. The oral exam is usually assessed by the pupil's teacher and an external examiner, based on a rating scale produced locally by either the school or the county authorities. The candidates for an oral exam will have two days to prepare a presentation on a given topic. However, they can be questioned on other competence aims of the subject curriculum after the presentation. As a result, the content assessed at an oral exam may be much more demanding compared to the written exam, seeing as candidates have both preparation material, notes and handouts available in a written exam. Bøhn's (2015) study of teachers' assessments of a pupil's performance at an oral exam, found a common ground among teachers regarding the criteria and constructs. On the other hand, there were varying opinions as to their relative importance. Bøhn (2015) therefore suggested that it would strengthen the validity of score interpretations to have a common rating scale in oral exams as well (2015: 9).

2.3.4 White papers

White papers (Meld.St.) are matters drawn up by the government for the purpose of presenting it to the Norwegian Parliament. They are usually in the form of a report and contain descriptions of work carried out in a particular field and future policy. White papers are presented in this thesis because they represent a status reproduction and dissemination of what the authorities consider to be important for the political and administrative control of the education system in Norway. The four white papers Meld.St. nr.30 (2003-2004), Meld.St. nr.16 (2006-2007), Meld.St. nr.20 (2012-2013) and Meld.St. nr.28 (2015-2016) in the field of education have been selected to include, due to their relevance in referring to the role of assessment.

In Meld.St. nr.30 (2003-2004), it is clearly stated that there are two forms of summative assessments: overall achievements and examinations. This report described exams in primary education as "especially quality assured tests that are centrally or locally/regionally administered" (2003-2004: 37, my translation). The report did not provide any other explicit information of the role exams have beyond this description. This was made a little clearer in Meld.St. nr.16 (2006-2007). This document presents an exam as a final assessment which is

meant to inform society, educational institutions and future employers of the competence acquired by the pupil (2006-2007: 78). A more extensive discussion of how the exam system works appeared in Meld.St. nr.20 (2012-2013):

The draft scheme means that the students should not have exams in all subjects, but that they should be exam-prepared in the subjects where the exam is a possible final assessment next to the overall achievement grade.

(Meld.St. nr.20 2012-2013: 65-66, my translation)

This gives a clearer explanation of the relationship between exams and overall achievement as it explains how the exam serves as an incentive for pupils to work well in all subjects, so they are prepared in case they are picked for an exam. This could imply that the exam has a managing role on how to ensure the widest possible basis for certification, which is carried out through the overall achievement assessment. Furthermore, Meld.St. nr.20 (2012-2013) brought up the importance of examiner training, first and foremost with a view to ensure that the examination grades are set in a fair manner:

National authorities engage practicing teachers to prepare centrally provided exams. The practicing teachers are also commissioned to be examiners. In order to ensure a fair assessment, it is important that the examiners evaluate responses as equally as possible. Therefore, emphasis is placed on good examiner training in all subjects. In the Knowledge Promotion, guidelines have been developed with the characteristics of goal achievement for examinations in all subjects. The guides are used in examiner training. Pupils are also encouraged to get to know them before the exam. Efforts to quality assure, administer and assess the exam are important for the pupils' legal security.

(Meld.St. nr.20 2012-2013: 66, my translation)

It is stated here that the exam has a role in supporting the pupils' learning by making them familiar with the guidelines containing characteristics of goal achievement. Overall, Meld.St. nr.20 (2012-2013) addresses the interaction between the exam's roles in management, certification and support of learning and teaching. The Ministry of Education and Research provided a more explicit discussion of the different roles of the exam system in Meld.St. nr.28 (2015-2016):

The ministry believes the current examination system fulfils several important functions. The exam is a form of quality assurance for the pupils because they receive an external assessment of their subject competence. Subjects with a centrally granted exam have identical set of tasks for everyone who comes up in the subject. It helps to give the pupils a more equal diploma because the grades are set on the same assessment basis. The exam results can also help teachers and the school to control and develop their own assessment work. Grades given for the exam are a feedback to the school on how external examiners evaluate pupils' exam performance, it can help the school and teachers adjust their own practice and achieve a fairer assessment practice. This applies to both the centrally and locally given exams.

(Meld.St. nr.28 2015-2016: 62-63, my translation)

This report expresses an expectation that the schools and teachers use the examination results to adjust the practice's final assessment. This is another expression of the fact that exams are used to guide schools' work on certification. The report Meld.St. nr.28 (2015-2016), states that surveys have been made about the relationship between final assessments in subjects and exams. The surveys showed that it is a relatively widespread practice that school owners and schools compare the schools' grades with exam grades in subjects to assess whether there are systematic deviations between them over time. The report makes a clear distinction between a final assessment grade and an exam grade:

It is important to emphasize that exam- and overall achievement grades are two different expressions of competence. An examination grade will be set on a more limited basis of assessment than an overall achievement grade. An exam represents one or very few forms of assessment (written, practical and/or oral) and it is an individual case. Nationally, the average exam grades for a centrally granted exam are usually slightly below the average overall achievement grades. If such a comparison is to be appropriate, it is the Ministry's assessment one must look at if differences over time should systematically deviate from the national average difference between overall achievement- and exam grades. Permanent deviations should be a warning to the school owner and school principal that it is necessary to change existing assessment practices. However, such analyses are not enough alone, but should be one of several sources of knowledge about own practice for use in a school's development work.

(Meld.St. nr.28 2015-2016: 63, my translation)

This is a clear example of how assessment, student results and learning outcomes have been given a far more central role in Norwegian education policy after the introduction of the

Knowledge Promotion (Mausethagen et al. 2018). The Ministry seems to be concerned with ensuring a sounder use of exam results, seeing that systematic deviations are to be regarded as a notice and should introduce changes in the practice of overall achievement assessments. The report suggests that the subjects Norwegian, maths and English should be mandatory for all pupils to have a written exam in (Meld.St. nr. 28 2015-2016: 64). By doing so, all pupils will have the same number of grades on their diploma. However, this suggestion has not been set in motion.

Tveit and Olsen (2018) analysed the four white papers presented above, and their findings lead to the conclusion that there has been very low awareness politically of the exam's roles beyond certification. They believe that there are a number of roles in the primary education exam, which should be clarified and defined as formal purposes of the exam system. They move on to say that "through such clarification, the exam can be further developed into a tool that is better adapted to these many roles" (2018: 24, my translation).

2.3.5 Utdanningsspeilet

The trends and themes highlighted in white papers from the government can be found in the annual reports from UDIR called *Utdanningsspeilet*. The researcher has looked into these reports from 2008-2019. Based on an initial review, the reports from 2008, 2009, 2011, 2014, 2016, 2017, 2019 were selected to include due to their relevance with final assessments.

The report from 2008 referred to surveys conducted by Statistics Norway (Hægeland et al. 2005) which identified some deviations between grades based on a pupil's social background. Their findings were that a pupil who had classmates with an advantageous social background got lower overall achievement grades, whereas it seemed to be easier to achieve a higher overall achievement grade if the pupil's classmates had a less favourable social background (Utdanningsspeilet 2008: 105). However, it was emphasized that this was not the case regarding exam grades. Analyses of school results from 2007 were also presented (Gravaas et al. 2008), which came to the same conclusions. The researchers came to this conclusion by comparing overall achievement- and exam grades. UDIR put great emphasis on such studies in their annual reports in the following years.

UDIR was also concerned with the relationship between overall achievement grades and exam grades in *Utdanningsspeilet* (2009). This report showed that the average of grades is generally higher for overall achievement grades than for a written exam, while oral exams had the highest average of grades (2009: 50). Ongoing work which UDIR received from the Norwegian Ministry of Education and Research from 2007 was also presented in their report. This work focused on initiating a number of measures to strengthen the assessment practice in Norway (2009: 96). UDIR also pointed to the work on assessment guidelines with characteristics of goal achievement, and that this has been compiled in most examination subjects (*Utdanningsspeilet* 2009: 104). These assessment guides “should have a learning-enhancing effect by allowing teachers to communicate the characteristics to the pupils before an exam.” (2009: 104, my translation) Consequently, emphasis was also placed on the role of the exam in supporting learning.

Utdanningsspeilet (2011) also referred to research done by looking at schools’ practices of overall achievement grading, where exam grades constituted the reference basis. This report gave the conclusion that there was a “significant variation in how elementary schools set overall achievement grades” (2011: 64, my translation). It was also emphasized that any written exam of five hours or an oral exam of half an hour could never show the same broad competence of a pupil as an overall achievement grade. This could be seen as an expression of the weaknesses in the role of the exam. Furthermore, it was stated that “since the examination is conducted with external assessment, the examination has an element of external quality assurance in it” (2011: 112, my translation). This implies that the exam’s external assessment contributes to a more quality assured form of certification than overall achievement assessments. It is also pointed out in this report that teachers do not get a view of their pupils’ exam grades before setting their overall achievement grade, meaning that the exam does not have any adjusting effect on the overall achievement grade of a pupil (2011: 113).

UDIR introduced a new way of looking at results in *Utdanningsspeilet* (2014), by comparing results from national tests and exams. They pointed out an improvement the last two years of the difference in overall achievement grades and exam grades. The difference between written exam grades and overall achievement grades had been reduced. This was explained by UDIR, saying that:

A larger proportion of pupils are given the same grade in their exam as their final achievement in the subject. At the same time, the proportion has increased of those who receive a higher grade on their exam than their final achievement, and there are fewer who receive a lower grade.

(Utdanningspeilet 2014: 76, my translation)

The report also referred to new research from Statistics Norway (2013), which showed that half of the pupils who were up for an exam got a lower grade for their exam than their overall achievement in a given subject. These two statements from *Utdanningspeilet* (2014) seem to be contradictory. *Utdanningspeilet* (2017) gave a more thorough explanation of final assessments, where a distinction is drawn between what an exam- and an overall achievement grade measure:

Both the overall achievement- and the exam grade in a subject are determined on the basis of the overall competence objectives in the curriculum, but they measure different aspects of the pupil's competence:

- The overall achievement grade must show the pupil's overall competence after completing the training and should be assessed on such a broad basis as possible.
- An examination grade shall express the pupil's overall competence as expressed in the examination.

(Utdanningspeilet 2017, Chapter 5.2, my translation)

There is reason to assume that this clarification is based on the fact that school owners and school leaders compare exam- and overall achievement grades to guide the school's assessment practice, and that this could be considered as problematic. With all reports gathered we see a pattern of UDIR stating that the exam has an important role in management of the school's work on certification. The reports from 2007-2017 show a difference in how the roles of exam have been described. The exam's role had a more descriptive discussion of school and school owners' use of exams to guide the overall achievement grade in *Utdanningspeilet* (2008, 2009, 2011). In *Utdanningspeilet* (2014), UDIR uses exam data to make comparisons and encourages schools and school owners to do the same. This use of exam results is however problematized in *Utdanningspeilet* (2016, 2017). *Utdanningspeilet* (2019) did not describe anything exam- or overall achievement assessment related other than that "In English, the proportion of pupils at the two lowest levels of mastery has decreased from 28 per cent in 2014 to 26 per cent in 2019." (Utdanningspeilet 2019, my translation).

2.4 Norwegian studies of deviations between overall achievement- and exam grades

From the evaluation of Reform 97, it emerged that teachers tend to evaluate pupils in different scales (Dale and Wærness 2006: 115). Dale and Wærness (2006) outline how the same prerequisites and level of knowledge from two different school classes can still lead to different assessments when these are conducted by two different teachers, because the assessment criteria are too vague (2006: 160). The evaluation of Reform 97 showed that teachers often asked unclear academic requirements, especially at lower secondary schools according to Dale and Wærness (2006). They point to “Mønsterplanen” of 1987 (M87) which emphasized that “in a school for all, the assessment of student work should not unilaterally revolve around the knowledge and skills that students acquire” (Dale and Wærness 2006: 56, my translation). This is in line with 1970s school politics, which was characterized by growing criticism of what was perceived as a “goal-means pedagogy” (2006: 52-55, my translation). Introducing new guidelines may not automatically change teachers’ attitudes and practical approaches, neither regarding their own role nor assessment methods. Dale and Wærness (2006) argue that this underlying attitude contributes to emphasizing the pupils’ prerequisites when assessing their academic level (2006: 56-59).

Hægeland et al. (2005) looked at grades among graduates from both lower- and upper secondary schools in Norway from the schoolyear 2003/2004. They used the deviations between overall achievement end exam grades at an individual level as a dependent variable. Hægeland et al. (2005) examined whether these deviations varied systematically between schools, both among graduates from lower and upper secondary schools. This led them to find a systematic variation of difference in grades between schools. They also found a tendency that schools with higher overall achievement- and exam grades in 2004, had so in 2002 and 2003 as well. However, they pointed out that this trend is weak, and can be explained by the fact that a smaller number of schools appears to have systematic divergent grading practices (Hægeland et al. 2005: 52).

Gravaas et al. (2008) examined primary school points, overall achievement- and exam grades among all pupils in Norway who graduated from lower secondary school from 2002 to 2007, with main emphasis on those who graduated in 2007. They also looked at relationships between grades, school size and ownership (private or public school), as well as social background, parent’s education, immigration background and gender. By comparing the

pupil's overall achievement- and exam grades at different schools, they revealed significant school-level differences. They interpreted this as a sign that some schools had a special kind or strict assessment practice when overall achievement grades were set and that this trend was systematic. Schools with higher overall achievement- than exam grades in 2007 had the same pattern in 2006 as well. However, they emphasized that this tendency was weak. Moreover, they found that pupils having parents without education above primary school, achieved lower grades at an exam than the pupils having parents with higher education. Social background thus plays a role not only for achievement in itself, but also for the deviations between overall achievement- and exam grades. They pointed out that this was in line with findings documented by Hægeland et al. (2005) (2008: 53). Gravaas et al. (2008) reported that boys in general, achieved lower grades for their exam than their overall achievement as opposed to girls.

Galloway et al. (2011) examined the deviations between overall achievement- and exam grades in Oslo, Bergen, Stavanger/Sandnes and Trondheim in their report *Karakterpraksis i grunnskoler* from schoolyear 2001/2002 until 2007/2008. Their report was based on previous analyses of differences in overall achievement- and exam grades (such as Gravaas et al. 2008; Hægeland et al. 2005). They looked at the relationship between the size of the schools and the difference in grades and found that pupils at small schools achieved lower grades in an exam to a greater extent than pupils at large schools.

In addition, they reported that schools with low average exam grades more often overestimated their pupils in determining their overall achievement grade, while the opposite was true for schools with high average exam grades. This would imply that teachers' grading practices rely on the level of the pupils at the individual school. The variations between the lower secondary schools exist across subjects and teachers. This implies that schools which give high or low grades relative to exam grades in one subject, also do so in another. Galloway et al. (2011: 35) interpreted this as a sign of relative characterization at schools but emphasized that the findings could also be explained by other conditions. A possible explanation of a decline in exam grades at small schools, may be that since small schools have fewer exam candidates, a single external examiner could give especially high or low grades, which have far greater significance for small schools than large schools (2011: 12). Galloway et al. (2011) expressed that there are considerable variations in the levels of grading practices in middle schools. Prøitz (2013) shared their view of the grading practices in

Norway, saying that “It is well documented that teachers assign grades in ways that diverge from recommended grading practices, sometimes at the expense of their interpretability” (2013: 555).

Hovdhaugen et al. (2018) interviewed teachers from 8 different schools in Norway to “highlight the differences between grading based on summative classroom assessment and being an external examiner” (2018: 2). It might be worth to mention that all of the interviewed teachers answered no to whether they had received formal training in assessment and grading as part of their education, or at the school they worked at. Hovdhaugen et al. (2018) found differences in how the teachers thought when giving final overall achievement grades. Some teachers looked at the pupil’s progress and final grades of the second term, while some felt that the first term was also of significance to include (2018: 10).

Tveit and Olsen (2018) looked into the roles of national exams in the Norwegian school system. Their conclusion was that the exam’s role had not been complementary enough in any of the reports from the Parliament during the period of 2003-2017 (2018: 24). Hovdhaugen et al. (2018) found several flaws with this idea, stating that the overall achievement grade is based on a teacher-pupil relationship characterized by the teacher’s in-depth knowledge of the pupil as a person, while the exam grade is conducted with mutual anonymity. Both overall achievement grades and exam grades could be seen to have strengths and weaknesses according to Hovdhaugen et al. (2018: 17).

3 METHODOLOGY AND MATERIALS

A mixed-method approach has been used for the purpose of this thesis and “involves both quantitative and qualitative research” (Johnson et al. 2007: 121). The present thesis involves a qualitative approach by interviewing six different teachers working at lower secondary schools in the western part of Norway. The six teachers interviewed have also assessed and graded an exam paper from the schoolyear 2017/2018 to get a view of their grading practice. On the other hand, there has also been a quantitative approach by analysing statistics of the average grade score of overall achievements and comparing them to the average grade score of final examinations at lower secondary schools in Norway by using the platform *Skoleporten* (2020). These statistics have been analysed to get an overview of the average deviations between overall achievement- and final exam grades. The statistics have been looked at both nationally and school wise, looking into three lower secondary schools in Norway. These schools will be referred to as Schools A, B or C hereafter. To get an even more detailed view, the researcher collected anonymous individual overall achievement- and exam grades in the English subject from Schools A and C. These statistics will be analysed to make a distinction between individuals, in addition to the statistics from *Skoleporten* (2020).

When combining qualitative and quantitative research, it provides a more elaborated understanding of the interviews when looking at statistics. Creswell (2014) states that the value of using multiple methods in research “resided from the idea that all methods had bias and weaknesses, and the collection of both quantitative and qualitative data neutralized the weaknesses of each form of data” (2014: 14-15). The present thesis would in the eyes of the researcher not get sufficient data by only using a qualitative or a quantitative approach. By looking at the actual grades it is possible to get a clearer picture of the deviations between overall achievement- and exam grades. However, in order to answer the research questions of what teachers emphasize when assessing pupils, it was also necessary to include interviews with teachers.

3.1 Qualitative data

The qualitative research method which has been implemented in this thesis are interviews and exam assessments done with and by six English teachers from three lower secondary schools in the western part of Norway. It would have been preferable to include more schools, but it proved to be difficult to get teachers to volunteer. A convenience sampling has therefore been

used in order to recruit participants. Dörnyei (2007: 98-99) gives the following explanation of convenience sampling:

[A]n important criterion of sample selection is the convenience of the researcher: members of the target population are selected for the purpose of the study if they meet certain practical criteria, such as geographical proximity, availability at a certain time, easy accessibility, or the willingness to volunteer.

In the present study, both geographical constraints and availability at a certain time were of importance when reaching out to potential participants. Principals from eight different lower secondary schools were emailed and asked to pass the email along to English teachers working for them. All principals responded by forwarding the email to possible candidates. There were, however, few teachers who responded to the email. Some of the candidates that did respond said that they were unable to help due to a busy schedule at the moment. In the end, six teachers from three different schools agreed to take part in the study.

The six teachers participated in semi structured interviews, meaning that “the interviewer provides guidance and direction (hence the '-structured' part in the name), but is also keen to follow up interesting developments and to let the interviewee elaborate on certain issues (hence the 'semi-' part)” (Dörnyei 2007: 136). This implies that the interviewer has a set of questions to ask all participants, though not always in the same wording or order. There are many advantages with having semi-structured interviews. One important advantage is that a flexible approach towards the interviewees is possible. It is also possible to go more in-depth in a semi-structured interview than in a structured interview, which has a rigorous set of questions. This implies that a semi-structured interview is more open for new ideas than a structured interview. An unstructured interview with no detailed interview guide could have made the participants more relaxed, but the researcher felt it was necessary to have an interview guide as a basis for the conversation in order to receive all participants’ views on their assessment- and grading practices.

An interview guide was therefore used based on a previous report by Hovdhaugen et al. (2014: 85-87). This report documents the project *Karakterpraksis i offentlige og private videregående skoler*, which the Nordic institute for Studies in Innovation, Research and Education (NIFU) carried out on behalf of UDIR. Relevant questions from this report was

implemented in the present thesis seeing as they have been tested before by a credible source. The interview guide consists of 24 questions, whereof six have follow up questions. 11 questions were written as complementary questions. However, most of the questions were designed in order to receive complementary answers, meaning it was expected that even though several questions did not specifically ask for it, the participant would still give an in-depth answer. This implies that the questions were open ended to receive subjective experiences, feelings and opinions from the teachers. The interview guide was constructed both in English (see Appendix 1) and Norwegian (see Appendix 2), giving the participants a choice of language for the interview. This option was provided by the researcher in order for the participants to be as comfortable as possible when being interviewed. Some might find it difficult to express what they want to convey when talking in a second language. The researcher did not want this to be a factor in how they expressed their opinions. Seemingly, all participants chose to conduct the interview in Norwegian.

The project was notified to the NSD – Norwegian Centre for Research Data² to ensure that legal requirements and recommendations were being followed. A qualitative approach regarding interviews is often personal and some ethical considerations has therefore been made because anonymity and privacy are important issues. The participants have knowingly been audio-recorded for later transcription. They were given a consent sheet to sign prior to the interview, which can be found in Appendix 9. The transcriptions and results from the interviews are therefore presented with full anonymity.

The researcher met all participants in person and conducted the interview with them while audio-recording. The interviews conducted lasted about 30-60 minutes. Four out of the six participants were met at their school by the researcher. These interviews were held in group rooms at their school building. Participant 5 was in maternity leave and was therefore visited in her home due to convenience. The interview was interrupted by Participant 5 having to check up on her child, but the interview proceeded as normal afterwards and the researcher did not see this as an interfering factor. Consequently, the researcher chose to include this interview. Participant 4 was interviewed in a public place; the meeting took place at a restaurant for her convenience. However, the interview took place at a time where there were

² Approval number: 690888

few customers around, which gave the interview the privacy it needed. This was also not seen as an interfering factor, which made it possible to include this interview as well.

The interviews were recorded using an audio-recording device. Those recordings were subsequently used in the transcription of the interviews. After the interviews were transcribed, the audio-recordings were deleted from the device as stated by the legal requirements from NSD. Summaries of the interviews can be found in Appendices 3-8. The researcher chose to write edited transcriptions of the interviews due to them all being in Norwegian. The interviews are therefore presented as summarised translations in Chapter 4. The researcher would have to transcribe everything and then translate it to English if a verbatim transcription had been used. The reason for not doing so, is because it is very time consuming and would not be possible within the framework of this master's thesis. Also, it would not necessarily be more accurate due to the fact that the interviews had to be translated either way.

Prior to conducting the interviews, Participants 1-6 were asked to write an assessment and grade an examination paper written by a pupil from School C. All participants were asked to comment on what was emphasized when grading the paper, and to justify their grade. The researcher wrote that it was preferable that the participants did this before the interviews took place. The written exam paper was from the schoolyear 2017/2018. This was collected through asking a principal for permission to use it anonymously in research. The examination paper was written with a computer whereas name and candidate number have been removed from the paper before using it in research. Those where the legal requirements for using it from the NSD. The exam paper can be found in Appendix 10.

3.1.1 Participants

The six participants were all teachers working at lower secondary schools in western Norway. The researcher tried to choose interview participants of different ages to get opinions across ages. The six Participants are from three different schools. Participants 1 and 2 are from school A, Participant 3 is from school B and Participants 4-6 are from school C. All participants turned out to be female in the present study.

Participant 1 was 23 years old and graduated from the university as a primary school teacher in spring 2019. She had worked as a teacher for 5 months. She was currently a teacher working with 8th grade pupils. Participant 2 was 46 years old and had been working as a

teacher for 11 years. Prior to this, she was a hairdresser. Participant 3 was 40 years old and had worked as a teacher for 14 years, starting in 2005. Participant 4 was 56 years old and had worked as a teacher for 16 years. Prior to this she had been an engineer working in the oil industry. Participant 4 believed that her background might make her different from other teachers in how they perceived the English subject. She noticed this when discussing with other teachers. They had different views of what was interesting to teach the pupils. Participant 5 was 28 years old and had worked as a teacher for 3 years. She was currently at home in maternity leave. Participant 6 was 61 years old and started working as a teacher in 1981. She had been an English teacher since 1995. This would imply that she has been working as a teacher for 39 years and as an English teacher for 25 years.

3.2 Quantitative data

Statistics from UDIR (Skoleporten 2020) are included in the analysis in order to get a broader view of deviations between overall achievement- and exam grades. There are two versions of *Skoleporten* (2020), one public for all, and one for school owners and school leaders which have full access to own results. The statistics used for the present thesis is the version which is public for everyone to view. Statistics of overall achievement- and exam grades from the schoolyears 2014/2015-2018/2019 will be analysed and discussed, both oral and written English. The presentation of these statistics is done graphically in diagrams, showing the average grade for the two assessments. The average overall achievement grades are compared to the average exam grades.

There are some limitations in the statistical material that must be taken into consideration in the analysis. Almost all pupils get an overall achievement grade. However, some pupils may receive an exemption from assessment if they have an individual training plan or have recently moved to Norway (Regulations to the Education Act, § 3-21). These pupils will not have an exam in the given subject either and are therefore not included in the overall achievement grade numbers. Furthermore, not all pupils who get an overall achievement grade, get an exam grade in English. This is true for both the written and oral English exam, due to the fact that only a certain number of pupils are normally selected for examination. All pupils are therefore included in the numbers for the average overall achievement grade in English, while the numbers for the average exam grades only include the ones who were selected for the exams.

Some schools have smaller pupil compositions than others. When a pupil composition at a given school is very low, some numbers are excluded from the public and not accounted for in the statistics. *Skoleporten* (2020) explains that if the number behind an indicator is 1, 2, 3, 4 or 5, then the number should be excluded. The indicator stated could be, for instance, the number of pupils, the average grade or the percentage of pupils. Furthermore, if all pupils for example at a school or in a municipality have the same grade, then the number should be excluded. If a number is excluded, other numbers that can be used to derive the excluded number must also be excluded. As a result, there are many factors to consider when analysing the statistics from *Skoleporten* (2020).

Seeing as the statistics from *Skoleporten* (2020) only shows an average of all overall achievement grades and the average of exam grades for those who had an exam in either oral or written English, the researcher wanted to look deeper into these numbers and find explanations for them. The researcher hoped to get a more detailed conception of the individual deviations' pupils receive. An average grade score can tell something entirely different than the reality, seeing as any given average is vulnerable and could be influenced by a small number of pupils receiving either a very high or low grade. As a result, the researcher contacted the principals of Schools A-C. They were asked to provide anonymous individual overall achievement- and exam grades for pupils at their school, who either had an oral or written exam in the English subject, in schoolyear 2018/2019. Both Schools A and C provided such information resulting in 26 pupils who had an oral examination, and 87 pupils who had a written examination in the schoolyear 2018/2019. School B did not have any pupils who had a written or an oral examination the given schoolyear and is therefore not included in the individual statistics. All pupils are presented with full anonymity and are called A1, A2 etc. from School A and C1, C2 etc. from School C.

4 PRESENTATION OF FINDINGS

The present study used both qualitative and quantitative data as mentioned in the methodology chapter (see p.27). The presentation of findings chapter is therefore divided into four subsections; 1) interviews, 2) exam assessment, 3) statistics from *Skoleporten* (2020) and 4) individual statistics. This division was made in order to view the different findings separately before putting them into context in the discussion chapter. As mentioned in the methodology chapter, the analysis is based on summarised translations (see p.30). This implies that the material presented in this chapter are summarised answers given by Participants 1-6. All material presented in subsections 4.1 and 4.2 are therefore the participants’ words and not the researcher’s unless stated otherwise.

4.1 Interviews

The interview guide consists of four main categories (see Appendices 1-2). The category regarding the participants background information has already been presented in the methodology chapter (see Section 3.1.1 p.30-31). The material has been further divided in the presentation of findings to distinguish between certain aspects. Moreover, the questions are grouped in a different manner here than in the interview guide because of that. The presentation of the interviews is therefore divided into five categories: 1) teacher experience, 2) grading practice, 3) teachers’ support when grading, 4) overall achievement grades and 5) the relationship between overall achievement- and exam grades. The questions answered from each category are presented in Figure 5. Continuous summaries of each participant’s interview can be found in Appendices 3-8. Questions 13, 20 and 21 will not be presented in the analysis. Question 13 was not emphasized seeing as all participants managed to provide answers for this question in questions 11 and 12. Questions 20 and 21 yielded no remarkable material to present and are therefore not included in the analysis.

Interview questions

| | |
|----------------------------|---|
| Teacher Experience: | 1. How much experience do you have with setting overall achievement grades in lower secondary schools? |
| | 2. Did you feel, after completing your teacher education that you were qualified for grading assessments? |
| | 3. Have you participated in courses/training in assessment/grading? Can you describe the experience? |
| | 4. Have you been an external examiner in English at a lower secondary school? |

| | |
|---|---|
| | <ul style="list-style-type: none"> a) In oral English? b) In written English? |
| | 5. Have you participated in examiner training? Could you describe the experience? |
| Grading practice: | 6. What do you emphasize when determining a grade in oral English? |
| | 7. What do you emphasize when determining a grade in written English? |
| Teachers' support when grading: | 8. Do you cooperate with colleagues on grading? |
| | 9. What kind of help or support do you get in the process of grading? |
| | 10. Do you think you have enough support in this work? |
| Overall achievement grades: | 11. Could you describe how you work with setting overall achievement grades in oral English at your school? |
| | 12. Could you describe how you work with setting overall achievement grades in written English at your school? |
| | 13. Do you use any aids to support the grading of pupils? (for example, tests, custom designed tests, survey tests, presentation, group assignments, home assignments, UDIR's examination assignments.) |
| | 14. To what extent do you use the national curriculum/local curriculum as a benchmark/ anchor point when setting an overall achievement grade? |
| | 15. Is there something you emphasize more/less, for example, if a pupil is in-between two grades? (competence, attendance, homework, effort, achievement or progression.) |
| | a) Do you have an example? |
| | 16. Do you use the entire grade scale? |
| The relationship between overall achievement grades and exam grades: | 17. How do you think the relationship between overall achievement grades and exam grades should be? |
| | 18. Would you say that exam grades are a quality assurance of how you and your colleagues have set the overall achievement grades? <ul style="list-style-type: none"> a) Has this led to changes in your own grading practice? |
| | 19. Do you think it is possible to get a fair assessment of the pupils? <ul style="list-style-type: none"> a) What is difficult, if any? |
| | 20. Are you familiar with UDIR's examination guide for pupils and teachers? |
| | 21. Do you use this document actively yourself in teaching planning or as information to the pupils? (bring and show) |

Figure 5: Categories and questions presented from the interviews.

4.1.1 Teacher experience

All participants were asked on how many occasions they had set overall achievement grades in English, a number that was highly variable among the participants. Participant 1 had not

given any overall achievement grades due to the fact that she completed her teacher education last year. Participant 2 had set overall achievement grades 3 times and Participant 3 could not recall how many times she had done it. Both Participants 4 and 6 had set overall achievement grades on 10 occasions. Lastly, Participant 5 who had been working as a teacher for 3 years had set overall achievement grades in 2 of those years.

None of the participants felt qualified to grade their pupils after completing their teacher education. Both Participants 3 and 5 mentioned that from their teacher education, they gained a lot of knowledge of the English language, but they never had any teaching in assessment and grading. On the other hand, Participant 1 who was newly educated explained herself as being lucky with her teaching practices. She had always been offered to look at papers to grade. When studying English at the university they were also given example texts written by pupils to grade. The common denominator was that all participants learned more from experience. Nevertheless, Participant 3 explained that the more she knew, the more she wanted to know and questioned everything.

When talking about whether the participants had participated in courses/training in assessment, the two eldest, Participants 4 and 6, were the only ones who said yes. The courses were, however, not recent. Participant 4 believed the course she attended was held in connection with a change in the curriculum and perhaps 10 years ago. Common to both courses Participants 4 and 6 attended was grading of papers written by pupils, which they discussed to see whether or not teachers think alike and share similar mindsets.

All participants except Participants 1 and 6 had experience with being an external examiner at oral English exams. Those who had participated as an external examiner in oral English had not participated in examiner training. Participant 2 was the only one who had been an external examiner in written English. She believed that she had participated as an external examiner in written English four times. The examiner training involved reading and grading example papers. After doing so, teachers would get together to discuss their grading and why they had chosen the different grades for the exam papers. Afterwards, they would gather to see examples that a selected test group has graded. Given the above, Participant 2 thought that these examiner trainings had been very useful.

4.1.2 Grading practice

All participants were asked what they emphasised when determining a grade in both oral and written English. Both Participants 4 and 6 emphasized communication as an important factor in oral English. Their pupils should be able to communicate and have a dialogue in English. Participant 6 also said that a pupil must have good content, reflection and a good language in order to achieve a high grade. However, there was certainly room for incorrect grammar, a few stops and some occasions of stuttering, according to Participant 6. Intonation was mentioned by Participant 1 as being most emphasized in oral English. It seemed to be most important for her, especially in the 8th grade because the pupils had a fact-based approach. By this she meant that the pupils showed a general lack of reflection and personal anecdotes.

On the other hand, Participant 3 was concerned with not being fooled by a pupil who has very good flow and intonation in the language. She learned that there were pupils who have a really good intonation but could not use the language in a proper way to communicate. Participant 3 emphasized the pupils' language comprehension in oral communication, meaning that the pupils managed to use the language, build sentences, that the sentences were good and that they had a good vocabulary. Participant 5 had similar thoughts that intonation was the least important factor emphasized in an oral grade. She seemed to focus on the pupils' vocabulary more than the other participants. Participant 4 was the only teacher who mentioned that she included the pupils' engagement in class when determining a grade in oral English.

Participant 2 felt that she did not emphasize anything more or less in oral English. She thought that a grade should reflect a combination of several factors. There were often professional aims where the pupils should show knowledge of certain countries. The pupils were for instance required to explain features of history and geography in Great Britain and the USA, which meant that the English language could not be looked at in isolation either. She stated that she had often tilted a grade up where a pupil has shown a lot of extra-linguistic knowledge. However, she also had to understand what the pupil was trying to convey. Given the above, the two factors language and content would be looked at in combination, according to Participant 2.

Likewise, Participant 2 also shared these thoughts when determining a grade in written English. Participant 1 looked at how her pupils wrote in terms of orthography and sentence structure, but she would also emphasize the importance of the overall impression of a text. This included how the text was structured and the formal requirements in terms of punctuation, font size and headlines. Communication came up as an important factor for Participant 4 in written English as well as in oral English. Her second factor was the pupils' language. She explained that there was a difference between making one error and having the same error throughout a text. This would imply that if a pupil managed to write a whole text with only writing small I's, then they according to Participant 4, were most certainly not able to achieve the highest grade 6, and barely a 5.

Participant 5 looked at the written assignment as a whole first. She looked at the structure, whether their texts had a title and clear paragraphs. After doing so she started by skimming the text to see whether they have answered the text correctly. She explained that it was very difficult because she has experienced having pupils who write good texts but the texts they have written did not correspond with the task. Participant 5 thought she was different from others, because she did not punish the pupils by giving them a 3 instead of a 4 just for not understanding the task properly.

When asked what Participant 6 emphasized when determining a grade in written English, she explained that she first tried to read the content and whether or not they have answered to what the task asked of them. Participant 6 usually set a grade for herself in terms of the content, without looking directly at the pupil's language and grammar. She said that she sometimes set a separate grade for the content if there were huge deviations between the language and the content. After doing so she looked into their language, and for the weaker pupils she also had to look at punctuation. What Participant 6 emphasized when determining a grade was therefore dependent on what level the pupils were at. For low-achieving pupils, she aimed at them trying to convey meaning and being able to communicate. However, for high-achieving pupils she liked to use a division of content and language.

4.1.3 Teachers' support when grading

When talking to the participants of the support they received when grading, there was one thing that was common for the majority. All participants expect Participant 5 had good support in talking and discussing with colleagues when they were in doubt. Participant 5 was

the only teacher who expressed that she felt alone as an English teacher in the classes she had worked with. She had not worked with a team, even though there were several English teachers working at School C. She claimed that, as a result, there were perhaps some pupils who had received a different grade than they should have. All other participants were working in teams with several teachers. Participant 6 also used a friend who was a former colleague for assistance when she needed to and expressed that it was a huge advantage to be working in teams when grading.

As a newly graduated teacher, Participant 1 would prefer more guidance in grading. She explained that she was provided with a supervisor, but he was not an English teacher and their sessions were only once a month. Participants 2, 3 and 4 expressed that there should be a better support system with grading pupils. UDIR should issue better guidelines according to Participants 2 and 3. Participant 2 felt that the characteristics of goal achievement from UDIR was not very clear on the differences between the different grades. Conversely, Participant 4 thought that there should be some sort of course regarding assessment once a year similar to the course she participated in several years ago. This should be provided especially for teachers working with pupils in the 10th grade, and the course should take place in March before an exam. Given the above, the participants used colleagues and UDIR for support, but no one knew of any other support apart from this and they were all open for more training in assessment.

4.1.4 Overall achievement grades

The participants were all asked how they worked with setting overall achievement grades in both oral and written English. Participant 1 wanted the oral grade to be based on the assessment situations that the pupils had had during all of lower secondary school, not only from their work in the 10th grade. With this in mind, Participant 1 also thought it was important to be verbally active in class. She emphasized that being verbally active in class did not necessarily mean that the pupils were actively raising their hands to answer questions during class. However, the pupils should show participation in group work and be able to discuss with classmates in groups. After all, there were several ways to show that they were engaged in the classroom according to Participant 1.

Participant 2 answered by saying she kept track of what the pupils were doing all the time. However, the pupils had to know what they were being assessed on and when. She explained

that they sometimes had mini talks in class with peers and small assignments. All in all, Participant 2 said that the assessment situations form the basis for the overall achievement grade. English was a progression subject according to Participant 2 and this applied both in oral and written English. What the pupils did in the last few assessments showed more of where the pupils were at in terms of achievement.

Participant 3 explained that the pupils only got one grade per semester at School B. They tried to avoid grading the pupils too much, due to the stress this puts on pupils. However, Participant 3 had her own system and documented the pupils' process along the way, where only by the end of the 10th grade she would be able to evaluate the pupils in terms of all competence aims and what they have achieved. Furthermore, she claimed that this was simple in some areas and difficult in others. Not all of the competence aims were as easy to assess, according to Participant 3. She explained that some of the competence aims were very big and diffuse. Even though she found it difficult, Participant 3 tried to have an overall assessment of what the pupils have achieved.

When being asked how she worked with setting overall achievement grades in oral English, Participant 4 explained that she first observed the pupils in class. She looked at who were verbally active and encouraged her pupils to be as verbally active as possible. Those who were verbally active towards her, were easy to place she said. Participant 4 said that the verbally active pupils would have an advantage. For instance, a pupil who was verbally active and tried to communicate in English, and who, based solely on language proficiency, should get the grade 2, would perhaps get a 3 overall because of him/her being verbally active.

Participant 5 answered by saying that the basis for her pupils' oral overall achievement grade lied in two assessments and was also dependent on the pupils being verbally active in class. Participant 5 normally had two oral assessments every year. It was usually a presentation and a subject conversation with the teacher in groups. The subject conversation was emphasized the most in the pupils' overall achievement grades.

Furthermore, Participant 6 tried to get everyone to talk English to bring out the pupils' efforts in class. She observed their activeness, language and the content the pupils managed to produce in class. Other than that, she supported her grades with presentations which had clear

competence aims regarding language and reflection. These presentations were set with clear guidelines and criteria to see if they were at a high or low level.

When moving on to overall achievement grades in written English, Participant 1 explained that the main focus here would also be the assessment situations. Secondly, she would also look at the pupils' engagement in class. It was easier for her to look at progression seeing as all texts were saved digitally and the teacher had access to see the pupils' progression.

Participant 2 said that she based her overall achievement grades on some small and some big assessments. Usually, she tended to use old exam assignments at the end of the schoolyear. This was done so the pupils could become acquainted with this type of assignment, because that was what met them eventually.

Participant 3 explained that she tried to give her pupils as many opportunities as possible to show their competence. She explained having a big focus on process-oriented work to understand how they worked, how they used their sources and if they were critical to the sources being used and how they referred to these sources. Furthermore, Participant 3 tried to document as much as possible of the pupils learning and thinking process, in order to give them a proper grade. Finally, she emphasized that the last thing the pupils did should matter the most. However, if a pupil had shown very little competence at the end and she knew that the pupil could do better than this, then she would try to give that pupil as many opportunities as possible to show his/her competence.

When being asked how Participant 4 worked with setting overall achievement grades in written English, she claimed using all sorts of methods. Everything from questions to glossaries. Participant 4 also had some papers as assessments, which she disliked correcting. She gave her pupils writing tasks of a chosen theme sometimes to ease her own workload. The pupils either wrote small texts or wrote texts together with peers. Another thing that she thought worked well were translations, which she also commented could be seen as old fashioned. Participant 4 saw the pupils language understanding right away when using translation tasks. They could write very bad sentences in English but if they were good at translations, then they were proficient in English according to Participant 4. Translations were also fairly easy to give feedback on as well. Participant 4 immediately saw whether her pupils followed the Norwegian structure or whether they were familiar with the language and wrote proper English. The pupils did not go in with high shoulders on such type of assignments as

opposed to when they were writing papers. She often translated a piece of writing from their homework into Norwegian and gave her pupils the task of translating it back to English. After doing so they could compare their translation against the homework text.

Participant 4 had a strong belief in glossary tests. She emphasized the importance of practicing new words in order to achieve a better vocabulary. Given the above, in written English she tried to document many assessments on every pupil to use as a basis for her overall achievement grades. If there was any doubt, she let the pupils have a new assessment in some sort of way. Participant 4 thought of herself as a kind teacher when grading. She did not know for certain whether she was too kind, but she sometimes felt like she was.

There were a lot of assessments in written English which could be seen as the basis for the pupils' overall achievement grade. When setting an overall achievement grade, Participant 5 emphasized the last assessment they had had. However, she also said that the overall achievement grade should benefit the pupil. An example was given where a pupil had received the grades 3+ and 4+. In this scenario she would give the pupil a 4 for an overall achievement. All in all, the grades should be as fair as possible.

When working on setting overall achievement grades in written English, Participant 6 tried to constantly have some small assessments. These small assessments were divided into tasks. An example given by Participant 6 was when working with earlier exams they were sometimes given the task of either writing a short answer or a long answer. She also explained to her pupils that homework was an important aspect of their grade. In their homework they got a chance to show that they had understood a text and that they managed to articulate themselves by answering questions regarding a text they had read. Accordingly, when Participant 6 was in doubt whether a pupil should get a 3 or a 4 for their overall achievement grade, she used homework as an extra bonus of documentation to either tilt their grade up or down. The last big test at the end of the semester counted the most of their grade according to Participant 6. However, Participant 6 would gladly look over the pupils' previous achievement in general if they were extremely unfortunate on this particular test. She explained that it was important to look closely at the pupils' last test when grading because pupils changed a lot through that year. The boys especially became more mature and reflected over the last year of lower secondary school according to Participant 6.

All participants felt they used the national curriculum as a benchmark when setting overall achievement grades. They all mentioned using the competence aims. Participants 3 and 5 were those who also mentioned using the characteristics of goal achievement template from UDIR when setting overall achievement grades. Participant 4 talked about the national curriculum as being open for interpretation and not firm and specific as to what the pupils should achieve and know. As a result, she often made her own arrangements when teaching subjects she was interested in. This meaning, she did not think she went beyond the national curriculum, but she did not follow all competence aims either. Participant 3 also shared a concern with the national curriculum. She felt there were too many competence aims. Some of the competence aims were also formulated in a way that was difficult to assess according to Participant 3.

The entire grade scale was used by all participants. When the participants were asked if there was something they emphasized more or less if a pupil is in-between two grades there were many similarities. The interviewer mentioned competence, attendance, homework, effort, achievement or progression as factors to emphasize. Competence was seen as a given for many participants, and everyone emphasized it. Apart from that, they all mentioned that they would emphasize progression. Participant 4 was the only one who would emphasize effort. Participant 2 explained that a lack of effort in class would not downgrade a pupil. On the other hand, if a pupil was in-between the grades of 4 and 5 and was active in class, he/she would benefit from that and perhaps receive the grade 5. Attendance seemed irrelevant for the participants regarding an overall achievement grade. Participants 1-4 would not emphasize homework in such a situation. However, Participants 5 and 6 mentioned homework as something they would emphasize if a pupil was in-between two grades. Participant 6 answered by saying that the quality of a pupil's homework would be emphasized when in doubt. She reasoned it by saying: if a pupil showed good competence at reading his/her homework, it was very different from just showing up in class and achieving a better grade.

4.1.5 The relationship between overall achievement- and exam grades

When being asked how the relationship between overall achievement grades and exam grades should be, Participant 1 believed that overall achievement grades should always give pupils the benefit of the doubt. She was asked whether it was okay if a pupil got a 3 for the overall achievement and a 5 for an exam. She believed that to be fair, if that was what the pupil

managed to accomplish. Although it would be more difficult if it was the other way around according to Participant 1.

Participant 2 said that the exam grade should hopefully be the same as the overall achievement grade. However, sometimes it depended on the task itself, how it was made. The theme might be very good, leading to the pupils having a lot to write about. If this was an area of interest for the pupil, then it was much more likely that the pupil could achieve a better grade. On the other hand, if the theme revolved around something a pupil had little knowledge of, then it would be more difficult to answer the exam. Participant 2 thought that this could be difficult to avoid. Nonetheless, she believed that the pupils should be at the same level regarding their structure and language. Her school has received feedback that their overall achievement- and exam grades have been very similar.

Participant 3 explained that those grades represented two completely different things. For the overall achievement, one actually assessed a pupil's 10 years' worth of competence. This was not possible to assess at an exam according to Participant 3. There were many factors that came into play during an exam. Factors such as how the tasks were designed, how the pupils' day was that day and how the pupils' nerves came into play during an exam. However, in an ideal world, there should be no deviations between the two grades according to Participant 3. They should especially match if the task for the exam was good and understandable. Participant 3 also explained that there was no strict syllabus to follow in the English subject. The theme given for an exam should therefore be constructed well enough so that every pupil managed to give a proper answer to the tasks. Accordingly, she thought this applied to genres as well. Participant 3 thought that it was strange that the written exam got to have an own grade on the pupils' diplomas after completing lower secondary school, because it was only one day's worth of work which was put into that grade. There were also many competence aims that the pupils did not get the chance to show competence in for an exam. Participant 3 said that looking at exam grades had not led to any changes in her own grading practice. If there were constant deviations, then Participant 3 would have to change her way of assessing. Some deviations were normal. However, she explained that it was normal to achieve a higher grade at an oral exam than it was in a written exam.

Participant 4 personally found it soothing if her overall achievement grades corresponded with the exam grades given. However, some deviations might occur. Participant 4 explained

the reason why being that the task was not suited or interesting for the pupil. Another reason might be that the pupils were nervous when having an exam. She claimed this to be very difficult and would happily get rid of the whole exam scheme. Participant 4 talked of the exams as being nonsense. Participant 4 felt that 90% of her pupils should get the same grade for the exam and the overall achievement, statistically speaking. Nonetheless, she had experienced that there were small deviations between the grades she set and the exam grades. She had had some pupils who have gone up a grade or two and some that had achieved a lower grade in an exam. Personally speaking, she did not like to have the exam as a measuring instrument of how she had set her grades. Participant 4 felt nervous on both her pupils' behalf and her own before an exam.

Participant 5 explained that her pupils got a higher grade at their exam the two years she set overall achievement grades. She thought that there might be another reason than her grading being wrong or too strict. The reason being that she used a lot of time with the pupils who were going to have an exam in English. They all got individual counselling and they worked very hard with grammar the two days prior to the exam. Participant 5 believed that this made a difference. She looked over their exams and saw that her pupils were more conscious in their writing. Furthermore, Participant 5 explained that this happened with the pupils who especially were in-between the grades of 3 and 4. The pupils might also put in more effort in their exam answer just because it was an exam and the nerves that follows.

Participant 5 did not think that there should be considerable deviations. However, it was natural if there were some deviations. First of all because it was a very stressful situation. Secondly, because the exam tasks vary greatly from time to time, she emphasized an exam from 3 years ago which was really 'out there'. The preparation material was difficult, Participant 5 did not understand everything herself. However, the exam the following year was very good. Participant 5 said that the help she provided her pupils with from earlier exams would not be the practice in the future. She explained that she did more than what was expected of a teacher to prepare for her pupils before an exam. It would therefore be interesting for her to see in the future whether the work she did those two years had a considerable effect. Participant 5 believed that her pupils would have gotten lower grades if she had not helped them as much as she did when preparing for the exam. Participant 5 liked the oral exam in English because it brought out the best in the pupils. She did however not think it was strange if there were some deviations between exam- and overall achievement

grades in written English. If there were deviations, Participant 5 did not think that it was the teachers' fault. She said that those who made the exams did not always have a clear picture of the reality.

Participant 6 thought that the average should be higher for overall achievement- than exam grades and that this situation would be perfectly normal. An exam is a very special situation, and as a pupil, they could have a good or a bad day. As a teacher there was more time to figure out the level of his/her pupils if they have followed them for 3 years she said. If the pupils had one bad assessment at school, they got a new opportunity to show what they were capable of, because teachers were after all looking to see what the pupils knew. By doing so, it was natural that the overall achievement- was higher than an exam grade according to Participant 6.

When being asked whether they perceived an exam to be a quality assurance, Participant 1 said that the examiners could either have a good day or a bad day seeing as they were also human. The examiners could be strict, or they could be kind when grading. Participant 1 stated that grading was difficult in general due to these factors. Participant 2 explained that the exam was a quality assurance in a way because there were teachers with no relationship at all towards the pupils who were grading the exams. She said that she was very happy when the exam results were in and they matched the overall achievement grades which she had set. Due to this, she realized that she did agree that the exam grades were a quality assurance of how she had set the overall achievement grades. On the contrary, seeing the exam grades given to her pupils had not led to any changes in her grading practice. However, being an external examiner herself, has made her more aware. She thought that the competence a teacher learned from such an experience was helpful when grading her pupils' overall achievement.

Both the oral and written English exam were not necessarily a quality assurance in Participant 3's eyes. She was of course thrilled if the grades harmonized. However, she changed her mind a little and said that it was in a way a quality assurance. What was devastating to Participant 3, was that there was such a big focus on the exams in the pupils' final year of lower secondary school. That one day of exam affected the teaching throughout the whole year. The pupils worked so hard to get to that one day of exams. Participant 3 did not see the point in itself of these exams. It was a stressful situation for the

pupils, and it was stressful to hear about the exam for a whole year, according to Participant 3. However, she did care for these exam grades saying that she was also stressed if there were huge deviations. All in all, the exam grades were a measurement tool for teachers to get some clues of their own grading practice according to Participant 3.

Participant 4 gave an example of a written exam two years ago which was very good and appealing for the pupils. The year before that was a difficult exam with a political theme. Participant 4 stated that the exam changed every year which made it difficult to use as a quality assurance or a measuring instrument of how she had set overall achievement grades. She did agree that an exam could be looked upon as a quality assurance but questioned why there was a need for it. The overall achievement grades had already been set when the pupils were having their exams. Consequently, she stated that an exam had never led to any changes in her own grading practice because she did not look into the pupils' exams, only the grades given. Participant 4 explained that this was not something she got paid to do and was thus not bothered by doing so either. It would not mind Participant 4 if an exam replaced the overall achievement grade. Then the pupils would have been more focused in all subjects in case they had an exam in the subject.

Before Participant 5 started to assess pupils and give them grades, she went through UDIR's assessment examples. She remembered sitting with a colleague looking at how these examples were graded. She then found out that there were greater expectations of the pupils than she would have thought. Participant 5 explained that a pupil had to have a very good vocabulary and write very precise to achieve a 4. As a result, she explained that the exams had led to a different way of thinking but it had not led to any changes in her grading practice. She felt that an exam was a quality assurance since she looked at those grades to see whether she was on the right track or not. This was something she automatically did, and something she believed all teachers automatically did.

Participant 6 answered yes straight away with no hesitation as to whether exam grades were a quality assurance of how she and her colleagues had set the overall achievement grades. She thought that it was good to see where these exam papers ended up on the grade scale. There were several teachers who had set the grade together and not only one, which was reassuring according to Participant 6.

At last, the participants were asked if they believed it was possible to get a fair assessment of the pupils. Participants 2, 3 and 5 definitely meant that it was possible, whereas Participants 1, 4 and 6 answered by saying it was difficult. Participant 1 explained that the hardest aspect of the job was to not let the personal relation between teacher and pupil interfere when grading. She explained it was especially difficult for her since her pupils at that moment were her first pupils and she cared for them very deeply. However, she liked to think that this benefited her pupils. Participant 4 found it difficult with the pupils who were not verbally active. It was difficult because they probably achieved a lower grade than they should because they were not able to show their competence as much in front of the teacher. Likewise, she also found it difficult to grade her pupils seeing as she often started working with them their last year of lower secondary school. Participant 6 felt that the grade pupils achieved was often not fair in terms of how they would work with the English language outside of school after having completed school. Even though Participant 3 felt it was possible to get a fair assessment of her pupils, she explained that there were many factors which should be considered when assessing. In English there was not a right or wrong answer such as in math. When sitting with colleagues and discussing papers, they often disagreed. As a teacher, she connected a bond with her pupils and cared for them. That may sometimes cause teachers to look for positive things, because as teachers; they cheered them on. Participant 3 thought that she was more kind than strict, even though the pupils sometimes thought otherwise she said.

4.2 Exam assessment

Participants 1-6 graded the same exam paper and gave a comment as to what was emphasized in the grade. The exam paper was written by a pupil in the schoolyear 2017/2018 at School C. This pupil received the grade 4 on this written exam and a 3 for the written overall achievement in 2018. An anonymous version of the exam paper can be found in Appendix 10. Each participant's assessment is presented separately.

4.2.1 Participant 1

Participant 1 gave the paper grade 4. Her assessment of the exam paper was given orally while being audio-recorded. She felt that this exam answer was of average goal achievement, because of the long answer in part 2. According to the task, the pupil was supposed to create a text about his/her own sense of belonging when using English compared to his/her mother tongues. The pupil was supposed to include one or more examples from the preparation

material. Nonetheless, Participant 1 felt that the pupil put the focus on other people. He/she managed to write some oblique points. The pupil used a lot from the examples, but Participant 1 did not feel that the pupil told anything of his/her own experience with the English language. She explained that the pupil did not seem to have understood the task completely. He/she never used a personal way of writing the text to answer the task properly.

The content at first in the long answer felt like repetition from the first short answer for Participant 1. In the short answer, the pupil managed to talk more about his/her own experience with the sense of belonging, as he/she should have done in the long answer. An example was when the pupil wrote “Even though I can speak many languages, but only my mother tongue gives me that strong feeling of belonging” (see Appendix 10). Furthermore, the pupil referred to sources. There were some typical Norwegian ways of writing. Examples given were when the pupil wrote “These things are constant, while can always change your sport over time” and “For example you will not stay in your football team for your whole life and you will change it one day. That way you will lose your sense of belonging to your team one day.” (see Appendix 10).

The pupil explained what a multicultural society was, which was very good according to Participant 1. She always told her pupils that it was wise to give a short explanation of what he/she was writing about. Some typos were found by Participant 1, but many of them were common mistakes such as have/has, plural s and so on. Unfortunately, the pupil slipped on his/her long answer and did not answer the task properly. All in all, Participant 1 would give the grade 4 on this exam paper. She would give that grade due to the pupil’s language, the sentence structure, and some very common mistakes.

4.2.2 Participant 2

Participant 2 gave the paper grade 3. She gave both an oral assessment while being audio-recorded and her notes. Participant 2 thought that the pupil answered to what the tasks asked of him/her in 1A and 1B, however not in task 2A. The text written to answer 2A was not an answer to the task at all according to Participant 2. This made her wonder whether the pupil actually answered task 2D and wrote the wrong task number in his/her exam answer. The reason she believed so was because the pupil wrote that he/she was going to talk about two

persons or two texts which was more in correlation with task 2D where the pupils were supposed to choose two texts.

2A involved talking about one's mother tongue, which the pupil did not mention at all in the text. Anyhow, it was not an answer to either 2A or 2D according to Participant 2, even if the pupil had written the wrong task number. The pupil had a lot of knowledge linguistically speaking. He/she had a language which communicated even though there were many parts which did not make sense. Moreover, the pupil had good language and communicated well in a way that he/she was well understood. He/she managed to introduce the tasks in a good manner and showed knowledge in text writing which outweighed the fact that he/she did not answer 2A correctly. Participant 2 would therefore have given this exam paper the grade 3. She explained that this grade was generous and definitely not a strong 3.

Furthermore, Participant 2 used the characterization of goal achievement from UDIR when grading the exam. She wrote that the short answers 1A and 1B were at the grade level 3-4 whereas the long answer was at the grade level 2. Some notes were also made in the actual exam paper. Participant 2 missed headings in both 1A and 1B. She commented that task 1B explained, but the pupil was not precise enough in his/her answer. The heading in the long answer 2A was also corrected by Participant 2, since the words were missing initial letters. Her other notes consisted of some verb tense errors in 2A and that the text had no conclusion.

4.2.3 Participant 3

Participant 3 gave the paper grade 2. She forgot to grade the exam prior to the interview. She was told that it could be sent by e-mail afterwards because the interview and the exam paper were not linked together. Firstly, Participant 3 sent an e-mail saying that she believed that this exam answer had gotten the grade 3. She had misunderstood the task and the researcher answered by clearly outlining the length and detail of the answer. Participant 3 sent a new mail with a more thorough assessment of the exam with the grade she would have given. Furthermore, she wrote that she realized that she had not read the tasks well enough the first time. When reading the tasks against the answers more thoroughly, she wrote that she would be even stricter. At first, Participant 3 thought that the pupil showed average goal achievement. The researcher interprets this as the pupil lying somewhere between the grades of 3 and 4.

Participant 3 explained that there were some linguistic errors, but the structure was fine. By saying linguistic errors, she did not refer to misspelling of words but sentence errors and grammatical weaknesses. The pupil could have had a better flow and better transitions. The way he/she introduced paragraphs and referred to the preparation material was weak.

Participant 3 explained that there were some problematic formulations such as “he came as gay” (see Appendix 10). However, the text communicated and brought forward some good points. Unfortunately, the pupil did not have an answer to the assignment text. He/she was supposed to discuss affiliation towards the use of English/mother tongue according to the task. Participant 3 did not feel that this was evident in the pupil’s answer in 2A. She would therefore have given this exam paper the grade 2, because in her view, it was not an answer to the given task.

4.2.4 Participant 4

Participant 4 gave the paper grade 4. She wrote an email with own thoughts and gave an oral assessment of the exam which was audio-recorded. In her written e-mail she wrote that she wondered whether the pupil had used time on this exam answer. She felt like this was an answer where the pupil had tried to finish the text as fast as possible in order to get home. The reason for that was because she found a lot of double words, wrong use of words, wrong use of prepositions and some errors on punctuation.

She had her own note about the language and wrote that the pupil was trying. The pupil was varying his/her language, experimenting with words and linking words. Moreover, he/she came with own thoughts, was able to reason and was prepared for this exam according to Participant 4. However, the text fell apart at the end and she believed that the pupil might have lost his/her concentration. She had an interesting thought, that if this was her pupil, he/she would possibly have received the grade 5 for overall achievement and a 4 on this exam. Her reason for saying this was because of the language. The text communicated well, but the language was not at the same level.

4.2.5 Participant 5

Participant 5 gave the paper grade 3. She wrote a very thorough assessment of the written exam commenting on language, structure and content. On the first short answer she wrote that

the pupil answered short and okay. The pupil could have gone more in-depth and explained why someone's mother tongue gives a feeling of belonging. She commented the vocabulary as mediocre and that he/she could have varied the language to a greater extent. Participant 5 thought it was nice that the pupil used terms related to the theme. Other than that, she felt that the pupil had mostly good spelling and grammar. There were some mistakes regarding sentence structure which affected the flow of the text, for example: "Even though I can speak many languages, but only my mother tongue gives me that strong feeling of belonging." (see Appendix 10).

On the second short answer, Participant 5 wrote that the text responded well to what the task asked for. The pupil had some knowledge of the theme and could explain why UK was a multicultural society. He/she also used terms which were related to the theme and had a good vocabulary. Participant 5 expressed that the pupil could continue to work on sentence structure, seeing as some of the sentences were characterized by Norwegian sentence structure.

At last, she commented on the long answer from part 2. Participant 5 commented on content, language, structure and gave a small comment regarding the use of sources. Firstly, commenting on the content, Participant 5 did not feel that the pupil answered to what task 2A asked for; Create a text about your own sense of belonging when using English compared to your mother tongue/tongues. She pointed out that the pupil explained that it was not necessarily the mother tongue or the first language that makes one feel belonging, but this was probably the only time the pupil elucidated the theme. Participant 5 wrote that the he/she had several good thoughts and valid points in his/her text but could have gone more in depth and reflected to a greater extent.

The pupil had good spelling and a fine vocabulary according to Participant 5. She thought it was nice that the pupil used some linking words (However, even though) to create a nice flow. He/she could still vary his/her language to a greater extent. Participant 5 felt that the pupil could work on his/her sentence structure. Some sentences were characterized by a Norwegian sentence structure, and thus the words came in a wrong order. The pupil could also use sentence starters to create a better flow in the text. Nonetheless, He/she had generally good control of grammar. Some errors occurred, especially regarding prepositions which the pupil could work on.

Finally, the structure was commented on. Participant 5 wrote that the pupil had a good introduction where he/she made it clear what the text was about, even though he/she did not answer the assignment. There was a decent structure in the main body; they adhered to a thought/theme in a paragraph. The long answer did however lack a clear conclusion. The pupil also showed good source reference. Based on the fact that the pupil felled short on the assignment in the long answer, as well as not having a clearer conclusion, Participant 5 would give this exam paper the grade 3.

4.2.6 Participant 6

Participant 6 gave the paper grade 3. When she first looked at the exam answer, she thought it had to be a 4 or 5, but after going deeper into it she felt that it was a 3 or 4. The pupil used “a sense of belonging” at least 14 times during the text. There were many minor errors in language and grammar, but in the end, it was the content that Participant 6 deducted the most. In 1A there was much repetition and little content. According to Participant 6, it would seem that he/she did not understand the content of the task clearly in 1B.

In the long answer task, the pupil used two good examples, but had a very unoriginal title. The introduction was poorly, and the text had no ending. Participant 6 would give this assignment 3/4 if this was her pupil in a formative assessment. She was very much in doubt about what she was going to give but ended up awarding the assignment with grade 3. However, she pointed out that this grade was very strong.

4.3 Statistics from *Skoleporten*

Skoleporten (2020), which is an integral part of UDIR, is used as the primary resource of gathering data. *Skoleporten* (2020) claims to “give schools and school owners (the local and county authorities) easy access to relevant and reliable information for use in local quality assessment activities” (Skoleporten³ 2020). Firstly, the researcher looked into the statistics from *Skoleporten* (2020), covering all lower secondary schools of Norway. This was done to get a better overview of the average deviations that existed. It was not possible to look at individual statistics due to privacy. The average grade score of overall achievements and examinations from year 2014/2015-2018/2019 have been looked at both nationally and school wise, which includes Schools A-C. As mentioned earlier, Participants 1 and 2 are from School A, Participant 3 from School B and Participants 4-6 are from School C. This was done to see whether there were any conclusions to be drawn between the interviews and statistics from each school. *Skoleporten* (2020) has been used as the primary source here as well. School B is a smaller school with a smaller pupil composition than Schools A and C, which entails that some numbers are exempted from publicity due to privacy matters.

The statistics are divided into two subsections; oral English and written English, which implies that a presentation of statistics in oral English both nationally and school wise is provided before a presentation of the written statistics. This was done to easier notice and compare similarities from the national average and the averages of Schools A-C in both oral and written English.

4.3.1 Oral English

Figure 6 portrays the national average grade scores in oral English regarding overall achievement- and exam grades. The coloured horizontal lines represent the average grade score of the five schoolyears combined. The numbers represented at the left margin cover the Norwegian grade scale which goes from 1 to 6, where 6 is the highest achievable grade. The graphs start at 2 instead of 1 to give a clearer visualization.

³ <https://skoleporten.udir.no/omskoleporten/english/skoleporten-the-school-portal?kode=55>
Accessed 13 February 2020.

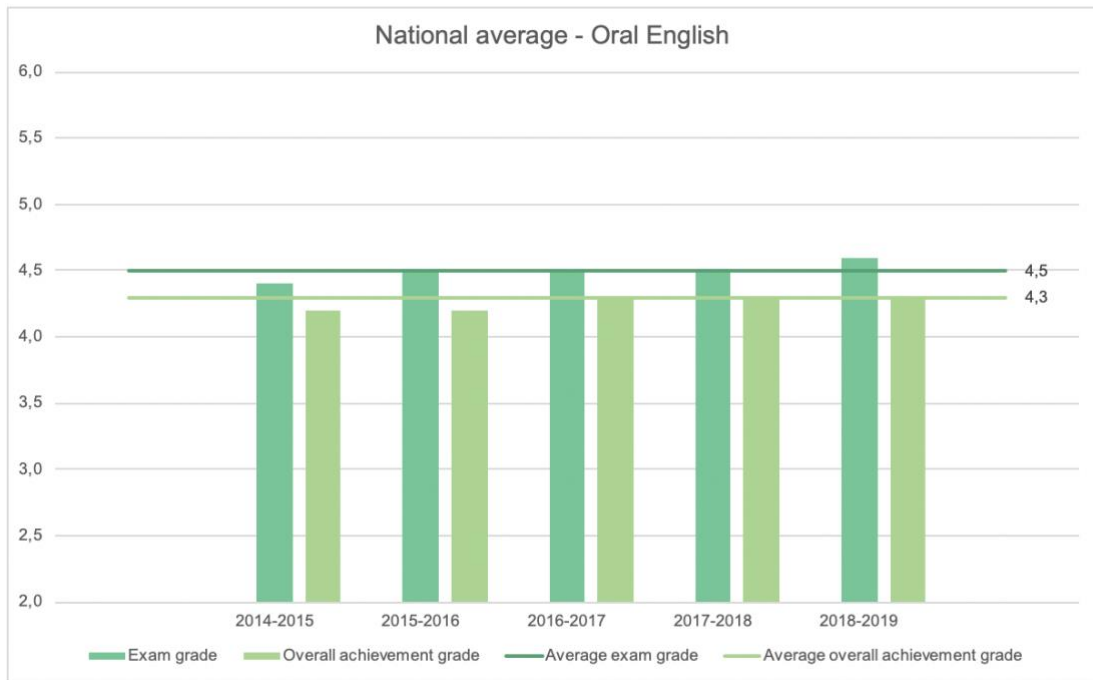


Figure 6: The national average overall achievement- and exam grades in oral English from schoolyears 2014/2015-2018/2019.

The national average grade score has been 0,2 higher at oral exams than for an oral overall achievement when looking at all five years combined. Similarly, every year represents a higher average grade score at oral exams in English. This could imply that pupils in general achieve higher grades for an oral exam in English than for their overall achievement.

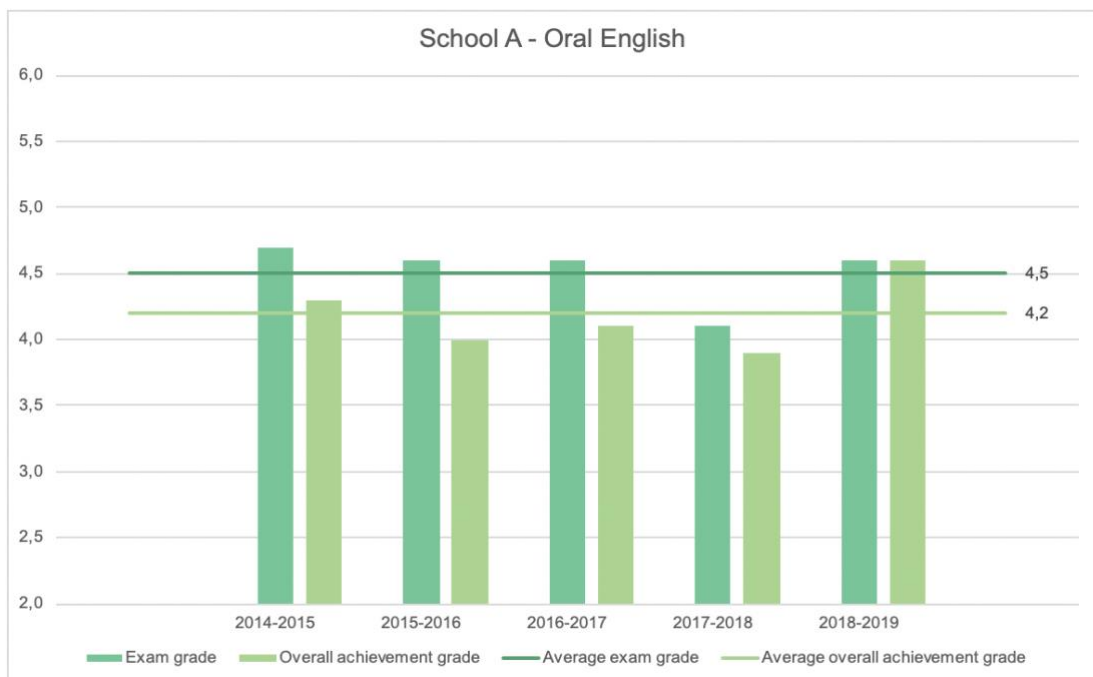


Figure 7: Oral English overall achievement- and exam grades at School A from schoolyears 2014/2015-2018/2019.

Figures 7-9 portrays the deviations between overall achievement- and exam grades in oral English at Schools A-C over the last five schoolyears. As can be seen in Figure 7, the average grade score was higher in oral exams at school A. However, School A had the same average grade score for both overall achievement- and exam grades in the schoolyear 2018/2019. All in all, School A had a deviation of 0,3 with all schoolyears combined, between the two assessments, where the pupils seem to have achieved higher grades in an oral exam.

School B is presented in Figure 8, where three schoolyears of exam data are absent due to privacy. The coloured horizontal lines are therefore not accurate in Figure 8. However, the schoolyears 2015/2016 and 2017/2018 show a higher average grade score at oral exams than an oral overall achievement. Furthermore, when exclusively looking at the average grade scores for an overall achievement at School B, there is a noticeable decrease in schoolyear 2018/2019.

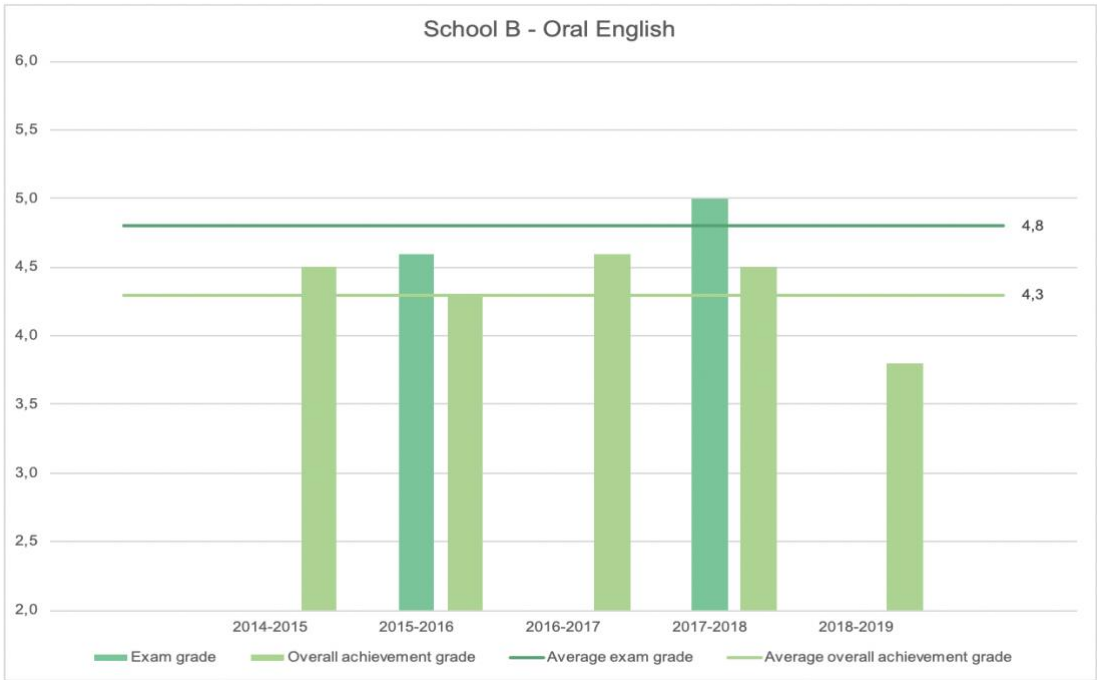


Figure 8: Oral English overall achievement- and exam grades at School B from schoolyears 2014/2015-2018/2019.

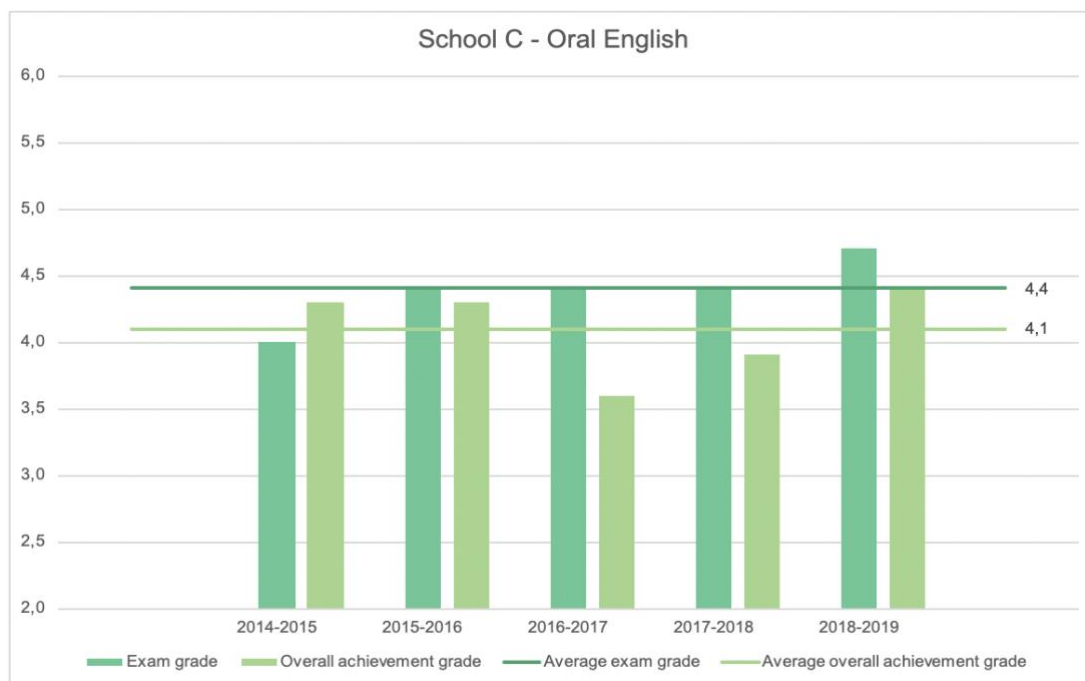


Figure 9: Oral English overall achievement- and exam grades at School C from schoolyears 2014/2015-2018/2019.

The average grade score for oral overall achievement- and exam grades at School C is presented in Figure 9. All numbers are provided and there is a clear average deviation of 0,3 here as well, over the last five years combined. The last four schoolyears show that the average grade score was higher at oral exams than for an overall achievement. However, the opposite was the case for schoolyear 2014/2015, where the average grade score was higher for the overall achievement. Nevertheless, Figures 7-9 portrays that in general, all schools had a higher average grade score at oral exams than for an overall achievement.

4.3.2 Written English

The national average grade scores in written English shows contradictory findings from the national average grade scores in oral English. Figure 10 also shows a deviation of 0,2 between the average overall achievement- and exam grade. However, in written English, pupils seem to achieve higher grades for their overall achievement than at an exam. This finding is consistent in all schoolyears presented in Figure 10.

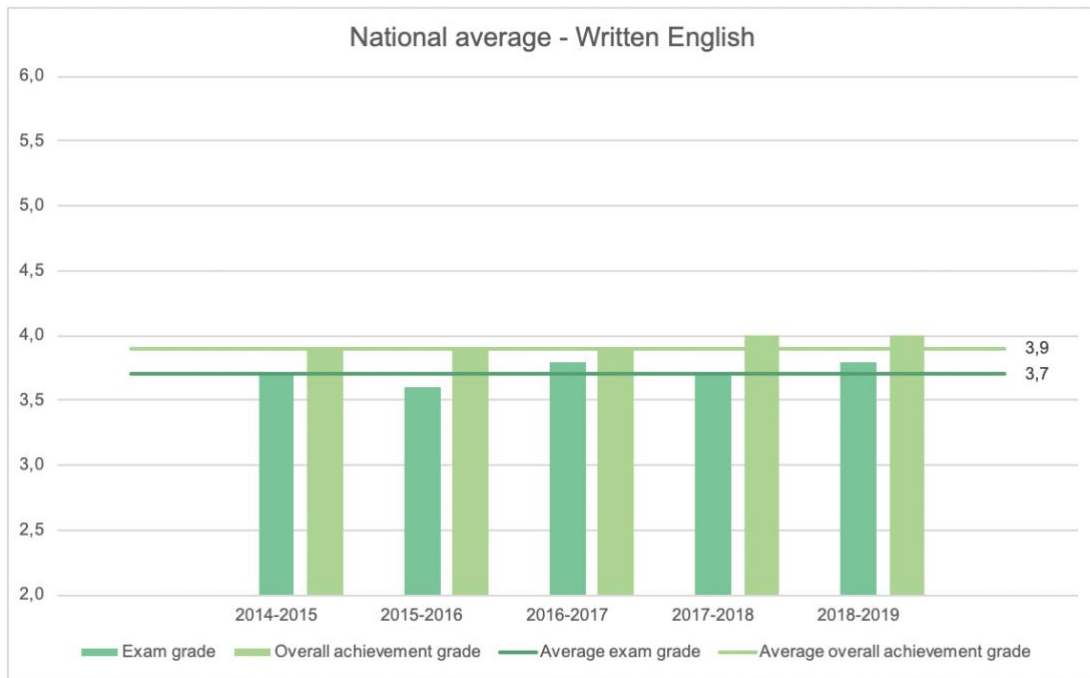


Figure 10: The national average overall achievement- and exam grades in written English from schoolyears 2014/2015-2018/2019.

Figures 11-13 portrays the deviations between overall achievement- and exam grades in written English at School A-C through the last five schoolyears. All three figures show varying results. Figure 11 shows that School A had 3 years where the average grade score was higher for the overall achievement than the exam as opposed to the schoolyears 2014/2015 and 2017/2018. The average grade score for the overall achievement was much higher compared to the exam in 2018/2019 at School A. All in all, over the past five years there has been a deviation of 0,4 where the pupils seem to have achieved higher grades for their overall achievement than their exam.

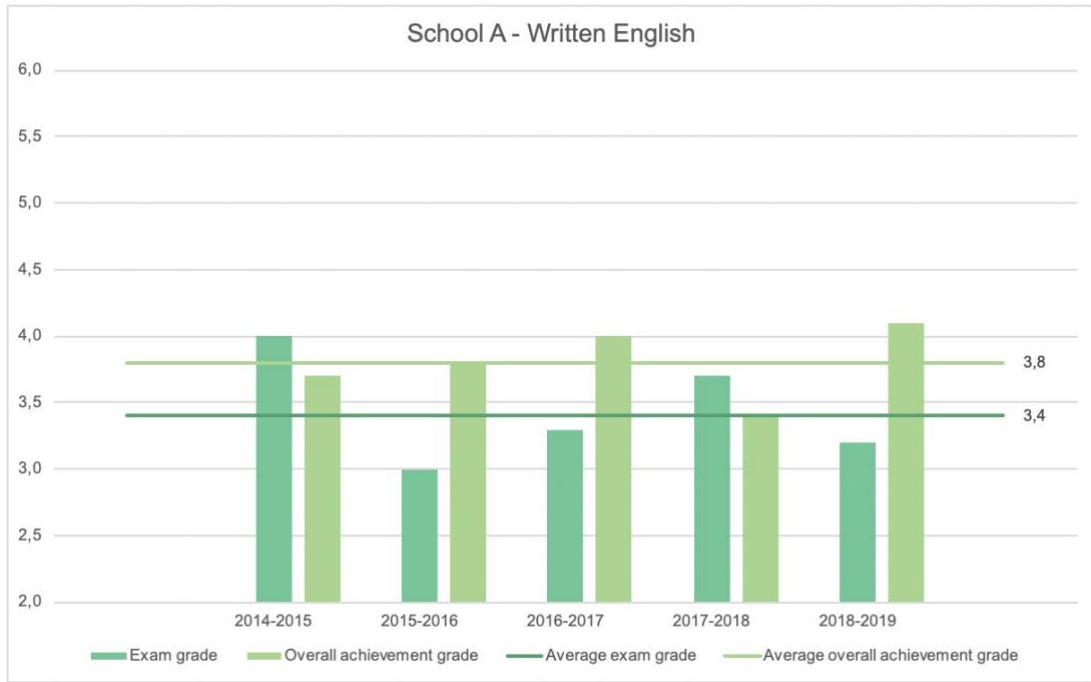


Figure 11: Written English overall achievement- and exam grades at School A from schoolyears 2014/2015-2018/2019.

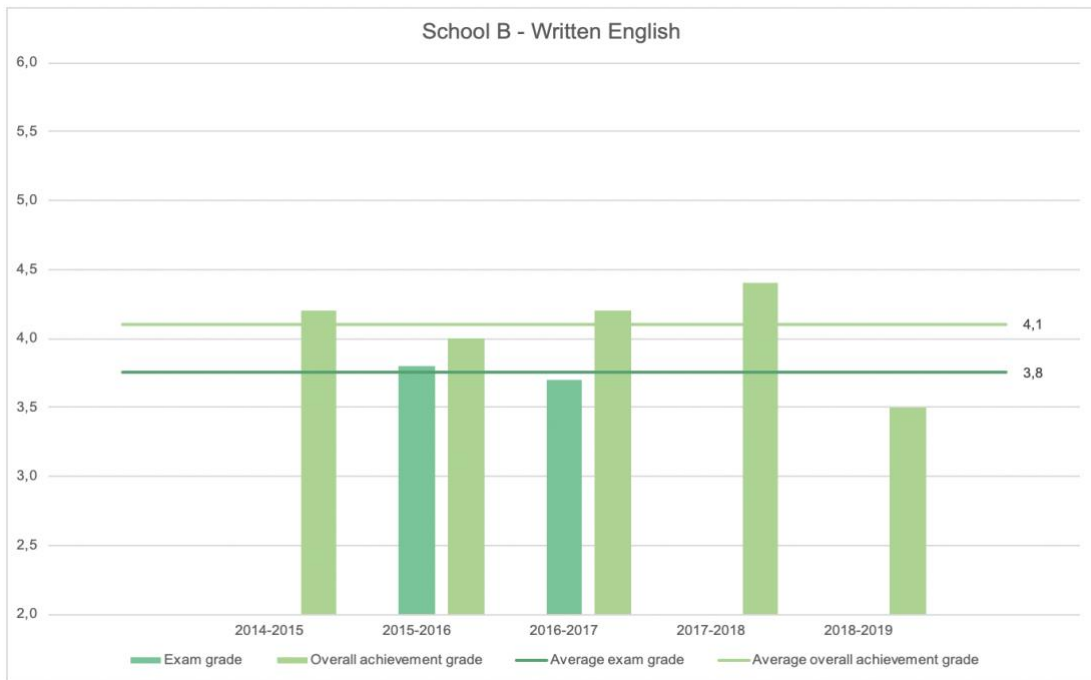


Figure 12: Written English overall achievement- and exam grades at School B from schoolyears 2014/2015-2018/2019.

There were only two schoolyears from school B which were presented with both the average grade score of written exams and overall achievements. The two schoolyears presented in Figure 12 shows that there has been a higher average grade score in overall achievements than in exams. However, the data collected was limited, and thus this cannot be considered as a representative trend for School B. Furthermore, Figure 12 conveys a low average grade score

in overall achievements from the schoolyear 2018/2019. This was also an issue at School B regarding oral English (see Figure 8).

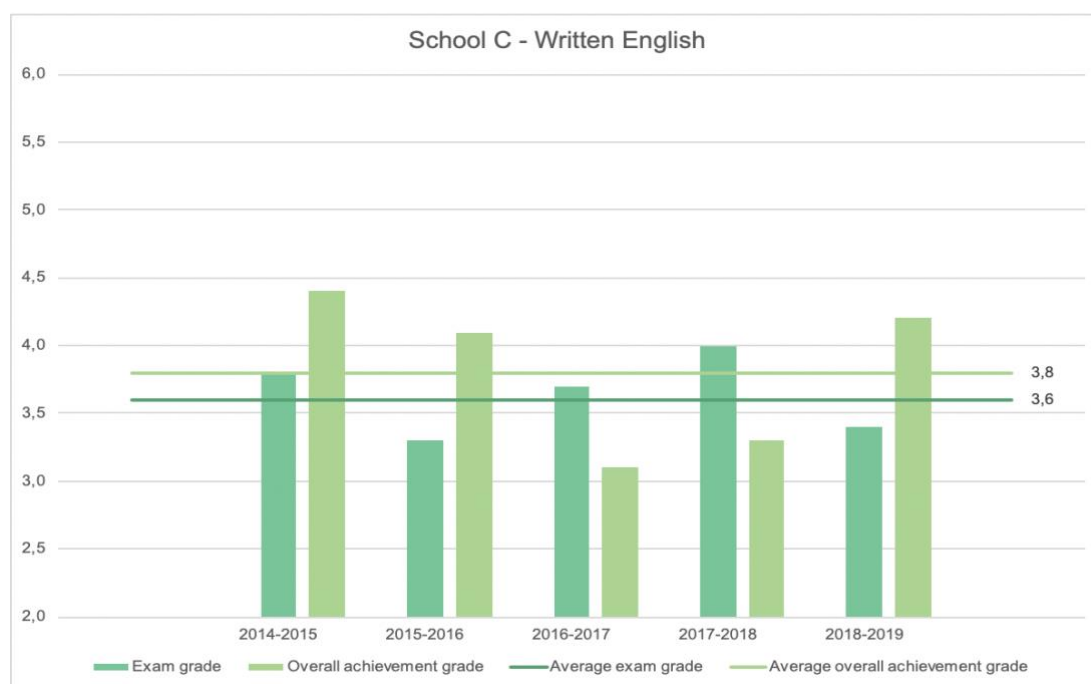


Figure 13: Written English overall achievement- and exam grades at School C from schoolyears 2014/2015-2018/2019.

Lastly, the results from School C can be seen in Figure 13. School C also portrays some varying results. The trend seems to be that pupils achieve higher grades for their overall achievement in written English at School C. However, the two schoolyears 2016/2017 and 2017/2018 portray reversed results, where the average grade score was higher at the written exam than for their overall achievement. Nevertheless, the average grade score was 0,2 higher for the overall achievement compared to the exam when looking at all five years combined.

4.4 Individual statistics

The individual statistics from Schools A and C are presented in Figures 14-17. These figures show the overall achievement- and exam grades from the pupils who either had an oral or written exam in the schoolyear 2018/2019 at Schools A and C. The white rows in the figures express that a pupil has received the same grade for their exam and overall achievement. The light red rows express that a pupil has received a lower grade for their exam than their overall achievement. If a pupil has received two or more grades lower for their exam than their overall achievement, the rows are darker red. The opposition are the green rows, which represent that a pupil has achieved a higher grade for their exam than for their overall achievement. The pupils are placed in order from lowest to the highest overall achievement

grades and the deviations are also descending from no deviations to the highest number of deviations to get a clear picture of how many pupils had deviations and what grades they received for the overall achievement.

4.4.1 Oral English

Oral English

| Pupil | Overall achievement | Exam | Deviation |
|-------|---------------------|------|-----------|
| A 3 | 2 | 2 | 0 |
| A 1 | 4 | 4 | 0 |
| A 5 | 4 | 4 | 0 |
| A 13 | 4 | 4 | 0 |
| A 4 | 5 | 5 | 0 |
| A 8 | 5 | 5 | 0 |
| A 9 | 5 | 5 | 0 |
| A 11 | 5 | 5 | 0 |
| A 12 | 5 | 5 | 0 |
| A 2 | 6 | 6 | 0 |
| A 10 | 6 | 6 | 0 |
| A 14 | 6 | 6 | 0 |
| A 16 | 6 | 6 | 0 |
| A 6 | 4 | 3 | -1 |
| A 7 | 4 | 3 | -1 |
| A 15 | 5 | 4 | -1 |

Figure 14: Overall achievement- and exam grades in oral English 2018/2019 at School A.

What can be seen from Figure 14 is that there were 16 pupils who had an oral examination at School A in schoolyear 2018/2019. 13 of those pupils achieved the same grade for both their oral overall achievement and their oral exam. Only 3 pupils achieved a grade lower for their exam and none of the pupils achieved a better grade on their exam.

Figure 15 shows that there were 10 pupils who had an oral exam in English at School C. 8 of those pupils received the same grade for their exam as their overall achievement. There were, however, 2 pupils who got a lower grade on their exam. The numbers in both Figures 14 and 15 show similar statistic where the majority achieve the same grade in both assessment forms, and a few achieve a lower grade on their exam than for their overall achievement in oral English.

Oral English

| Pupil | Overall achievement | Exam | Difference |
|-------|---------------------|------|------------|
| C 3 | 3 | 3 | 0 |
| C 2 | 4 | 4 | 0 |
| C 5 | 4 | 4 | 0 |
| C 8 | 4 | 4 | 0 |
| C 4 | 5 | 5 | 0 |
| C 1 | 6 | 6 | 0 |
| C 7 | 6 | 6 | 0 |
| C 9 | 6 | 6 | 0 |
| C 10 | 5 | 4 | -1 |
| C 6 | 6 | 5 | -1 |

Figure 15: Overall achievement- and exam grades in oral English 2018/2019 at School C.

4.4.2 Written English

Written English

| Pupil | Overall achievement | Exam | Difference |
|-------|---------------------|------|------------|
| A 2 | 2 | 2 | 0 |
| A 7 | 2 | 2 | 0 |
| A 12 | 3 | 3 | 0 |
| A 13 | 3 | 3 | 0 |
| A 15 | 3 | 3 | 0 |
| A 56 | 3 | 3 | 0 |
| A 43 | 4 | 4 | 0 |
| A 54 | 4 | 4 | 0 |
| A 36 | 5 | 5 | 0 |
| A 34 | 6 | 6 | 0 |
| A 1 | 2 | 1 | -1 |
| A 3 | 3 | 2 | -1 |
| A 11 | 3 | 2 | -1 |
| A 5 | 4 | 3 | -1 |
| A 6 | 4 | 3 | -1 |
| A 9 | 4 | 3 | -1 |
| A 10 | 4 | 3 | -1 |

| | | | |
|------|---|---|----|
| A 16 | 4 | 3 | -1 |
| A 24 | 4 | 3 | -1 |
| A 25 | 4 | 3 | -1 |
| A 32 | 4 | 3 | -1 |
| A 35 | 4 | 3 | -1 |
| A 39 | 4 | 3 | -1 |
| A 44 | 4 | 3 | -1 |
| A 46 | 4 | 3 | -1 |
| A 18 | 5 | 4 | -1 |
| A 22 | 5 | 4 | -1 |
| A 27 | 5 | 4 | -1 |
| A 42 | 5 | 4 | -1 |
| A 48 | 5 | 4 | -1 |
| A 53 | 5 | 4 | -1 |
| A 55 | 5 | 4 | -1 |
| A 26 | 6 | 5 | -1 |
| A 33 | 6 | 5 | -1 |
| A 41 | 6 | 5 | -1 |
| A 47 | 6 | 5 | -1 |
| A 4 | 4 | 2 | -2 |
| A 19 | 4 | 2 | -2 |
| A 51 | 4 | 2 | -2 |
| A 52 | 4 | 2 | -2 |
| A 8 | 5 | 3 | -2 |
| A 20 | 5 | 3 | -2 |
| A 21 | 5 | 3 | -2 |
| A 23 | 5 | 3 | -2 |
| A 29 | 5 | 3 | -2 |
| A 30 | 5 | 3 | -2 |
| A 31 | 5 | 3 | -2 |
| A 37 | 5 | 3 | -2 |
| A 40 | 5 | 3 | -2 |
| A 49 | 5 | 3 | -2 |
| A 50 | 5 | 3 | -2 |
| A 14 | 6 | 4 | -2 |
| A 17 | 6 | 4 | -2 |
| A 28 | 6 | 4 | -2 |
| A 45 | 6 | 4 | -2 |
| A 38 | 6 | 3 | -3 |

Figure 16: Overall achievement- and exam grades in written English 2018/2019 at School A.

What can be seen in Figure 16 is that 56 pupils had a written exam in English at School A in schoolyear 2018/2019. None of the pupils achieved a higher grade on their exam as opposed to their overall achievement grade. 10 pupils achieved the same grade in both assessment forms. That would roughly imply that only one-fifth of the pupils had no deviations. 26 pupils got one grade lower on their exam than their overall achievement, which would constitute almost half of the group. The most fascinating find here is that 20 pupils out of 56 received either 2 or more grades lower on their written exam than their overall achievement. This would constitute one-third of the group who participated in the written English exam at School A. The numbers presented here are also evident in Figure 11, with an average grade score of 4.1 on the pupils' overall achievement, and an average grade score of 3.2 on the pupils written exam at School A in 2018/2019.

Written English

| Pupil | Overall achievement | Exam | Difference |
|-------|---------------------|------|------------|
| C 17 | 3 | 4 | 1 |
| C 1 | 4 | 5 | 1 |
| C 4 | 2 | 2 | 0 |
| C 22 | 2 | 2 | 0 |
| C 5 | 3 | 3 | 0 |
| C 10 | 3 | 3 | 0 |
| C 19 | 3 | 3 | 0 |
| C 28 | 3 | 3 | 0 |
| C 3 | 5 | 5 | 0 |
| C 7 | 5 | 5 | 0 |
| C 8 | 5 | 5 | 0 |
| C 16 | 5 | 5 | 0 |
| C 11 | 2 | 1 | -1 |
| C 18 | 2 | 1 | -1 |
| C 6 | 3 | 2 | -1 |
| C 15 | 4 | 3 | -1 |
| C 23 | 4 | 3 | -1 |
| C 27 | 4 | 3 | -1 |
| C 9 | 5 | 4 | -1 |
| C 12 | 5 | 4 | -1 |
| C 13 | 5 | 4 | -1 |
| C 20 | 5 | 4 | -1 |
| C 24 | 5 | 4 | -1 |
| C 25 | 5 | 4 | -1 |
| C 26 | 5 | 4 | -1 |

| | | | |
|------|---|-----------------------|----|
| C 30 | 5 | 4 | -1 |
| C 2 | 4 | 2 | -2 |
| C 14 | 5 | 3 | -2 |
| C 29 | 5 | 3 | -2 |
| C 21 | 6 | 4 | -2 |
| C 31 | 2 | FA (failed to appear) | |

Figure 17: Overall achievement- and exam grades in written English 2018/2019 at School C.

31 pupils at school C had a written exam in English the schoolyear 2018/2019 which can be seen in Figure 17. C31 did not show up for the exam and did therefore not receive a grade. 18 pupils out of 31 received a lower grade at their exam than for the overall achievement. Only 2 pupils managed to get a higher grade for their written exam and 10 pupils received the same grade on both assessments. At School C there were only 4 out of 31 of pupils who received 2 or more grades lower on their exam than on their overall achievement, as opposed to school A where there were 20 out of 56 pupils who experienced the same thing.

5 DISCUSSION

The discussion chapter is divided according to the research questions stated in the introduction, in order to discuss the data that has been presented in relation to the main aims of the thesis. As stated in the introduction, the given research questions were:

- Are there deviations between overall achievement grades and exam grades in the English subject and what possible factors might influence this deviation in grading?
- What importance does teachers attach to the difference between the two forms of assessment?
- How do teachers approach grading in the English subject at Norwegian lower secondary schools?
- How do teachers' grading practices in the English subject correspond to the recommendations and regulations for grading?

The first research question is divided into two separate sections seeing as there are two questions to discuss and answer. The various findings from both the interviews, examination grades and statistics will be implemented in the different research questions in order to discuss the various findings from different perspectives. Furthermore, the findings will be examined and discussed in light of the theoretical background, presented in Chapter 2.

5.1 Deviations between overall achievement- and exam grades in the English subject

The present thesis has shown that there are systematic deviations between average overall achievement- and exam grades both nationally and school wise. There seems to be a pattern where pupils achieve higher exam grades in oral English compared to their overall achievement grades. The national average in oral English shows a consistently higher average grade score at exams than overall achievements over the last five years. This was also shown to be the trend in Schools A-C when looking at the average grade scores (see Figures 7-9). The deviations presented from the past five schoolyears seems to be similar to what Galloway et al. (2011) reported in their study of grading practices in Norwegian primary schools. They found that schools with low average exam grades more often overestimated their pupils in determining their overall achievement grade, while the opposite was true of schools with high average exam grades. When looking at the average grade scores of Schools A-C in oral

English (see Figures 7-9), it is possible to make assumptions that this was the case for these schools as well.

Considering the individual pupils and their grades at Schools A and C, there are findings to support different trends than what was shown in the average grade scores. In Figure 14, which shows individual grades in oral English from School A, there were 3 pupils and at School C (see Figure 15) there were 2 who received a lower grade for the oral examination. None of the pupils achieved a higher grade for the oral exam. These numbers were somewhat surprising seeing as Participant 3 talked of the oral exam as being ‘the pupils’ time to shine’ and that the national average grade scores at examinations are higher than the average grade scores of overall achievements. Participant 5 also explained that the oral examinations brings out the best in pupils. However, these results are based on a small set of data and it is therefore not possible to make any assumptions or conclusions. Nonetheless, some thoughts to explain these numbers could perhaps be that the external examiners were stricter than the teacher who set their overall achievement grade. The external examiners could have different expectations than the teacher who made the exam tasks. Another element which could factor into this deviation, is that the pupils were perhaps nervous during the examination. It must also be taken into consideration that due to a convenience sampling, Schools A and C are from the same area and could have had the same external examiner in oral English.

In written English there seems to be a pattern where pupils achieve lower exam grades as opposed to their overall achievement grades. The national average in written English shows a consistently higher average grade score for overall achievements than exams, over the last five years. However, the average grade scores at Schools A-C show various results. What could be seen in Figures 11 and 13 was that the pupils at Schools A and C had a higher average grade score at a written exam than for the written overall achievement the schoolyear 2017/2018. Both Participants 4 and 5 talked of this exam as being a good exam, whereas the exam from three years ago (2016/2017) was more difficult. The exam itself could be a factor to consider, the tasks change from year to year, insinuating they might be easy to handle for some and difficult for others.

Furthermore, the individual scores in written English (see Figures 16 and 17) are examples of the deviations one might find at any given school. Figure 16 shows a poor average on the pupils’ exam compared with their overall achievement grades at School A. At School A, 20

pupils out of 56 experienced a decrease of 2 grades or more on their exam as opposed to their overall achievement. When dividing the individual differences from Figures 16 and 17 by the number of pupils, the results show that the pupils at School A combined, had a decline of 0,6 points, whereas the pupils from School C had a decline of 1,2. As can be seen from Figure 17, there were 18 out of 31 pupils who lowered the grade score at School C. 4 of those pupils achieved two grades lower in their written exam than for their overall achievement. There does not seem to be a pattern concerning a certain grade seeing as there were deviations from the grades 2-6 from both Schools A and C. However, the pupils who experienced a decrease of 2 grades or more on their exam compared to their overall achievement had a grade between 4 and 6 for their overall achievement. Nonetheless, this could be seen as quite natural since a pupil with a lower overall achievement grade than a 4, would have to get a 1 in order to achieve an exam grade 2 grades lower than their overall achievement grade, which is highly unlikely.

All in all, the national average grade scores would convey that pupils achieve lower grades for a written English exam compared to their overall achievement grades. There was evidence of this from all findings present in this thesis, thus it would seem that pupils achieve lower grades for a written exam compared to their overall achievement grade both nationally and school wise. The researcher was, however, surprised by the degree of deviations shown at Schools A and C. The national average grade scores in oral English show contradictory findings where pupils achieve higher grades for an oral exam as opposed to their overall achievement. This was also proven to be the fact when looking at Schools A-C (see Figures 7-9). However, these numbers do not show the whole picture, as can be seen from Figures 14 and 15, where none of the pupils achieved a higher grade on their oral exam at Schools A and C. The reason why the average grade score is portrayed higher for the oral English exams is because there are less pupils accounted for compared to the average grade scores for overall achievement grades. The findings in the present thesis correspond with what Gravaas et al. (2008) found in their report of grades in lower- and upper secondary schools. They found that the average grades for a written exam are generally somewhat lower than the overall achievement grades in the same subject (Gravaas et al. 2008).

5.2 The possible factors which might influence a deviation in grading

The factors influencing deviations between overall achievement- and exam grades are difficult to establish. Evidence to support the various factors of the particular deviations

presented in this study are not evident for the researcher to discuss but will be looked into in light of theory and the data collected for the present thesis. As explained in Chapter 5.1, the exam itself could be an important factor of deviations. However, there are many other factors that could explain a deviation between overall achievement- and exam grades. One possible factor for the trends in written as opposed to oral exam grades is that it might be easier for teachers to distance themselves from the pupil when assessing a written exam as opposed to an oral exam. Participant 2, who had been an external examiner in written English, could agree that she assessed differently when grading an exam paper, than she did when grading her own pupils. She expressed that as teachers we put on a different role when an exam was involved. This corresponds with the relationship of the dual roles of coach and judge, which is discussed by Bishop (1992) and Wilson (1996) (see Chapter 2, p.8).

School size and pupil composition are factors which could influence a deviation in grading. Schools B and C presented in this thesis may be categorized as small lower secondary schools in a Norwegian context, seeing as they are placed in rural areas and hold fewer than 300 pupils. School A is also placed in a rural area but holds more than 300 pupils and is therefore a bigger school. Galloway et al. (2011) gave a possible explanation that since small schools have fewer exam candidates, a single external examiner could give especially high or low grades which have far greater significance for small schools than large schools (2011: 12). Another important factor to keep in mind when talking of averages, is that an average is very vulnerable when it is based on few exam candidates, the reason being that deviations have a much larger impact on the average when the size of the school is smaller. This would imply that if there are few pupils having an exam, some low grades would influence the average greatly. It was also presented by Galloway et al. (2011) that schools with low average exam grades more often overestimate their pupils in determining their overall achievement grade, while the opposite is true of schools with high average exam grades. This would imply that teachers' grading practices rely on the level of the pupils at the individual school. Furthermore, this depends on the composition of pupils in a given class. It would seem that the teachers at School A overestimated their pupils when setting their overall achievement grades in written English (see Figure 16), since only 10 out of 56 pupils had no deviations.

The present researcher wanted to have a closer look into the distribution of grades given at Schools A and C to see whether there were any patterns to draw conclusions from. The distribution of grades at Schools A and C can be viewed in Figures 18 and 19.

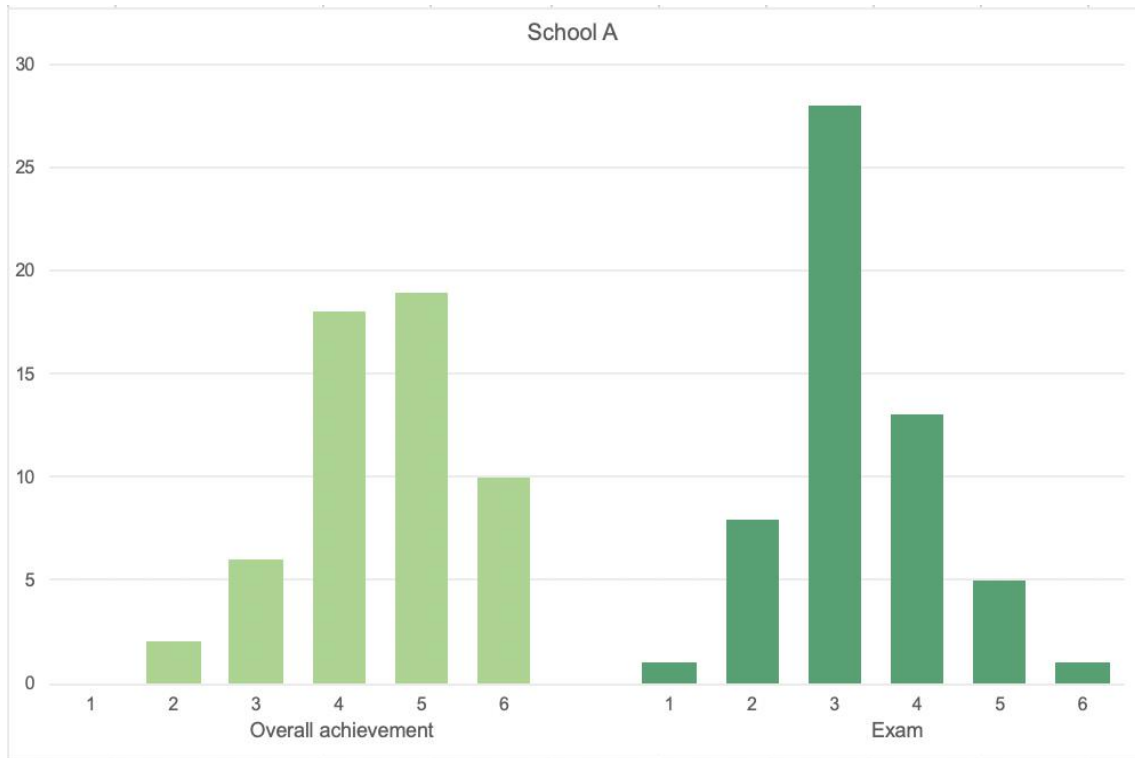


Figure 18: The distribution of grades from Schoolyear 2018/2019 in written English at School A.

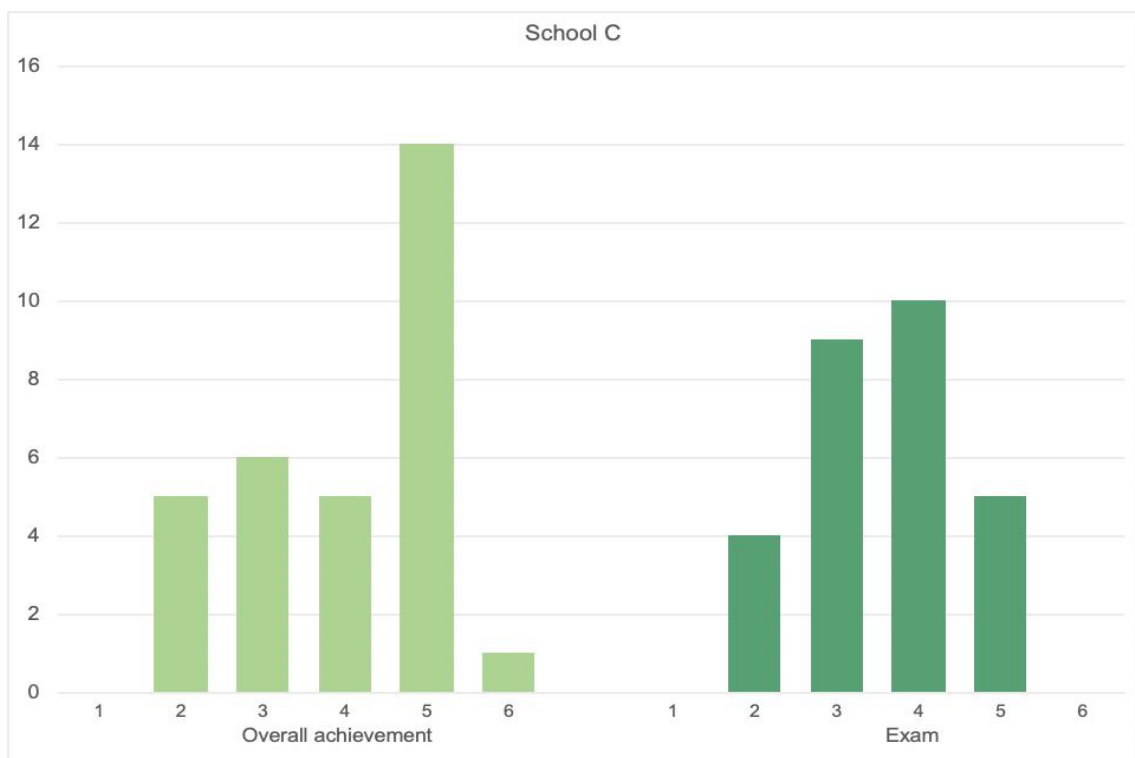


Figure 19: The distribution of grades from schoolyear 2018/2019 in written English at School C.

When looking at the distribution of grades at Schools A and C, there is a more even distribution of exam grades than of overall achievement grades. The distribution of exam grades seems to be very similar to a bell-shaped curve (Bøhn 2018: 233; see also Chapter 2.2, p. 10). A normal distribution was most commonly used in norm-referenced assessment, which used to be a frequently applied practice in Norway, but it was abandoned from the 1970s onwards. Nevertheless, From Figures 18 and 19 it is apparent that the distribution of exam grades resembles a bell-shaped curve. The researcher has no evidence to support whether or not this normal distribution was intentional. However, these figures illustrate teachers' assessment practices as an important factor in order to explain deviations between overall achievement- and exam grades.

The exam paper graded by Participants 1-6 will also be presented as an example to show different factors which might influence a deviation. One interesting factor that made the researcher choose this as an example paper, was that this pupil achieved the grade 3 for the written overall achievement and the grade 4 for the written exam, which is not consistent with the trends presented in the national statistics. However, this pupil had the exam in schoolyear 2017/2018 at School C. From the statistics at School C it is apparent that there was a much higher average grade score of written exam grades than overall achievement grades that year (see Figure 13). Participant 5 was the English teacher at School C who set the overall achievement grades schoolyear 2017/2018. She explained that she had set overall achievement grades her first two years of teaching, which were the schoolyears 2016/2017 and 2017/2018. Those are the only two schoolyears where there was a higher average grade score in exam grades than overall achievement grades at School C from the last five schoolyears. This could imply that Participant 5 might have a different approach when grading than a teacher with more experience. Participant 5 might have underestimated her pupils. Looked at this in context, Participant 5 was the only teacher who expressed feeling alone as an English teacher in the classes she was working with. As a result, there were perhaps some pupils who had received a grade they really should not have, she claimed. This would imply that newly qualified teachers should perhaps receive more guidance and support when assessing. This corresponds with Brookhart (2013) and Wyatt-Smith and Klenowski's (2013) studies where they express that teachers may not always need more competence regarding assessment, but there may be a need for more support and guidance in teachers' work on grading.

Assessment is something many teachers struggle with. All teachers felt a need for more support and guidance of assessment. Whether this support or guidance was in the form of courses or better guidelines when assessing varied. This indicates that the teachers themselves were not sure of what the best solution was. None of the participants felt qualified to grade their pupils after completing their teacher education. This raises a question whether the teacher educations should look at new arrangements for future students in order to prepare them for the tasks they are faced with, in this particular instance grading, when starting to work as a teacher. In light of recent research, there were 112 principals from all over the country who responded to a survey in 2016. This survey intended to reveal what the principals call for in newly qualified teachers (Mellingsæter 2016). The principals graded the newly qualified teachers from 1-6 and found a common ground where the newly qualified teachers were good in subjects but lacking social skills. The newly qualified teachers were also graded in assessment skills, where the average grade from all 112 principals was 3.49 out of 6. This would imply that the overall impression of newly educated teachers' assessment skills could be improved. Common to many principals were the idea of having longer teaching practices in order to prepare the students for the reality of being a teacher (Mellingsæter, 2016). Some principals even suggested having a whole year of teaching practice. However, Dean Knut Patrick Hanevik at the Faculty of Education and International Studies at the University of Oslo and Akershus would rather quality assure and develop the existing 110 days of teaching practice rather than increase the number of days (Mellingsæter, 2016). None of the participants had similar recommendations. Instead, they felt the need for more didactics in courses, seeing as they felt qualified in the English language but not with the tasks teachers are faced with at work.

All participants claimed to have good support in colleagues, except Participant 5. Only Participants 4 and 6 believed to have participated in courses focused on grading. Both of the courses seemed to have taken place a long time ago. This questions whether there is a lack of courses or if there is little engagement within the schools to participate in grading related courses. Nevertheless, Participants 1-4 and 6 emphasised the importance of having colleagues to discuss with. This corresponds with Prøitz and Borgen (2011) and Haugstveit et al. (2006) who state that the opportunity to assess in collaboration with other teachers is important.

There was also a lack of support among supervisors for newly graduated teachers. This became apparent during the interviews where it was explained that Participant 5 received

supervision meetings once a week, while Participant 1's supervision meetings were only scheduled once a month. In addition to the frequency of these meetings, the supervisors may not be educated in the same subjects as the apprentice, such as Participant 1's supervisor who was not an English teacher. This factor made it more difficult for Participant 1 to discuss English papers with her supervisor. Seeing as Participants 1 and 5 had a difference in frequency of supervision meetings, it raises a question as to whether there should be guidelines nationally regarding a support system for newly graduated teachers. In conclusion, a lack of a support system could therefore be seen as a possible factor of deviations between overall achievement- and exam grades, seeing as the likelihood of under- or overestimating pupils' achievement could be higher when working individually. Participant 5's experiences from year 2016/2017 and 2017/2018 could be seen as an example of this.

5.3 The importance teachers attach to differences between overall achievements and exams

The six teachers interviewed were affected by a deviation between the two different assessment forms. They were all nervous about a deviation but had never changed their own grading practice because of it. All participants expressed that the overall achievement- and exam grades should be fairly equal in an ideal world. Participant 2 believed that all pupils should be at the same level regarding their structure and language. By saying that, it seemed that content and knowledge could be the deciding factors of whether a pupil achieved the same grade for an exam or a grade higher or lower. This relies further on the tasks given at an exam.

The overall achievement grade is based on a teacher-pupil relationship characterized by the teacher's in-depth knowledge of the pupil as a person, while grading a written exam is conducted with mutual anonymity. Both overall achievement grades and exam grades can be seen to have strengths and weaknesses, according to Hovdhaugen et al. (2018: 17). This is emphasised by UDIR in Meld.St. nr.28:

It is important to emphasize that exam- and overall achievement grades are two different expressions of competence. An examination grade will be set on a more limited basis of assessment than an overall achievement grade. An exam represents one or very few forms of assessment (written, practical and/or oral) and it is an individual case. Nationally, the average exam grades for centrally granted exams are usually slightly below the average overall

achievement grades. If such a comparison is to be appropriate, it is the Ministry's assessment one must look at if differences over time should systematically deviate from the national, average difference between overall achievement- and examination grades. Permanent deviations should be a warning to the school owner and school principal that it is necessary to change existing assessment practices. However, such analyses are not enough alone, but should be one of several sources of knowledge about own practice for use in a school's development work.

(Meld.St. nr.28 2015-2016: 63, my translation)

Similarly, Participant 3 explained that an exam- and an overall achievement grade represented two completely different things. She claimed that for the overall achievement, a teacher actually assessed a pupil's 10 years' worth of competence. This was not possible to assess at an exam, according to her. She explained that there were many factors that came into play during an exam. Such factors could be how the task was designed, how the pupils' day was that day and how the pupils' nerves came into play during an exam. However, in an ideal world, there should be no deviations between the two grades, especially as long as the task for the exam was good and understandable. Participant 3 also explained that there was no strict syllabus to follow in the English subject. The theme given for an exam should therefore be constructed well enough so that every pupil managed to give a proper answer to the tasks. Accordingly, she thought this applied to genres as well. Participant 3 thought that it was strange that the written exam got to have an own grade on the pupils' diploma after completing lower secondary school, because it was only one day's worth of work which formed the basis for that grade. Participant 1 felt that overall achievement grades should give pupils the benefit of the doubt and that it would be normal if a pupil got a lower grade for an exam. On the other hand, it would be more difficult if a pupil achieved a better grade for an exam, she said.

Participant 4 called exams 'nonsense' and would gladly get rid of the whole exam scheme. She felt that 90% of her pupils should get the same grade for an exam and an overall achievement, statistically speaking. Nonetheless, she had experienced that there were small deviations between the grades she set and the exam grades. Participant 4 thought the exam seemed to change every year which made it difficult to use as a quality assurance or a measuring instrument of how she has set her overall achievement grades. She did agree that an exam could be looked upon as a quality assurance but questioned why there was a need for

it. She made a good point by saying that the overall achievement grades have already been set when the pupils are having their exams. However, it does not seem that she understood that this deviation could be taken under consideration for the years to come.

Participant 5 explained several reasons as to why a deviation between the two assessments could be acceptable. She thought that there might be another reason than her grading being wrong or too strict, the reason being that she spent a lot of time with the pupils who were going to have an exam in English. They all got individual counselling and they worked very hard with grammar two days prior to the exam. Participant 5 believed that this made a difference. She looked over their exams and saw that her pupils were more conscious about their writing. Furthermore, Participant 5 explained that this happened especially with pupils who were in-between the grades of 3 and 4. The pupils might also put in more effort in their exam answer just because it is an exam and the nerves that go along with that situation. An exam undertakes a more formal assessment, whereas it is not based on factors such as the pupil's individual prerequisites, background and relative level (see Chapter 2.2, p.11).

Participants 2, 3, 4 and 5 aimed at the exam itself as being the largest factor in deviations between pupils' overall achievement- and exam grades. It does not seem that these teachers thought of the external examiners or themselves as factors for deviations between overall achievement- and exam grades. This could be due to not wanting to hurt their profession's reputation.

5.4 Teachers' approach with grading in the English subject

When looking at the teachers' grading practices in oral English, there was a difference in what they emphasized. Many mentioned that pupils needed to be able to communicate. However, there was a difference in how Participants 1, 3 and 5 emphasized intonation. One of the competence aims is to "use the central patterns for pronunciation, intonation, word inflection and different types of sentences in communication" (UDIR 2013). On the other hand, there is also a competence aim expecting the pupils to "understand and use a general vocabulary related to different topics" (UDIR 2013). Nonetheless, Participant 5 seemed to emphasize vocabulary more than the other participants in oral communication. When pupils of Participants 1, 3 and 5 would have an oral exam, one could therefore assume that there might rise some deviations from their overall achievement grade due to the difference in emphasis.

Participant 4 was the only participant who emphasized engagement in oral English. It could be fair to assume that engagement concerns the pupils' effort in class.

When looking into what Participants 1-6 emphasized when determining a grade in written English, there were many similarities. Participants 1 and 3 talked of the overall impression as an important factor in written papers. However, the most important factors which would constitute a grade in written English was content and language. Participant 6 was the only participant who explained that she determined a grade depending on what level the pupils were at. For low-achieving pupils, she hoped they were able to convey meaning and communicate and for high-achieving pupils she liked to use a divided approach and look into their language and content.

The differences in grading became more apparent when looking into what the participants emphasized when a pupil was in-between two grades for their overall achievement. Participants 1 and 4 were the only ones who would emphasize effort. Participants 5 and 6 would emphasize homework, while for instance Participant 3 explained that this should only influence their grade in order and conduct and not in the subject. Progression was the one thing all participants mentioned as a factor that would be emphasized, and attendance was not seen as an important factor overall. This is interesting, seeing as "the prerequisites of the individual, absence or conditions related to order and conduct of the pupil, apprentice or the learning candidate should not be drawn into the assessment in subjects" (Regulations to the Education Act, § 3-3). The teachers are thus following the regulations when it comes to attendance. However, homework is not mentioned here, which may imply that this could be drawn into the assessment of pupils. These findings would imply that teachers more often than not set grades according to their own views of what is correct, and not according to what is recommended practice. The interviews thus showed that factors such as effort and progress was considered by some teachers alongside the pupils' achievements when deciding grades. Furthermore, seeing as there are 10 competence aims under both the oral- and written communication categories, it could be difficult to emphasize them all equally. However, the upcoming curriculum with 19 competence aims altogether do not contain any categories such as oral communication or written communication. This could make it harder for teachers to know what to emphasize when determining a grade in both oral and written work in English.

Hovdhaugen et al. (2018) found differences in how the teachers thought when giving final overall achievement grades. Some teachers looked at the pupil's progress and final grades of the second term, while some felt that the first term was also of significance to include (2018: 10). Participants 2, 3 and 6 also expressed that; what the pupils did in the last few assessments showed more of where the pupils were at in terms of achievement. However, Participant 1 wanted the grade to be based on the assessment situations that the pupils had had, not only from their work in the 10th grade but for all of lower secondary school.

The examination paper graded by Participants 1-6 here serves as an illustration to discuss the teachers' approach when grading. The same exam paper managed to get three different grades from the participants. In reality, the exam paper received the grade 4, while the pupil's overall achievement grade was 3. Figure 20 shows that Participants 1 and 4 gave the same grade as the pupil received on his/her exam. The exam paper received the grade 3 by Participants 2, 5 and 6 which would be in correlation with the pupil's overall achievement grade, whereas Participant 3 gave the grade 2.

| Participant: | P1 | P2 | P3 | P4 | P5 | P6 |
|---------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| Grade: | 4 | 3 | 2 | 4 | 3 | 3 |

Figure 20: Participants 1-6 assessment of an exam paper from School C in schoolyear 2017/2018.

Many participants seemed to agree that the pupil did not write a proper answer according to the question in task 2A. The difference here is that Participant 3 felt that the pupil should receive the grade 2 because of it, whereas the others were not so harsh. However, Participants 4 and 6 did not mention that the pupil had not answered task 2A correctly. This questions whether this is something they look for or not. Participant 4 felt that the text communicated, but the language was not at the same level. Participant 2 was the one who really felt that the pupil might have written the wrong task number but would still give the exam the grade 3. She was also the only one who commented on missing headlines in tasks 1A and 1B, even though the examination guide does not refer to this as a requirement in either the document or in the characteristics of goal achievement template (UDIR 2020).

Even though there are many similarities in their assessment there were also some differences. Only 2 of them gave the grade that was actually awarded, which might tell us that the external examiners did not punish the pupil as harshly on the content in the long answer. This exam paper is evidence that teachers' grading practices might not be stable and that a shared

assessment culture is missing. Bøhn (2018) expresses the need for a shared assessment culture for teachers in order to improve the validity and reliability of teachers' assessment practices. He explains that:

The development of a shared assessment culture requires that teachers have a common understanding of: (i) What should be tested (the competence aims and criteria that are involved) and (ii) How performance should be assessed with respect to the competence aims and the criteria.

(Bøhn 2018: 236)

In light of the graded exam paper, one could assume that Participants 1-6 do not have a shared assessment culture even though they all have access to the same criteria from an examination guide. When compared to the individual oral statistics it was apparent that there were smaller deviations (see Figures 14 and 15). In Henrik Bøhn's doctoral dissertation (2016), he looked into how EFL teachers in Norway rated oral English exams at upper secondary schools. Bøhn (2016) found that "the teachers understood the main constructs in the same way, but they disagreed on some of the more specific performance aspects, such as 'pronunciation'." (Bøhn 2016: iii). Bøhn let 80 teachers from three counties evaluate the same, videotaped oral exam. The grades they came up with ranged from 2 to 4. However, most of them landed on the grade 3. One statistician who evaluated the results did not think the spread was very large, considering that there were no common assessment criteria and guidelines. This is interesting in view of the written exam implemented in this thesis, which actually has common assessment criteria and still managed to spread between the grades 2-4 from the small group of six teachers.

Perhaps Yorke (2011) is on to something stating that grading becomes more judgmental and less a matter of measurement "if broad categories are used as the basis of grading" (2011: 251), seeing as the written exam paper graded by Participants 1-6 received 3 different grades. In addition, there were large deviations presented from Schools A and C in written English, whereas oral English had small occasions of deviations at Schools A and C between exam- and overall achievement grades. It is evident from the interviews that there has to be clearer guidelines on what to emphasize for an overall achievement grade. However, in view of the grading of this particular exam paper, one could question whether there is a need for better guidelines or training on how to grade.

5.5 Grading practices in light of recommendations and regulations

Recommendations and regulations for grading in the English subject can be found in Regulations to the Education Act, the subject curriculum, examination guides and characteristics of goal achievement templates, all of which are presented in the theory chapter. As stated in the theoretical background, teachers are supposed to set overall achievement grades based on achievement exclusively (Dyrness and Dyrness 2008). In Regulations to the Education Act, it is evident that the basis for assessment in subjects are the competence aims in the curriculum (Regulations to the Education Act, §3-3). However, as Participant 3 points out, some competence aims could be perceived as difficult to assess. This raises a question as to how the various teachers in Norway address the different competence aims when setting overall achievement grades. We also find that “the prerequisites of the individual, absence or conditions related to order and conduct of the pupil, apprentice or the learning candidate should not be drawn into the assessment in subjects” in Regulations to the Education Act (§3-3). Nonetheless, as also mentioned in Section 5.4 above, Participant 6 explained that she determined a grade depending on what level the pupils were at. For low-achieving pupils, she hoped they were able to convey meaning and communicate and for high-achieving pupils she liked to use a two-pronged approach and look into their language and content. From this it is possible to assume that the pupils’ prerequisites are drawn into assessments by Participant 6.

On the other hand, there is a contradictory emphasis in the regulations that the pupils must attend class and participate actively (Regulations to the Education Act, §3-3). Effort could very easily be compared to a pupil’s participation in class. This would make it hard for teachers to know how to address pupils’ effort in class. There seemed to be common agreement between Participants 1-6 that competence and achievement were to be included in an overall achievement grade. However, a factor such as effort was also mentioned by several participants. The main essence in the pupils’ effort from the interviews, was that they should be rewarded for it and not punished if there was a lack of effort. The findings here would then correspond with findings from Guskey (2011), who states that teachers in fact “combine aspects of students’ achievement, attitude, responsibility, effort, and behavior into a single grade that’s recorded on a report card” (2011: 19).

Recommendations for grading shown in the examination guide (see Figure 4) and through the characteristics of goal achievement forms (see Figures 2 and 3) were perceived as unclear by

Participants 2 and 3. They thought that the school system needed clearer guidelines on how to assess their pupils. Bøhn (2016), who calls for national templates, assessment guides and achievement characteristics in oral examinations, expressed in an interview with *Aftenposten* that it was not a good idea to create templates too rigid either (Nipen 2018), the reason being that assessments could get too instrumental. Furthermore, he explained that it was important for teachers to still use discretion (Nipen 2018). With this in mind, the guidelines might be fairly clear when assessing written English, in order for teachers to also use discretion. However, to be able to address whether a pupil should receive the grade 2, 3, or 4 might be the toughest to handle. The grade 1 is barely mentioned in the examination guide and not mentioned at all in the characteristics of goal achievement, which makes it harder for teachers to know how much or how little a pupil must show in terms of competence in order to receive the lowest grade. There could be clearer borders concerning that grade. However, Participant 2 expressed that there were more than two external examiners who viewed an exam paper prior to it receiving the lowest grade. Nevertheless, there should also be a quality assurance regarding overall achievement grades, especially for pupils receiving the lowest grade.

Participants 1-6 were asked if they were familiar with the examination guide for teachers and pupils. There seemed to be a common understanding of the content from the examination guide. However, Participant 5 was not quite sure what the examination guide contained, seeing as she asked if the interviewer was referring to the competence aims. Many participants admitted that they could use the examination guide more. Nonetheless, Participant 6 did not use this in the classroom, saying that some things go automatically after having worked as a teacher for so long. This questions whether Participant 6 is aware that examination guides change from year to year, seeing as assessments and assessment practices are always changing. Participant 1 did not have a close relationship with this guide since she recently started working as a teacher. According to her, there is perhaps too much information for the pupils in this document. If she were to go through this with her pupils, she would have made a simplified version of it to include only the most central and important aspects of it to share with the pupils when they are in the 10th grade.

6 CONCLUSION

The present thesis was a mixed method study of deviations between overall achievement- and exam grades in the English subject in Norwegian lower secondary school. The researcher has looked into several aspects of assessment practices both internationally and in Norway. Materials from six semi-structured interviews with teachers from lower secondary schools and their assessment of an exam paper were used to gain a perspective of teachers' grading practices. In addition, public statistics of the average of overall achievement- and exam grades from the platform *Skoleporten* (2020), and individual grades from two lower secondary schools have been collected and analysed in order to see whether there are systematic deviations between the two types of assessments.

The data presented have shown that there have been deviations between overall achievement- and exam grades in the English subject the past five years in Norwegian lower secondary school. The differences between oral English and written English were less striking, seeing as this seems to be an understood trend in the Norwegian education system. Whether this trend in the English subject where pupils achieve lower grades in a written examination and higher grades in oral examinations is acceptable, was not evaluated in this thesis. However, all participants explained that some deviations might occur and did not see this as unusual. Overall, the national average grade scores convey that pupils achieve lower grades for a written English exam compared to the overall achievement grades. The national average grade scores in oral English show contradictory findings where pupils achieve higher grades for an oral exam as opposed to their overall achievement. Whether or not this was true for the majority of pupils having an English exam was not apparent in the present thesis. Factors influencing such deviations have been looked into and discussed, where exam tasks, school size, pupil composition, teachers and their assessment practices have all been looked into as factors of importance.

All teachers interviewed for the purpose of this thesis had similar grading practices. Nonetheless, there were some differences in what was emphasised for an overall achievement grade in both written and oral English. The differences became more apparent when looking into what the participants emphasized when a pupil was in-between two grades for their overall achievement. Some participants emphasized effort, where others emphasized homework. There seemed to be a common ground regarding progression, seeing as all

participants would emphasize it, if a pupil was in-between two grades. These findings would imply that teachers more often than not set grades according to their own views of what is correct and not what is recommended practice (see Yorke 2011: 251; see also Chapter 2.1, p.6).

All participants felt that a difference between overall achievements and exams were of importance. However, they had never changed their own grading practice due to a deviation between the two assessment forms. There was a unanimous agreement that the two assessment forms measure different achievements. The two assessment forms are comprised of different premises in grading itself. One happens on the basis of a teacher-pupil relationship characterized by a form of dialogue and the teacher's in-depth knowledge of the pupil as a person, while the other is conducted with mutual anonymity and a so-called examiner look. The two assessment forms each have their strengths with several reasons to defend, but also weaknesses that could be criticized. At the same time, an ever-recurring question is whether the Norwegian education system needs both an overall achievement- and an exam grade to ensure legitimacy in grading. This thesis suggests that both grades contribute to addressing important theoretical assessment issues separately, and that there are good reasons for both forms of assessments. However, the idea that one quality assures the other is far more problematic. It is therefore a question as to whether there should be alternative ways to quality assure the overall achievement and examination, both separately and based on the different premises of the two assessment forms.

To view teachers' grading practices in light of their own perceptions and the statistics of overall achievement- and exam grades alone had some limitations. It is difficult to generalize these grading practices, experiences and views of overall achievement- and exam grades seeing as there were only six female English teachers from western Norway represented. An equivalent distribution of gender, age and geographical areas was not apparent in the present thesis due to convenience sampling. Seeing as all participants were female, there were no male teachers to compare experiences and perspectives to. Some groups are therefore underrepresented in the present thesis, which could impact the diversity of this research. However, Skjong (2018), in an article presented in *Utdanningsnytt*, explains that only one of four teachers in Norway are men. This could be an explanation as to why there were no male volunteers. Perhaps a male English teacher would have different views to share, although any teacher, either male or female, has the same guidelines and regulations in their occupation.

There should therefore not be considerable shortcomings in the findings of this thesis because of this distribution.

There are several findings in this study which would support that grading becomes more judgemental and less a matter of measurement “if broad categories are used as the basis of grading” (Yorke 2011: 251). An oral exam seems to achieve fewer deviations than an exam paper in regard to overall achievements. The researcher therefore has a recommendation for future exam arrangements in written English where the pupils still have an exam like they do today. However, instead of getting a grade on that exam by an external examiner, the external examiner should know the pupil’s overall achievement grade. The external examiner could then assess whether or not this overall achievement grade is within the limits of reason and can be defended. The pupil’s overall achievement grade could then be tilted up or down a grade or remain the same by the external examiners. This would still keep the ‘fear’ of having to have an exam and having to be prepared, but at the same time not have the potential to make a big impact such as if the pupil was having a bad day and is dropped two or three grades. There will be dilemmas here as well, and the problems need to be discussed, but this would host a solution that addressed the benefits we want an exam to give, without penalizing pupils too much and adding such pressure on this particular day. This solution shares similarities with other assessment situations such as defending a master’s thesis.

The present thesis managed to get individual overall achievement- and exam grades from two schools. In order to make any generalisations of deviations and compare them between schools, there should ideally have been a larger pool of quantitative data. If the researcher would have been able to get nationwide data from all lower secondary schools in Norway it would have been possible to make any generalisations across schools and areas. However, this was not the intent of the present thesis, seeing as qualitative data was involved as well. Nevertheless, it is an interesting recommendation for further studies. The pupils’ views were also not included in the present study. For further study it would have been interesting to get the perspective of the pupils regarding their perception of an overall achievement grade as opposed to an exam grade. To get an insight into the pupils’ thoughts regarding their own achievements for both assessments forms could bring out perspectives that have not yet been analysed.

REFERENCES

- Aasen, P., Møller, J., Rye, E., Ottesen, E., Prøitz, T. S. & Hertzberg, F. (2012). *Kunnskapsløftet som styringsreform – et løft eller et løfte? Forvaltningsnivåenes og institusjonenes rolle i implementeringen av reformen*. Report 20/2012. Oslo: NIFU.
- Arasian, P. W. & Jones, A. M. (1993). 'The teacher as applied measurer: Realities of classroom measurement and assessment'. *Applied Measurement in Education*. 6. (241-254).
- Bachman, L. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.
- Bishop, J. H. (1992). 'Why U.S. students need incentives to learn'. *Educational Leadership*. 49(6). (15–18).
- Black, P. & William, D. (1998). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: GL Assessment.
- Black, P. & William, D. (2003). 'In praise of educational research: formative assessment'. *British Educational Research Journal*. 29(5). (623–637).
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, S. (2010). 'Validity in teachers' summative assessments'. *Assessment in Education: Principles, Policy & Practice*, 17(2). (215-232).
- Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Brookhart, S. M. (1991). 'Letter: Grading practices and validity'. *Educational Measurement: Issues and Practice*. 10(1). (35-36).
- Brookhart, S. M. (1994). 'Teachers' Grading Practices: Meaning and Values'. *Journal of Educational Measurement*. 30(2). (123-142).
- Brookhart, S.M. (2003). 'Developing Measurement Theory for Classroom Assessment Purposes and Uses'. *Educational Measurement: Issues and Practice*. 22(4). (5-12).
- Brookhart, S. M. (2013). 'The use of teacher judgement for summative assessment in the USA'. *Assessment in Education: Principles, Policy & Practice*. 20(1). (69-90).
- Brown, D. B. & Abeywickrama, P. (2010). *Language assessment: Principles and Classroom Practices*. (2nd edition). White Plains, NY: Pearson Education.
- Bøhn, H. (2015). 'Assessing Spoken EFL Without a Common Rating Scale'. *SAGE Open*. 5(4). (1-11).
- Bøhn, H. (2016). 'What is to Be Assessed? Teachers' Understanding of Constructs in an Oral English Examination in Norway'. (Doctoral dissertation). Oslo: University of Oslo.

- Bøhn, H. (2018). 'General Perspectives on Assessment'. In Bøhn, H., Dypedahl, M. & Myklevold, G. (eds), *Teaching and Learning English*. Oslo: Cappelen Damm. (231-247).
- Cheng, L. & Sun, Y. (2015). 'Teachers' grading decision making: Multiple influencing factors and methods'. *Language Assessment Quarterly*. 12(2). (213-233).
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. (4th edition). Thousand Oaks, California: Sage.
- Dale, E. L. & Wærness, J. I. (2006). *Vurdering og læring i en elevaktiv skole*. Oslo: Universitetsforlaget.
- Duncan, C. R. & Noonan, B. (2007). 'Factors Affecting Teachers' Grading and Assessment Practices'. *The Alberta Journal of Educational Research*. 53(1). (1-21).
- Dyrness, R. & Dyrness, A. (2008). 'Making the grade in middle school'. *Kappa Delta Pi Record*. 44(3). (114-118).
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies*. Oxford: Oxford University Press.
- Galloway, T. A., Kirkebøen, L. J. & Rønning, M. (2011). *Karakterpraksis i grunnskoler. Sammenheng mellom standpunkt- og eksamenskarakter*. Report 4/2011. Oslo/Kongsvinger: Statistisk sentralbyrå.
- Gravaas, B. C., Hægeland, T., Kirkebøen, L. J. & Steffensen, K. (2008). *Skoleresultater 2007. En kartlegging av karakterer fra grunnskoler og videregående skoler i Norge*. Note 2008/24. Oslo-Kongsvinger: Statistisk sentralbyrå.
- Guskey, T. (2011). 'Five obstacles to grading reform'. *Educational Leadership*, 69(3). (17-21).
- Harlen, W. (2005). 'Teachers' Summative Practices and Assessment for Learning Tensions and Synergies'. *The Curriculum Journal*, 16(2). (207-223).
- Harlen, W. & Gardner, J. (2010). 'Assessment to support learning'. In Gardner, J., Harlen, W., Hayward, L., Stobart, G. & Montgomery, M (eds). *Developing Teacher Assessment*. Maidenhead, England: McGraw-Hill/Open University Press. (15-28).
- Hasselgreen, A. & Ørevik, S. (2018). 'Assessment and Testing'. In Fenner, A.-B. & Skulstad, A. S. (eds). *Teaching English in the 21st century: Central issues in English didactics*. Bergen: Fagbokforlaget. (361-384).
- Haugstveit, T. B., Sjølie, G. & Øygarden, B. (2006). *Vurdering som profesjonsfaglig kompetanse. Rapport fra et KUPP-prosjekt*. Report nr. 5. Elverum: Høgskolen i Hedmark.
- Hovdhaugen, E., Seland, I., Lødding, B., Prøitz, T. S. & Vibe, N. (2014). *Karakterer i offentlige og private videregående skoler: En analyse av eksamens- og standpunkt-*

- karakter i norsk og matematikk og rutiner for standpunktvurdering i offentlige og private videregående skoler*. Report 24/2014. Oslo: NIFU.
- Hovdhaugen, E., Prøitz, T.S. & Seland, I. (2018). 'Eksamens- og standpunkt karakterer – to sider av samme sak?' *Acta Didactica*. 12(4). doi: <http://dx.doi.org/10.5617/adno.6276>.
- Hughes, G. (2014). *Ipsative assessment: Motivation through Marking Progress*. Houndsmills, Basingstoke: Palgrave MacMillan.
- Hægeland, T., Kirkebøen, L.J & Raaum, O. (2005). *Skoleresultater 2004. En kartlegging av karakterer frå grunn- og videregående skoler i Norge*. Note 2005/31. Oslo: Statistisk Sentralbyrå.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). 'Toward a Definition of Mixed Methods Research'. *Journal of Mixed Methods Research*. 1(2). (112–133).
- Kunnskapsdepartementet. (2019). *Nye læreplaner skal gi elevene tid til mer fordypning*. Pressemelding nr: 259-19. Accessed 20 January 2020. <https://www.regjeringen.no/no/aktuelt/nye-lareplaner-skal-gi-elevne-tid-til-mer-fordypning/id2678138/?expand=factbox2678140>
- Mausethagen, S., Prøitz, T. S. & Skedsmo, G. (2018). *Elevresultater mellom kontroll og utvikling*. Oslo: Fagbokforlaget.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education*. (2nd edition). Thousand Oaks, California: Sage.
- McMillan, J. H., Myran, S. & Workman, D. (2002). 'Elementary teachers' classroom assessments and grading practices'. *The Journal of Educational Research*. 95(4). (203-213).
- Mellingsæter, H. (2016). *Rektorer om nyutdannede lærere: Gode i fag, dårligere på sosial kompetanse*. Aftenposten. 18 June 2016. Accessed 9 April 2020. <https://www.aftenposten.no/norge/i/EvWvA/rektorer-om-nyutdannede-laerere-gode-i-fag-daarligere-paa-sosial-kompetanse>
- Mousavi, S. A. (2009). *An encyclopedic dictionary of language testing* (4th edition). Tehran: Rahnama.
- Nipen, K. (2018). '«Flytende engelsk», mente én sensor. «Dårlig flyt», sa den andre. Ingen vet hvem som har rett'. *Aftenposten*. 18 August 2018. Accessed 5 February 2020. <https://www.aftenposten.no/norge/i/kajjpa/flytende-engelsk-mente-en-sensor-daarlig-flyt-sa-den-andre-ingen-vet-hvem-som-har-rett>
- Prøitz, T. S. & Borgen, J. S. (2010). *Rettferdig standpunkt vurdering. Det (u)muliges kunst?* NIFU STEP-rapport 16/2010. Oslo: NIFU STEP.
- Prøitz, T. S. & Borgen J. S. (2011). 'Rettferdig standpunkt vurdering – på grensen av det mulige?' *Bedre skole*. 1/2011. (69-73).

- Prøitz, T. S. (2013). 'Variations in grading practice – subjects matter'. *Education Inquiry*. 4(3). (555–572).
- Regulations to the Education Act. (2006). FOR-2006-06-23-724. Accessed 13 January 2020. <<https://lovdata.no/dokument/SF/forskrift/2006-06-23-724>>
- Scriven, M. (1967). 'The methodology of evaluation'. In Tyler, R. W., Gagne, R. M. & Scriven, M. (eds). *Perspectives of curriculum evaluation*. 1. Chicago, IL: Rand McNally. (39–83).
- Skjong, H. (2018) 'Andelen mannlige lærere i grunnskolen synker stadig'. *Utdanningsnytt*. Accessed 3 March 2020. <<https://www.utdanningsnytt.no/fagartikkel-grunnskole-likestilling/andelen-mannlige-laerere-i-grunnskolen-synker-stadig/170094>>
- Skoleporten. (2020). Oslo: Utdanningsdirektoratet. Accessed 12 February 2020. <<https://skoleporten.udir.no/>>
- Statistics Norway (2013). Karakterer ved avsluttet grunnskole, 2013. Accessed 12 January 2020. <<https://www.ssb.no/utdanning/statistikker/kargrs/>>
- Stiggins, R.J., & Conclin, N.F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: SUNY Press.
- Stortingsmelding nr. 30 (2003-2004). *Kultur for læring*. Oslo: Utdannings- og forskningsdepartementet.
- Stortingsmelding nr. 16 (2006-2007). *Og ingen stod igjen. Tidlig innsats for livslang læring*. Oslo: Kunnskapsdepartementet
- Stortingsmelding nr. 20 (2012-2013). *På rett vei*. Oslo: Kunnskapsdepartementet.
- Stortingsmelding nr. 28 (2015-2016). *Fag – fordypning – forståelse – en fornyelse av kunnskapsløftet*. Oslo: Kunnskapsdepartementet.
- Tierney, R. D. (2015). 'Altered grades: A grey zone in the ethics of classroom assessment'. *Assessment Matters*. 8. (5-30).
- Tveit, S. & Olsen, R. V. (2018). 'Eksamens mange roller i sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring'. *Acta Didactica Norge*. 12(4). Art. 18.
- UDIR - Norwegian Directorate for Education and Training. (2007). *Sluttrapport. Oppdragsbrev nr. 6 - 2007 om tiltak knyttet til individvurdering i skole og fag- og yrkesopplæring*. Oslo: Utdanningsdirektoratet. Accessed 15 January 2020. <https://www.udir.no/globalassets/upload/forskning/5/bedre_vurderingspraksis_sluttrapport_til_kd.pdf>
- UDIR - Norwegian Directorate for Education and Training. (2013). *The English subject curriculum*. Accessed 17 November 2019. <<https://www.udir.no/k106/ENG1-032?plang=eng>>

- UDIR - Norwegian Directorate for Education and Training. (2017). *Engelsk: kjennetegn på måloppnåing*. Accessed 18 November 2019. <<https://www.udir.no/laring-og-trivsel/vurdering/sluttvurdering/Engelsk-kjenneteikn-pa-maloppnaing/>>
- UDIR - Norwegian Directorate for Education and Training. (2019). *Standpunktvrdering*. Accessed 17 November 2019. <<https://www.udir.no/laring-og-trivsel/vurdering/sluttvurdering/standpunktvrdering/>>
- UDIR - Norwegian Directorate for Education and Training. (2020). *Eksamensveiledning – om vrdering av eksamensbesvarelser*. Accessed 13 March 2020. <[file:///Users/janehodne/Downloads/V2020_Eksamenveiledning%20\(3\).pdf](file:///Users/janehodne/Downloads/V2020_Eksamenveiledning%20(3).pdf)>
- Utdanningspeilet. (2008). *Tall og analyse av grunnopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Accessed 15 January 2020. <https://www.udir.no/globalassets/filer/tall-og-forskning/rapporter/2009/5/us08_110909.pdf>
- Utdanningspeilet. (2009). *Tall og analyse av grunnopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Accessed 16 January 2020. <https://www.udir.no/globalassets/filer/tall-og-forskning/rapporter/2009/5/utd_speilet_09_u_blanksider_net.pdf>
- Utdanningspeilet. (2011). *Tall og analyse av grunnopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Accessed 16 January 2020. <https://www.udir.no/globalassets/filer/tall-og-forskning/forskningsrapporter/utdanningspeilet_2011.pdf>
- Utdanningspeilet. (2014). *Tall og analyse av grunnopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Accessed 16 January 2020. <<http://utdanningspeilet.udir.no/2014/wp-content/uploads/2014/06/Utdanningspeilet-2014-utskriftsversjon.pdf>>
- Utdanningspeilet. (2016). *Tall og analyse av grunnopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Accessed 16 January 2020. <http://utdanningspeilet.udir.no/2016/wp-content/uploads/2016/06/Utdanningspeilet_2016.pdf>
- Utdanningspeilet. (2017). *Tall og analyse av grunnopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Accessed 16 January 2020. <<http://utdanningspeilet.udir.no/2017/innhold/del-5/5-2-grunnskolepoeng-og-karakterer-etter-10-trinn/>>
- Utdanningspeilet. (2019). *Tall og analyse i grunnopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Accessed 17 January 2020. <<https://www.udir.no/tall-og-forskning/finn-forskning/tema/utdanningspeilet-2019/>>
- Wilson, R. J. (1996). *Assessing students in classrooms and schools*. Scarborough, Canada: Allyn & Bacon.

- Wyatt-Smith, C. & Klenowski, V. (2013). 'Explicit, latent and meta-criteria: Types of criteria at play in professional judgement practice'. *Assessment in Education: Principles, Policy & Practice*. 20(1). (35-52).
- Yesbeck, D. M. (2011). 'Grading practices: Teachers' considerations of academic and non-academic factors'. (Doctoral dissertation). Richmond, VA: Virginia Commonwealth University.
- Yorke, M. (2011). 'Summative assessment: dealing with the 'measurement fallacy''. *Studies in Higher Education*. 36(3). (251–273).

APPENDIX 1: Interview guide – English version

Background information

1. Age:
2. Gender:
3. Years working as a teacher:

Experiences with setting grades

4. How much experience do you have with setting overall achievement grades in lower secondary schools?
5. Did you feel, after completing your teacher education, that you were qualified for grading assessments?
6. Have you participated in courses/training in assessment/grading?
 - a. Could you describe the experience?
7. Have you been an external examiner in English at a lower secondary school?
 - a. In oral English?
 - b. In written English?
8. Have you participated in examiner training?
 - a. Could you describe the experience?

Overall achievement grades

9. Could you describe how you work with setting overall achievement grades in oral English at your school?
10. Could you describe how you work with setting overall achievement grades in written English at your school?
11. Do you use any aids to support the grading of pupils? (for example: simple tests, custom designed tests, survey tests, presentations, group assignments, home assignments, the UDIR's examination assignments, etc.)
12. To what extent do you use the national curriculum/local curriculum as a benchmark/anchor point when setting an overall achievement grade?
13. What do you emphasize when determining a grade in oral English?
14. What do you emphasize when determining a grade in written English?
15. Is there something you emphasize more/less, for example, if a pupil is balancing between two grades? (competence, attendance, homework, effort, achievement or progression.)

- a. Do you have an example?
- 16. Do you use the entire grade scale?
- 17. Do you think it is possible to get a fair assessment of the pupils?
 - a. What is difficult, if anything?
- 18. Do you cooperate with colleagues on grading?
- 19. What kind of help or support do you get in the process of grading?
- 20. Do you think you have enough support in this work?

The relationship between overall achievement grades and exam grades

- 21. How do you think the relationship between overall achievement grades and exam grades should be?
- 22. Would you say that exam grades act as quality assurance of how you and your colleagues have set the overall achievement grades?
 - a. Has this led to changes in your own grading practice?
- 23. Are you familiar with the UDIR's examination guide for pupils and teachers?
- 24. Do you use this document actively yourself in teaching planning or as information to the pupils? (bring and show)

APPENDIX 2: Interview guide – Norwegian version

Bakgrunnsinformasjon:

1. Alder:
2. Kjønn:
3. Hvor mange år har du arbeidet som lærer?

Hvilke erfaringer har du med karaktersetting:

4. Hvor mye erfaring har du med standpunktkaraktersetting i Engelsk på ungdomsskolen?
5. Følte du etter ferdig lærerutdanning at du var kvalifisert til å sette karakterer?
6. Har du deltatt i kursing/opplæring i vurdering/karaktersetting?
 - a. Hvordan var det?
7. Har du vært ekstern sensor ved eksamen i Engelsk?
 - a. I muntlig Engelsk?
 - b. I skriftlig Engelsk?
8. Har du deltatt på sensor opplæring?
 - a. Hvordan var det?

Standpunktkarakterer:

9. Kan du beskrive hvordan du arbeider med å sette standpunktkarakter i Engelsk muntlig her på skolen?
10. Kan du beskrive hvordan du arbeider med å sette standpunktkarakter i Engelsk skriftlig?
11. Bruker du noen hjelpemidler for å støtte karaktersettingen av elevene? (F.eks. tester, egne utformede prøver, kartleggingsprøver, fremføringer, gruppeoppgaver, hjemmeoppgaver, UDIR's eksamensoppgaver.)
12. I hvilken grad bruker du nasjonal læreplan/lokal læreplan som målestokk/forankringspunkt når du setter standpunktkarakter?
13. Hva legger du vekt på når du skal bestemme en karakter i Engelsk muntlig?
14. Hva legger du vekt på når du skal bestemme en karakter i Engelsk Skriftlig?
15. Er det noe du vektlegger mer/mindre for eksempel om en elev står i vippeposisjon mellom to karakterer? (kompetanse, oppmøte, lekser, innsats, prestasjon eller progresjon.)

a. Har du et eksempel?

16. Tar du i bruk hele karakterskalaen?

17. Synes du det er mulig å få rettferdig vurdering av elevene?

a. Hva er eventuelt vanskelig?

18. Samarbeider du med kollegaer om karaktersetting?

19. Hva slags hjelp eller støtte får du i arbeidet med karaktersetting?

20. Mener du at du har god nok støtte i dette arbeidet?

Forholdet mellom standpunktkarakter og eksamenskarakter:

21. Hvordan mener du forholdet mellom standpunktkarakterer og eksamenskarakterer skal være?

22. Vil du si at eksamen er en kvalitetssikring på hvordan du og dine kollegaer har satt standpunktkarakterer?

a. Har det ført til endringer i hvordan du setter standpunktkarakterer?

23. Er du kjent med UDIR's eksamensveiledning til elever og lærere?

24. Bruker du dette dokumentet aktivt selv i undervisningsplanleggingen evt. som informasjon til elevene? (ha med og vise).

APPENDIX 3: Interview summary – Participant 1

Participant 1 was a 23-year-old female who graduated from the university as a primary school teacher in spring 2019. She has worked as a teacher for 5 months. Currently, she is a teacher working with 8th grade pupils. She has little experience with setting overall achievement grades due to her age. However, she worked as a substitute teacher last schoolyear in the 9th grade and had to set their overall achievement grades in Norwegian, but not in English.

Participant 1 had experience as an English teacher with formative assessment from these last 5 months in English. When being asked if she felt qualified for grading assessments after her teacher education, she said that she was nervous when encountering her first assessment to grade. She therefore consulted her colleagues when in doubt. In view of her answer, she locked assessments while grading them on her iPad. By doing so she could correct the pupils' texts without them seeing and could consult colleagues when doubting whether a pupil should get the grade 3 or 4 for instance. She also found it helpful to ask colleagues when being unsure whether a pupil have missed/misunderstood the task.

She has the sole responsibility of the subject English for the entire grade. However, the first assessment was made together with her colleagues. Participant 1 explained that they were 3 English teachers who worked as a team. Participant 1 taught one of the six classes from that grade. A colleague of hers has also newly finished her teacher education and started working as a teacher last fall. The third colleague had been working at the school for 12 years and had much experience in the field as a teacher. The two newly graduated teachers therefore often leaned on this colleague for help and guidance.

Participant 1 explained herself as being lucky with her teaching practices, whereas she has always been offered to look at papers to grade. Furthermore, she felt very included with setting grades in her teaching practices. When studying English at the university they were also being given example texts written by pupils to grade. They had a lot of practical didactics in the English subjects. Participant 1 boasted her 2 semesters of English saying that those were the best semesters of her entire teacher programme. Their teachers were very clear. She had not participated in any courses or training in assessment other than what she had gone through in her teacher programme at the university. Due to Participant 1's age, she had not

been an external examiner in either oral or written English. Consequently, she had not participated in any examiner training either.

Since Participant 1 had little experience with overall achievement grades, she was being asked hypothetical questions for the future of how she would work with setting overall achievement grades in oral and written English at her school. When being asked of the oral overall achievement grade, she wanted the grade to be based on the assessment situations that the pupils had had, not only from their work in the 10th grade. In light of these assessment situations, she wanted to be able to look back at where the pupils started in the 8th grade and look at their progression throughout these 3 years. Participant 1 wanted to be able to document their progression, both their oral skills and writing skills. Due to great technology she either filmed the pupils' presentations or they would hand in an audio-recording or a video of themselves for oral assessments. She explained that if she was able to document this well enough, it would be easier to look at their progression.

With this in mind, Participant 1 also thought it was important to be verbally active during class. She emphasized that being verbally active during class does not necessarily mean that the pupils are actively raising their hands to answer questions in class. However, the pupils should show participation in group work and be able to discuss with classmates in groups. After all, there were several ways to show that they were engaged in the classroom, she claimed.

Participant 1 moved on to say that their school strongly believed in cooperative learning and used this as a method in the classroom. The pupils in the classroom were usually paired in 4 groups. To be verbally active in front of the whole class was easier when letting the pupils talk in groups beforehand. She explained that it was much safer to tell what the group have discussed rather than what the pupils thought themselves. Participant 1 would include pupils' activity engagement from these situations when giving an overall achievement grade. Homework was not an important aspect to include when giving an oral overall achievement grade according to Participant 1. Nevertheless, if a pupil had not done any homework in the subject, that would be another thing. The pupil's homework involved the effort the pupil had in the subject.

When being asked how Participant 1 would work on setting written overall achievement grades, she explained that the main focus here would also be the assessments situations. Secondly, she would also look at the pupils' engagement in class. It was easier for her to look at progression seeing as all texts were saved digitally and the teacher has access to see the pupils' progression. When being asked a follow up question as to whether she would emphasize the last semester more than prior semesters, she answered that it was the last semester the pupils got to show everything they have learned. However, if the pupils had had a steady progression over the three years, then it should definitely benefit the pupil.

Participant 1 would use presentations and group assignments to support the overall achievement grades she had set. She explained that she had not used any general assessment tests in particular. However, the pupils had a national test, but Participant 1 had not taken this test into any consideration. She used this test to look at the pupils' level and skills in English because she had little prior knowledge of this. Moreover, Participant 1 explained that she also used adapted tests for those who needed it. The researcher gave a follow up question as to whether those pupils could then achieve all grades from the grade scale. Participant 1 explained that she would not take away the opportunity for them to get the grade 5 if they were worthy of it. However, these pupils had adapted tests for a reason, and it was not realistic that they would achieve such a high grade. On the other hand, the high proficiency pupils in English get extra tasks to work on when being done with the work from class. Participant 1 said that when setting overall achievement grades on those pupils, she would have to really look at those extra tasks because they showed what the pupil is capable of beyond the curriculum. They had already shown that they know everything else being discussed in class.

Participant 1 was also asked to what extent she would use the national curriculum/local curriculum as a benchmark when setting an overall achievement grade? Since she had not set any overall achievement grades yet, this was also a hypothetical situation. With that in mind, she answered that she always used the competence aims when designing an assessment.

When deciding a grade in oral English, Participant 1 definitely emphasized the pupils' intonation. Intonation seemed to be most important for her, especially in the 8th grade because the pupils have a fact-based approach. By saying this she meant that the pupils have a lack of reflection and personal anecdotes. When deciding a grade in written English she looked at

some of the same qualities. She looked at how they write in terms of orthography, sentence structure, but she would also emphasize the importance of the overall impression of a text. This included how the text was structured and the formal requirements in terms of punctuation, font size and headlines.

Participant 1 would emphasize progression a lot if a pupil would be balancing between two grades. Competence is also something she would take into consideration. An example would be if a pupil got an assignment which was not suited for him/her, but Participant 1 knew that the pupil's competence indicated that he/she could have shown a lot more than what the pupil did from that assignment. Attendance was not something she would take into much consideration. All in all, she would emphasize progression, competence and effort if the pupil worked hard in class but did not master everything.

When being asked if she would use the whole grade scale, Participant 1 explained that it was difficult to set the grade 1. She has now set the grade 1 two times which hurt a little. She reflected on this by saying she thought it was difficult because she was new and not very experienced. Furthermore, Participant 1 also thought it was difficult because of the culture presented at the school. She had the impression that one does not give a pupil the lowest grade unless they actually could have written their name. She moved on to explain that it takes a lot to get the grade 1. An example was given of an assignment in the subject English specialisation. The pupils were supposed to write half a page whereas one pupil only wrote 2 sentences. This pupil got the grade 1, which was also because she knew that this pupil was able to do more than that.

On the contrary, Participant 1 also said that she has given multiple high grades such as 6 on assignments. She explained that it was frustrating for her to listen to teachers who said they did not give the grade 6 unless it was an overall achievement grade or unless the pupils got multiple grades consisting of 5+ in a row. The highest grade should always be achievable according to Participant 1. She explained that the grade did only show the pupil's competence at that particular day/assignment or on that particular theme or topic. A grade on a single assignment did not necessarily show a pupil's competence in the whole subject as such, but rather at one assessment.

When being asked whether she thought it was possible to get a fair assessment of the pupils, she responded by saying it was difficult. The hardest aspect of the job was to not let the personal relation between teacher and pupil interfere when grading. Participant 1 explained it was especially difficult for her since her pupils at that moment were her first pupils and she cared for them very deeply. However, she liked to think that this benefited her pupils, since she tried to convey that she liked all her pupils. She tried to show this towards them as well. An example given by Participant 1 was when a pupil is feeling sick on a writing day and tries to finish early to go home to bed. This is a difficult situation whereas this pupil might not get a fair assessment. She asked a rhetorical question whether that performance should count as much as the others. In conclusion, Participant 1 felt that this was wrong.

When looking into the help and support Participant 1 got when grading she said that she always had colleagues to talk to. From her knowledge, they did not get any other support apart from this. When having meetings and such at school, they did not have the time to talk about grading in a given subject. This had therefore never been a topic at their school meetings. All in all, she would prefer some guidance in grading as a newly graduated teacher. However, she did not know who could provide that for her. Participant 1 was provided with a supervisor, but he was not an English teacher and their sessions were only once a month.

When being asked how the relationship between overall achievement grades and exam grades should be, Participant 1 believed that overall achievement grades should always give pupils the benefit of the doubt. She was asked whether it was okay if a pupil got a 3 for the overall achievement and a 5 for an exam. She believed that to be fair if that was what the pupil had managed to accomplish. Although it would be more difficult if it was the other way around according to Participant 1. When being asked whether Participant 1 perceived an exam to be a quality assurance she said that the examiners could also have either a good day or a bad day seeing as they were also human. The examiners could be strict, or they could be kind when grading. Participant 1 found grading difficult in general due to these factors.

Participant 1 was familiar with UDIR's examination guide for pupils and teachers. She did not have a close relationship with this guide but had gone through it. There was perhaps too much information for the pupils in this document according to Participant 1. If she were to go through this with her pupils, she would make a simplified version of it to include the most central and important aspects of it to when they were in the 10th grade.

APPENDIX 4: Interview summary – Participant 2

Participant 2 was a 46-year-old female who has been working as a teacher for 11 years. Prior to this, she was a hairdresser. She had 3 years of experience giving overall achievement grades. Participant 2 felt far from qualified to grade assignments after completing her teacher education. She said that teachers are really dependent on others and to get a team that wish to share and help. It was important for her to feel like she was met with kind and helpful teachers around to ask for help. Furthermore, Participant 2 said that it was not easy to ask for help if a teacher felt like he/she was being met with unhelpful teachers. She still thought it was difficult to grade pupils. That was one of the worst things she did as a teacher.

Participant 2 had not participated in any other coursing or training in grading apart from examiner training. She did not believe there have been any courses or training available in grading that she had heard of. Other than that, Participant 2 had been an external examiner in both oral and written English. She received examiner training when being an external examiner in written English, but never in oral English. Participant 2 believed she had participated as an external examiner 4 times. The examiner training involved reading different papers to grade. After doing so teachers would get together with other English teachers and discuss their grading and why they have chosen the different grades for the exam papers. Afterwards, they would gather to see many examples which a selected test group has graded. Given the above, Participant 2 thought that these examiner trainings had been very useful.

The grade she worked with has been working as a team. Her colleagues have read assignments together, discussed assignments and helped each other. That has been a good help for Participant 2.

When being asked how Participant 2 worked with setting overall achievement grades in oral English, she said that they have been working theme based these last few years. That was to say; they chose texts, tasks, and assignments that were linked to the theme they were working on. They were having more subject conversations with the pupils as an assessment than presentations. Participant 2 believed that the pupils were more comfortable to talk freely in a conversation than they were in presentations. An advantage of having conversations was that the pupils could not practice on what to say when having a conversation. That way, a teacher got a clearer picture of the pupils' competence of the theme. Participant 2 was then asked

more specifically what she included in her overall achievement grades in oral English. She answered by saying she kept track of what the pupils were doing all the time. However, the pupil had to know what they were being assessed on and when. She explained that they sometimes have a mini talk in class with peers and small assignments. All in all, Participant 2 said that the assessment situations form the basis for the overall achievement grade. What the pupils did in class often corresponded with the grade they achieved in assessments. She explained that a lack of effort in class would not pull the pupils down a grade. On the other hand, if a pupil was in-between the grades of 4 and 5 and was active in class, he/she would benefit from that and perhaps receive the grade 5. Participant 2 explained that it was important to convey this way of thinking towards the pupils.

English was a progression subject according to Participant 2 and it applied both in oral and written English. What the pupils did in the last few assessments showed more of where the pupils were at in terms of achievement. However, some pupils might be very unfortunate on the last assignment. That could be very challenging if a teacher was starting out with his/her pupils in the 10th grade, because he/she did not know them that well. It was harder for a teacher to see if a pupil was unfortunate or having a bad day if he/she has only been with them for a short amount of time according to Participant 2. Apart from that, she thought that some small and some big assessments formed a good picture of where the pupils were on the grade scale.

Likewise, in written English, Participant 2 based her overall achievement grades on some small and some big assessments. Usually, she tended to use old exam assignments at the end of the schoolyear. This was done in order for the pupils to become acquainted with this type of assignment, because that is what meets them eventually.

Participant 2 definitely used the UDIR's examination assignments and custom designed tests to support her grading of pupils. General assessment tests other than the required national tests, were not something she had used with her pupils. Participant 2 explained that it was important that the assignments the pupils received, were linked to the theme they have been working on. Sometimes she designed tasks for the pupils herself or she used Cappelen Damm and other resources to find relevant tasks. Ultimately, Participant 2 stated that the further the pupils got in the 10th grade, the more focus there would be on old exam assignments.

When being asked to what extent she used the national/local curriculum as a benchmark when setting overall achievement grades, Participant 2 said that she used the competence aims all the time. However, some were used more than others she claimed. Some of the competence aims were consistent all the time, while others were more specific to a part of the syllabus. Participant 2 expressed that she used the national curriculum a lot, especially the competence aims. She revealed that she could be better at looking at the purpose of the subject in the curriculum more often.

Participant 2 felt that she did not emphasize anything more or less in oral English. She thought that a grade should be a combination. There were often professional aims where the pupils should show knowledge and history of for example countries, which meant that we could not look at just the English language either. She stated that she has often tilted a grade up where a pupil shows a lot of extra-linguistic knowledge. However, she also has to understand what a pupil is trying to convey. Participant 2 said that she cannot avoid looking at either language or knowledge. The two factors should be looked at in a combination. In a continuous assessment, Participant 2 could give a split grade, for instance 2/3. She often gave comments on such assessments stating that the content was at a higher level than the language. She shared these thoughts in both oral and written English assignments. When Participant 2 had been an examiner she explained that one part involved the linguistics, another the content and the last one involved the text structure. Her grade was based on a combination of these parts, but it was important that the text communicated when reading it.

When being asked if there was something she emphasized more or less if a pupil was in-between two grades, Participant 2 explained that attendance was not relevant at all. If a pupil had not been in class, then the pupil would not know what was happening in class either. This would be evident prior to setting the overall achievement grades she said. The pupils' competence, however, was very important. Competence and progression were the two factors that Participant 2 would emphasize the most if a pupil was in-between two grades. She explained that these pupils often got the chance to write a text or perform a small task to show their competence. However, this was rarely done by Participant 2. She felt that after a year with the pupils, a teacher should be able to give them an overall achievement grade without being too uncertain. Likewise, Participant 2 claimed that an old exam assignment would reveal a lot of what the pupils knew and what they did not.

Participant 2 used the entire grade scale when setting overall achievement grades. She said that she has to sometimes, it was not a choice. If a pupil was at a very low level, then she would have to give that pupil 1. She has experienced giving some exam papers the grade 1 as well, saying that it was not just her setting that grade when they met up to discuss.

Furthermore, Participant 2 explained that if they gave an exam paper the grade 1, that exam paper would be sent to a selected test group for reviewing. At least 5 other teachers would then look through it to see if that pupil should receive the grade 1. When looking at the other end of the grade scale, Participant 2 said that a 6 was definitely achievable.

One would like to think that one gives a fair assessment of the pupils, Participant 2 said. She experienced when getting other teachers to read papers, that they had a similar mindset. However, being coloured by her pupils was hard to avoid. Participant 2 explained that a teacher's job was very subjective. She also believed that teachers were kinder with their own pupils than for example when being an external examiner. It was human to be affected by the efforts and hard work a teacher observes from his/her pupils. Nonetheless, when sending a text to a fellow colleague, which do not have a relationship with the pupil, she experienced that they thought alike on what that particular pupil achieved. Participant 2 therefore believed that it was possible to get a fair assessment of the pupils.

When talking about the support Participant 2 got in the process of grading she brought up her colleagues. Her fellow colleagues were her support system when grading. She explained that assessing was perhaps one of the most difficult tasks as a teacher. Participant 2 certainly thought it would have been good with even more support when grading. Perhaps there should be clearer guidelines to follow from UDIR. The characteristics of goal achievement was not very clear on the differences between the different grades. Participant 2 thought that we could never be fully trained in assessing, so it was important to always stay updated.

Participant 2 was asked how she thought the relationship between overall achievement grades and exam grades should be. She said that an exam grade should hopefully be the same as the overall achievement grade. However, sometimes it may depend on the task itself; how it was made. The theme might be very good, leading to the pupils having a lot to write about. If this was an area of interest for the pupil, then it was much more likely that the pupil could achieve a better grade. On the other hand, if the theme revolves around something a pupil had little knowledge of, then it would be more difficult to answer the exam. Participant 2 thought that

this could be difficult to avoid. Nonetheless, she believed that the pupils should be at the same level regarding their structure and language. Her school had received feedback that their overall achievement- and exam grades have been very similar.

Participant 2 explained that the exam was a quality assurance in a way because there were teachers with no relationship at all towards the pupils who were grading the exams. She said that she was very happy when the exam results were in and they matched the overall achievement grades which she had set. Due to this, she realized that she did agree that exam grades were a quality assurance of how she had set the overall achievement grades. Nonetheless, seeing the exam grades given to her pupils had not led to any changes in her grading practice. However, being an external examiner herself, has made her more aware. She thought that the competence a teacher learns from such an experience was helpful when grading her pupils' overall achievement.

Participant 2 was familiar with the UDIR's examination guide for pupils and teachers, explaining that she looked at this guide when grading exam assignments. She used it in class prior to writing days so the pupils could get a glimpse of what was expected of them. They did not go through it in detail in class. However, it was very useful to see what aims were being looked at and what the criteria were. Some parts of the guide were being used she explained.

APPENDIX 5: Interview summary – Participant 3

Participant 3 was a 40-year-old female who has worked as a teacher for 14 years, starting in 2005. She did not feel qualified to grade pupils after completing her teacher education.

English was not Participant 3's primary subject, but she did complete 30 ECTS Credits in the English subject in her undergraduate teacher education. After completing her undergraduate teacher education, she studied a new semester of English and received 30 more ECTS Credits in English. From her courses at the university she gained a lot of knowledge of the English language, but they never had any teaching in assessment and grading. She started learning how to grade pupils when she started working and explained that she still to this day did not find it easy. In other words; the more she knew, the more she wanted to know and questioned everything.

Participant 3 explained that she had not participated in any formal course/training in assessment/grading. However, they had had sessions at her school where the primary focus was grading. Grades and assessment had been a theme when having competence enhancement for teachers in the afternoon meetings at Participant 3's school. This had happened on occasion, but it was not inserted into any system and there has not been a huge focus on this topic. What Participant 3 has learned regarding grading has been taught through colleagues where they have met, discussed and sent texts to each other. She also used UDIR to look at pupils' goal achievement.

When being asked whether Participant 3 had been an external examiner, she answered by saying she has been an external examiner in oral English, which implied that she had not been an external examiner in written English. Consequently, she had never participated in any examiner training prior to the exams. Participant 3 explained that only the external examiners for written English received training, but oral examiners were just thrown into deep water.

After receiving background information and talking about Participant 3's experience, the conversation moved on to overall achievement grades. Participant 3 was asked how she worked with setting overall achievement grades in oral English. She explained that the pupils only got one grade per semester at their school. They tried to avoid grading the pupils too much, due to the stress this puts on pupils. However, Participant 3 has her own system and documents the pupils process along the way. She used Visma Flyt to write half year evaluations of the pupils. Participant 3 wrote down what the pupils had shown of competence

and formative feedback each semester. She explained that first by the end of the 10th grade, she would be able to evaluate the pupils in terms of all competence aims and what they have achieved. Furthermore, she claimed that this was simple in some areas and difficult in others. Not all competence aims were as easy to assess. She explained that some of the competence aims were very big and diffuse. Even though it was difficult, Participant 3 tried to have an overall assessment of what the pupils have achieved. Moreover, she said that the pupils received more grades in the 10th grade because the pupils should then be aware of where they are at and what they should work on to improve.

Participant 3 moved on from talking general to more specific of written English. She tried to give the pupils as many opportunities as possible to show their competence. Participant 3 had a big focus on process-oriented work to understand how they worked, how they used their sources and if they were critical to the sources being used and how they referred to their sources. Participant 3 tried to document as much as possible of the pupils learning and thought process in order to give them a proper grade. Participant 3 explained that the last thing the pupils did should matter the most. However, if a pupil has shown very little competence at the end and she knew that the pupil could do better than this, then she tried to give that pupil as many opportunities as possible to show his/her competence.

When asked if she used any aids to support her grading of pupils, Participant 3 felt she used them all such as; tests, custom designed tests, general assessment tests, presentations, group assignments, home assignments and the UDIR's examination tasks. The less commonly used by her was probably general assessment tests which were formally made.

Participant 3 explained that pupil participation was very important. She wanted her pupils to have a say in how the assessments were and what the criteria should be. She could present a theme or topic to her class and share some thoughts of what she thought they should do and how this project should end. An example given by Participant 3 was that the project should end with the pupils making picture books about fantasy literature. From there she wanted the pupils to decide together; what tools could be used to make this, how could we do this, are we able to do something both written and oral. As a result, Participant 3 got a lot of input from the pupils and the opportunity to adapt the task as best as possible.

Technology played an important part for Participant 3 which gave her a lot of opportunities. She has pupils who managed to have a presentation in front of class and some who did not.

By using technology, these pupils were able to record a presentation and send to the teacher. The pupils were much more relaxed when doing so and could talk more freely, which could be difficult in front of the class. Participant 3 pointed out that when having an oral English exam, the pupils did not have a presentation in front of a whole group, but 2 examiners, which was very different. It did not feel natural for Participant 3 to force all pupils to have a presentation in front of their classmates. She wanted them to try of course, but the presentation was not the goal in itself. Participant 3 wanted the pupils to show her their oral competence.

Old exam assignments were not used frequently in the 8th grade by Participant 3. The pupils should be familiar with how an exam works, the preparation material and what was expected of them. Prior to the written exam, all pupils would have gone through two to three old exam assignments over the course of lower secondary school. She tended to use them when they have mock exams in English. Participant 3 always used one of the old exam assignments as a long-term process. This meant that they worked thoroughly through it together in class for several weeks. By doing so, pupils got the time to understand the tasks properly and ask questions along the way when writing their texts. She felt this was a good way to let the pupils get to know a preparation material and how they should work with it to prepare themselves for an exam.

Participant 3 used the national curriculum to a high extent as a benchmark when setting overall achievement grades. She looked into the pupils' achievement based on the competence aims from UDIR. However, Participant 3 emphasized that they were about to change when the new curriculum comes next fall. Participant 3's problem with the competence aims has been that there were too many. Some of the competence aims were also formulated in a way that was difficult to assess. She talked about the UDIR's characteristics of goal achievement, saying that she used the form when setting overall achievement grades. Optimistically, Participant 3 explained that it was going to be great just having to set one overall achievement grade with the new curriculum, but that she had to think in a different manner. However, she thought it was sad for example if a pupil was very good orally in English but had poor writing skills in English. Then she would not be able to give the pupil that good grade in oral English. Instead, the grade would be a mix of the pupil's oral and written skills. Participant 3 believed that teachers will be forced to think differently with the upcoming new curriculum.

Furthermore, the conversation led to what Participant 3 emphasized when determining a grade in oral English. Participant 3 was concerned with not being fooled by a pupil who has very good flow and intonation in the language. She has learned that there were pupils who have a really good intonation but who could not use the language in a proper way to communicate. Language understanding was very important in oral communication. By saying this she meant that the pupils managed to use the language, build sentences, that the sentences were good and that they had a good vocabulary. A pupil could get a high grade if he/she showed good confidence and understanding of the language, even if there were some glimpses of Norwegian intonation at times. However, she was concerned with the pupils' intonation as well, that they master the sounds of the words. It was not required of them to sound like a native, because they were after all Norwegian pupils. Participant 3 summed up by saying that a grade in oral English would be an overall assessment of all these factors.

When moving on to what Participant 3 emphasized when determining a grade in written English, she said that there as well, the pupils needed to use the language to convey something reasonable. The pupils should be able to build sentences here as well, and to have an understanding of the English grammar. Using the English language to convey content was very important to Participant 3. The extent to which they managed to do so would help determine their grade. There would also be an overall assessment of these factors when grading according to Participant 3. She gave an example saying that if a pupil built good sentences, showed understanding and wrote good texts, then he/she could achieve a high grade even though there were some typos. The grade depended on how much competence the pupil managed to show, in terms of these factors.

Participant 3 would emphasize progression if a pupil was in-between two grades for the overall achievement. That a pupil has shown a development would help tilt the pupil a grade up. She gave an example of a pupil who had had two really good assessments but not done any homework. If such a situation should occur, Participant 3 said that she would have to look at the competence the pupil had shown and not punish him/her for what they did not show. It was not fair to give a pupil a lower grade than they have shown in competence according to Participant 3.

Homework and effort in itself should not have anything to say on the pupils' grade in English. Participant 3 explained that this should only influence their grade in order. However, Participant 3 claimed that their effort in class and with homework was often related to what they achieved in the subject. Participant 3 often gave the pupils homework which gave them the opportunity to show their competence. An example of that was when the pupils' homework was to read in an audio file and submit it to the teacher. After having done so, they were supposed to listen to themselves read, and listen carefully to how they read. They were supposed to note down errors or words that were difficult to pronounce and after having done so they were going to make a new audio file of them reading. Those who put in work and did their homework would benefit from it. Participant 3 believed this was a good way to also capture the competence of the pupils who were normally quiet in class. In such cases, homework could be emphasized when Participant 3 was in doubt between two grades to give a pupil.

Participant 3 said that she has used the whole grade scale when setting overall achievement grades, but it hurt to give the lowest grade. She has not given many of the highest grade 6, but some. However, the grade 1 has been given in both exams and for overall achievements. Participant 3 tried to avoid giving a pupil 1 for an overall achievement at all costs, but sometimes she had to. Participant 3 tried to give the low achieving pupils as many opportunities to show competence in the subject before setting this grade. She also thought it was difficult to know what lies within the highest grade 6 for an overall achievement grade.

She had an example of a situation where she was really unsure whether she had given the right grade for a girl. Participant 3 gave this girl a 5 for the overall achievement, and the girl asked her what she had to do in order to get the highest grade. Participant 3 said that she answered with a really general answer, because she did not know quite herself what was missing. When this girl had her oral exam in English, she received the highest grade and was so happy. Participant 3 felt unsure as to whether she had been too strict with the girl. A colleague of Participant 3 said that if she was unsure, then she had probably given her the right grade for the overall achievement. This happened several years ago, and Participant 3 still thought of this girl to this day.

Participant 3 did not believe that it was unattainable to achieve the highest grade. However, in written English more so than in oral, the highest grade was more difficult to achieve

according to Participant 3. She thought that this was unfair, because teachers assessed at a slightly higher level than they should, especially when assessing written exams.

Participant 3 felt that it was possible to get a fair assessment of the pupils and the grades they achieved were normally fair. On the other hand, there were many factors that should be considered when assessing. In English there was not a right or wrong answer such as in math. When sitting with colleagues and discussing papers, they often disagreed. As a teacher, she connected a bond with her pupils and cared for them. That may sometimes cause teachers looking for positive things, because we cheer them on. Participant 3 thought that she was more kind than strict, even though the pupils sometimes thought otherwise she said.

Participant 3 did not feel that the support she got in the process of grading was good enough. It could have been better with clearer guidelines to follow. She had good support from her colleagues, but other than that, she used the matrix from UDIR. Participant 3 always tried to write down good and concrete criteria for assessments. This would make it easier for her when assessing and grading the pupils. Moreover, one is quite alone as a teacher she said. Sometimes, teachers were lucky enough to be two teachers in the same class and could collaborate. Right now, they were two English teachers with their own class at the same grade. Participant 3 and her colleague made lists of the pupils they were in doubt of, which they discussed when grading.

When being asked of how the relationship between exam grades and overall achievement grades should be, Participant 3 thoroughly explained that those grades represented two completely different things. For the overall achievement, a teacher actually assessed a pupil's 10 years' worth of competence. This was not possible to assess at an exam according to Participant 3. There were many factors that came into play during an exam. Such factors were how the task was designed, how the pupils' day was that day and how the pupils' nerves came into play during an exam. However, in an ideal world, there should be no deviations between the two grades. They should especially match if the task for the exam was good and understandable. Participant 3 also explained that there was no strict syllabus to follow in the English subject. The theme given for an exam should therefore be constructed well enough so that every pupil managed to give a proper answer to the tasks. Accordingly, she thought this applied to genres as well. Participant 3 thought that it was strange that the written exam got to have an own grade on the pupils' diploma after completing lower secondary school, because it was only one day worth of work which was put into that grade. There were also

many competence aims that the pupils did not get the chance to show competence in for an exam. Participant 3 said that looking at exam grades had not led to any changes in her own grading practice. If there would have been constant deviations, then Participant 3 would have to change her way of assessing her pupils. Some deviations were normal. However, she explained that it was normal to achieve a higher grade at an oral exam than it was with a written exam. It has always been a day where the pupils got the chance to shine and Participant 3 looked back on oral exams as good experiences.

Both the oral and written English exam was not necessarily a quality assurance in Participant 3's eyes. She was of course thrilled if the grades harmonized. However, she changed her mind a little and said that it was in a way a quality assurance. What was devastating to Participant 3 was that there was such a big focus on the exams in the pupils' final year of lower secondary school. That one day of exam affected the teaching throughout the whole year. The pupils worked so hard to get to that one day of exams. Participant 3 did not see the point in itself of these exams. It was a stressful situation for the pupils, and it was stressful to hear about the exam for a whole year. However, she did care for these exam grades saying that she was also stressed if there were huge deviations. All in all, the exam grades were a measurement tool for teachers to get some clues of their own grading practice according to Participant 3.

Participant 3 was familiar with the UDIR's examination guide for pupils and teachers. She used a lot of time in classes with the pupils on this examination guide. Moreover, she felt that everything should be available for the pupils before an exam. They went through this examination guide in class as well as the technical report from the English subject. Graded exam papers were used as examples frequently by Participant 3 to show the pupils how one might answer the exam and what was required from them.

APPENDIX 6: Interview summary – Participant 4

Participant 4 was a 56-year-old female who has worked as a teacher for 16 years. Prior to this she has been an engineer working in the oil industry. Participant 4 believed that her background might make her different from other teachers in how they perceived the subject English. She noticed this when discussing with other teachers. They had different views of what was interesting to teach the pupils. Participant 4 believed to have set overall achievement grades for at least 10 of her 16 working years. She explained that she often works at the 10th grade and stay there because of her background. Participant 4 did not feel qualified to grade assessments right after her teacher education. She explained that she learned through experience. However, it was always scary to grade she said, and especially if she did not know the pupils well.

Participant 4 had participated in a course in grading in Stavanger. She thought the course was held in connection with a change in the curriculum and perhaps 10 years ago. Accordingly, she reckoned it was the municipalities who organized it. They were given exam papers to grade. After grading they would compare their own thoughts with the grades actually given at an exam. Participant 4 thought of this course as a very good experience. She explained that the teachers had good discussions, they even argued against each other on certain things.

Furthermore, Participant 4 had been an external examiner in oral English, but not in written English. Consequently, she has not participated in any examiner training. She explained that she had been an external examiner in oral English about 7 times. She tried to be an external examiner in oral English as often as she got the opportunity.

When being asked how Participant 4 worked with setting overall achievement grades in oral English, she explained that she first observed the pupils in class. She looked at who were verbally active in class and encouraged her pupils to be as verbally active as possible. Those who were verbally active towards her were easy to place she said. However, there were perhaps two-thirds of the class who were not verbally active in class. Those pupils were harder to place, because they were being labelled as if they did not manage the language. Participant 4 said that the verbally active pupils would have an advantage. For instance, a pupil who was verbally active and tried to communicate in English but should perhaps get the grade 2, would perhaps get a 3 overall because of him/her being verbally active.

Those who were silent in class were harder to place. Participant 4 explained that when doing oral activities such as speed dating, these pupils might become very silent when she was nearby. She had experienced that if she gave the pupils a task of making a podcast of a theme, she got a clearer picture of the pupils' levels. They talked more freely on tasks like this than in discussions. Participant 4 experienced that discussions as an assessment could have a negative effect for some pupils. It was rather difficult to give a grade in oral English for the pupils who were not engaged in oral activities. Moreover, she explained that this also applied later in their work life. If a person had a lot of good ideas but kept quiet, then he/she would have nothing to show for it.

When being asked how Participant 4 worked with setting overall achievement grades in written English, she claimed using all sorts of methods. Everything from questions to glossaries. Participant 4 also had some papers as assessments, which she disliked correcting. She gave her pupils writing tasks of a chosen theme sometimes to ease her own workload. The pupils would either write small texts or write texts together with peers. Another thing that she thought worked well were translations, which she also commented could be seen as old fashioned. Participant 4 saw the pupils language understanding right away when using translation tasks. They could write very bad sentences in English but if they were good at translations, then they were proficient in English according to Participant 4. Translations were also fairly easy to give feedback on as well. Participant 4 immediately saw whether her pupils followed the Norwegian structure or whether they were familiar with the language and wrote proper English. The pupils did not go in with high shoulders on such type of assignments as opposed to when writing papers. She often translated a piece of writing from their homework into Norwegian and gave her pupils the task of translating it back to English. After doing so they could compare their translation against the homework text.

Participant 4 has a strong belief in glossary tests. She believed it was important to practice new words to actually achieve a better vocabulary. Given the above, in written English she tried to document many assessments on every pupil to use as a basis for her overall achievement grades. If there was any doubt, she would let the pupils have a new assessment in some sort of way.

Participant 4 found it important to give her pupils feedback on oral assessments prior to giving them grades. By doing so she minimized the risk of surprise for the pupils. She usually asked them what they felt and whether the grade Participant 4 wanted to give them was fair or not. Furthermore, she also minimised the risks of complaints by doing so. However, she rarely adjusted the grade, but she got an idea of how her pupils thought.

Participant 4 thought of herself as a kind teacher when grading. She did not know for certain whether she was too kind, but she sometimes felt like she was. This was also reflected on her grading of the exam paper, seeing as she was one of the two teachers who gave the highest grade 4.

When being asked what aids Participant 4 used to support her grading of pupils, she answered that she has used the UDIR's examination assessments. She has also used examples of written exam papers to show an example of what was expected of the different levels of achievement. Participant 4 stuttered at the word aids, saying that she did not know of any aids that existed. She explained that general assessment tests could have been helpful, but she did not have any available at hand. She used presentations and tests/tasks designed by herself, because it was after all she who has to set their grades. However, the most important aid she used in the support of her grading were other teachers to discuss with when in doubt.

She felt that she got the support she needed in grading through her colleagues. Participant 4 explained that she was in a good team right now with two other teachers. Her colleagues may change from year to year, but she has always felt that she has had enough support through her colleagues. Nonetheless, she did think that there should be some sort of course regarding assessment once a year similar to the course she participated in several years ago. Especially for teachers working with pupils in the 10th grade, and the course should take place in March. She explained that it was important because a lot of changes happen, and the expectations of the pupils' achievement changed somewhat over time. However, a teacher often has several subjects he/she teaches. The teachers should probably know what subjects they are having pupils in either for a written or an oral exam before such a course. Teachers were very busy at the end of the spring semester in the 10th grade and not everyone would bother to go on an assessment course for all of their subjects if there was only one of them their pupils had an exam in.

Participant 4 was asked to what extent she used the national or a local curriculum as a benchmark when setting an overall achievement grade. She based her teaching on the textbook, which was national. Participant 4 also added a lot of information to the textbook. Furthermore, she explained that she looked into what she has taught the pupils when setting overall achievement grades. She did not believe she went beyond the national curriculum, but she did not follow all competence aims either. Participant 4 talked about the national curriculum as being open for interpretation and not firm and specific as to what the pupils should achieve and know. As a result, she often made her own arrangements when teaching about subjects she was interested in. Participant 4 believed she has a different approach than many other teachers who base all their teaching on the textbook. She explained that she has a Scottish husband and was also therefore internationally invested. This might affect her teaching style.

Participant 4 emphasized communication when determining a grade in oral English. It was important that the pupils could convey the essence of the topic they were talking about. Engagement was also something Participant 4 valued. Communication was also emphasized when determining a grade in written English. Lastly, It was important to be able to read the pupils' texts and be able to understand what they were trying to convey. The second factor were the pupils' language. Participant 4 explained that if a pupil managed to write a whole text with only writing small I's, then they were most certainly not able to achieve the highest grade 6, barely a 5. The reason for that was that the pupil has not bothered to read through his/her own text at the end. The pupil has not done his/her job thoroughly enough and they all have enough time in an assessment or an exam to check for communicative errors. Participant 4 also explained that there was a difference between making one error and having the same error throughout the text.

Participant 4 emphasized progression and effort the most if a pupil was in-between two grades for an overall achievement. She said that if a pupil put a lot of work into the subject, then he/she should be rewarded for it. They would normally also have a progression if they put a lot of effort into the subject, so these two factors are closely linked. Participant 4 checked if her pupils have done their homework which would count towards their effort in the subject. If she was in doubt whether to give a pupil a 3 or a 4, she often gave them a new opportunity to show their competence. Participant 4 explained that she was always open with her pupils and told them about the situation regarding this opportunity. It was important that the pupils knew

when they were being assessed and on what grounds. The pupils' competence was a given Participant 4 said, seeing as that was what she assessed when setting overall achievement grades.

Participant 4 had an example of a boy who did not like to be verbally active in class. His way of writing was characterized by a Norwegian structure. Nevertheless, It was a pupil who put in the effort and had the capacity to master the English language better based on his effort. However, language did not come easy for this pupil. Participant 4 wondered how she could guide this pupil to progress. She noticed that he talked better than he wrote. She therefore advised this pupil to write like he talked and to not overthink when writing. She believed and felt that this pupil has had a progression through this year. The pupil himself had said to Participant 4 that he thought it helped him a lot to watch British TV series with English subtitles instead of Norwegian subtitles. That way, he got the opportunity to listen and read in English at the same time. Participant 4 hoped that the pupil would have that progression and manage to achieve a 4 in written English by the end of the schoolyear.

Participant 4 was asked if she used the entire grade scale when setting overall achievement grades. She replied by saying that she did. She had set a lot of 6's, saying that the grade has been used more and more. Participant 4 has had pupils that should achieve the grade 8 if it existed. There were fewer pupils who received the lowest grade she pointed out. Accordingly, she rarely gave a pupil 1, because every pupil managed something in English. Participant 4 felt that the pupils nowadays were very good in English, a lot better than earlier years. She also explained that they have always been good in her teaching years because she started her teaching career so late. However, if she were to compare her pupils to her own generation, there was a huge gap in competence. She felt that the overall achievement grades were presented by a bell-shaped curve before. This curve was characterised by having many pupils receiving the grade 4, few pupils receiving 6 and few pupils receiving the grades 2 and 1. However, now there seemed to be a division in overall achievement grades. Participant 4 explained that this was not accurate at her school, but she brought up another lower secondary school as an example. At this school there was a division between those pupils who were high achievers and those who were low achievers. There was a huge gap between these pupils at that school. She did not think that the low achieving pupils were bad in English, but they did not put in any effort of trying.

Participant 4 believed that digital tests with digital grading would be easier for teachers and the pupils' grades would be more objective. Furthermore, she said that the pupils needed better teaching and it required more from teachers now than before to prepare challenging and befitting lessons for the pupils. Thus, the after-work of grading should not be as demanding. Teachers should emphasize having creative and good lessons and that required a lot since the pupils' level were so high according to Participant 4.

When talking about if it was possible to get a fair assessment of the pupils, Participant 4 said it was difficult with the pupils who were not verbally active. It was difficult because they might achieve a lower grade than they should because they were not able to show their competence as much in front of the teacher. Likewise, she also found it difficult to grade her pupils seeing as she often started working with them their last year of lower secondary school. Participant 4 did not know exactly what the pupils knew and felt, which made her wonder sometimes if she was demanding more from them than she should. She often assumed that the pupils had knowledge of something they should have learned, but that was not always the case. From experience she said that grading her pupils was a lot easier when being with them for two years instead of one.

Participant 4 personally found it soothing if her overall achievement grades corresponded with the exam grades being given. However, some deviations might occur. Participant 4 explained the reason why being that the task was not suited or interesting for the pupil. Another reason might be that the pupils were nervous when having an exam. She claimed this to be very difficult and would happily get rid of the whole exam scheme. Participant 4 talked of the exams as being nonsense. She felt that 90% of her pupils should get the same grade for an exam and an overall achievement statistically speaking. Nonetheless, she has experienced that there were small deviations between the grades she had set and the exam grades. She has had some pupils who have gone up a grade or two and some that have achieved a lower grade in an exam. Personally speaking, she did not like to have the exam as a measuring instrument of how she has set the overall achievement grades. Participant 4 felt nervous on both her pupils' behalf and her own before an exam. She has never experienced having a discussion of arguments for and against having an exam and would suggest this as being a fun exercise for teachers.

Participant 4 gave an example of a written exam two years ago which was very good and appealing for the pupils. The year before that was a difficult exam with a political theme. Participant 4 thought the exam seemed to change every year which made it difficult to use an exam as a quality assurance or a measuring instrument of how she had set her overall achievement grades. She did agree that an exam could be looked upon as a quality assurance but questioned why there was a need for it. The overall achievement grades have already been set when the pupils were having their exams. Therefore, she stated that an exam has never led to any changes in her own grading practice because she did not look into the pupils' exams, only the grades being given. Participant 4 explained that this was not something she got paid to do and was thus not bothered by doing so either. It would not mind Participant 4 if an exam replaced the overall achievement grade. Then the pupils would have been more focused in all subjects in case they had an exam in the subject.

Participant 4 did not feel familiar enough with UDIR's examination guide for pupils and teachers right now. She usually got familiar with the examination guide in April before an eventual exam. She did however have roughly knowledge of the content in the examination guide. Participant 4 used the examination guide in the classroom but not in the middle of a schoolyear. She went through the examination guide with her pupils in May as the exam approached, normally before the pupils' last and final writing day in English.

APPENDIX 7: Interview summary – Participant 5

Participant 5 was a 28-year-old female who has worked as a teacher for 3 years. She had 2 years of experience with setting overall achievement grades. She did not feel qualified to grade assessments, with her teacher education in mind. She had no experience with grading assessments from her teacher education and was never shown any examples of graded assessments in her 5 years long teacher program. However, she felt safe regarding the language after an education like hers. Participant 5 explained that there was a huge difference from what she learned at her teacher education and what a pupil at a lower secondary school was supposed to learn. She read multiple books in her teacher education that was of little use in her own teaching practice. Participant 5 had to ask colleagues as often as she could when she was in doubt and try to find out what the different grades required of competence. She often looked at exam answers at UDIR to see examples of the different grades.

Participant 5 had not participated in courses regarding assessment or grading. However, she had been to some seminars where they had discussed how to use digital aids when grading. Furthermore, they would look at assessing pupils, but the primary focus had never been on grading itself. Participant 5 had been an external examiner in oral English two times, but never in written English. She believed that one must apply to be an external examiner in written English and that some experience is required as a teacher to become an external examiner in written English. As a result, Participant 5 has not participated in any examiner training.

When being asked how Participant 5 worked with setting overall achievement grades in oral English, she answered by saying that the basis for her pupils' oral overall achievement grade lied in two assessments and if the pupils were verbally active in class. She sometimes said to her classes that it was extra important to read the homework a particular week because they were going to discuss it in class. The two assessments were usually a presentation and an oral conversation with the teacher in groups. The oral conversation was emphasized the most in the pupils' overall achievement grades. Participant 5 claimed to give the pupils good time to prepare for this kind of assessment. The pupils often received similar questions to practice on before the assessment. Vocabulary was also emphasized a lot she stated, and that the pupils were able to reflect and discuss a given topic. The pupils should be able to lead the conversation in such a way that Participant 5 did not need to follow up with questions all the

time. Participant 5 explained that it was important that the pupils were able to have a dialogue, which was also stated in the curriculum. Moreover, she looked at intonation but that was not as important as the other qualities. At last, Participant 5 mentioned that grammar was also important, she corrected a lot of was/were errors which was very important to control in order to communicate properly.

There were a lot of assessments in written English which could be seen as the basis for the pupils' overall achievement grade. Participant 5 usually started the year by having a short task either at school or as a written submission. Process writing has been a huge focus for Participant 5 the last year. To clarify, this was when the pupils started writing a text and got feedback on it before finishing the text and submit it to be graded. Participant 5's pupils always have a writing day before Christmas and before summer break.

Participant 5 looked at the written assignment as a whole first. She looked at the structure, whether their texts had a title and clear paragraphs. After doing so she started by skimming the text to see whether they have answered the text correctly. She explained that it was very difficult because she has experienced having pupils who write good texts but the texts they have written did not correspond with the task. Participant 5 thought she was different from others, because she did not punish the pupils by giving them a 3 instead of a 4 just for not understanding the task properly. Given the above, she explained that the language and communication should be in focus. However, it was of importance that the pupils understood a task as well. An example could be that Participant 5 gave a pupil 4- instead of a 4 if he/she has not answered a task properly. Participant 5 also talked about content as being important. She felt that the pupils' content was often very vague and superficial.

Furthermore, she was a strong believer in having a topic sentence for each paragraph and to elaborate on that sentence. After this, Participant 5 looked into grammar. She described herself as very picky when a pupil wrote a small "I". This showed that a pupil did not have any control of the language. Participant 5 then went on to look at punctuation. The last thing Participant 5 mentioned was vocabulary. She noticed very quickly if a pupil had an advanced or simple vocabulary. This was also something which affected the pupils' grade.

When setting an overall achievement grade, Participant 5 emphasized the last assessment they have had. However, she also said that the overall achievement grade should benefit the pupil.

An example was given where a pupil has received the grades 3+ and 4+. In this scenario she would give the pupil a 4 for an overall achievement. All in all, the grades should be as fair as possible according to Participant 5.

Participant 5 had through several other questions answered that she used group assignments, tests, presentations and home assignments. She did not use general assessment tests because she did not have enough bearing on them herself. Participant 5 did not use it since the pupils already have a national general assessment test to go through. UDIR's examination papers were used a lot by Participant 5. She often went through some exam papers with her pupils and she has noticed that UDIR was very thorough when it came to verbs such as "explain", "discuss" or "retell". It was very important to understand those verbs to be able to answer a written exam.

Participant 5 stated that she used the national curriculum a lot when setting overall achievement grades. She often used the characterization of goal achievement template from UDIR when grading. However, she often simplified the competence aims for both her and the pupils' benefit.

When being asked if there was something Participant 5 emphasized more or less if a pupil was in-between two grades, she replied by mentioning homework and progression. If a pupil did his/her homework and showed an effort, they showed that they really wanted to achieve something. She often told her pupils when there was an important homework which would be emphasized more than others. Moreover, progression was important, and a pupil should be rewarded for progressing for the better. Participant 5 had an example where a pupil worked really hard the last semester which gave count and the pupil actually progressed. Participant 5 gave that pupil the higher grade because of that. She explained that a teacher was supposed to assess a pupil's competence at the end of the schoolyear and not put too much focus on previous semesters. Furthermore, Participant 5 also thought that pupils should be rewarded for contributing in class when thinking of the oral overall achievement grade.

She stated to use the entire grade scale when setting overall achievement grades, at least up until now. However, there were some situations where she had not always given the lowest or the highest grade. Participant 5 did not believe that a 6 was unachievable, the pupils were only

in the 10th grade. The whole grade scale was made to be used from the lowest to the highest grade.

When being asked whether it was possible to get a fair assessment of the pupils, Participant 5 answered yes. She explained that some would perform better at presentations while others would perform better in a conversation. Since Participant 5 tried to have one assessment of both scenarios she felt that it was possible to give her pupils a fair assessment. In oral English, she stated that everyone should be able to stand in front of people and talk. However, there were some who struggled with presenting something in front of others. Participant 5 said that it was possible for them to hand in video presentations instead. Although, she tried to limit these options because it was very time consuming and the pupils needed the challenge. Given the above, Participant 5 explained that the pupils were sometimes affected by the situation around the assessment which might result in lower achievements.

Participant 5 stated that it was harder to get a fair assessment in written English. She asked the questions; when do we write a text in our work life where we do not get the opportunity to look over it? Are we actually preparing them for life after school? It was not normal to use five hours of a day to write a text where the pupils did not get any help and they did not have access to internet. This only happened in a school setting. Participant 5 therefore thought that schools were expecting too much of pupils. A pupil could have a good or a bad day, there were many factors to consider when grading written assignments according to Participant 5. She explained it as strange that the school system was not further along than this. However, Participant 5 never mentioned the teacher as a factor in this, but the situation and school system as a problem for getting a fair assessment of the pupils.

When being asked if Participant 5 cooperated with colleagues, she explained that she has very often been alone as an English teacher with the classes she had been working with.

Unfortunately, she has not been working in teams with others. She was very lonely as a newly graduated teacher the first two years. As a result, there were perhaps some pupils who had gotten a grade they should not have gotten she claimed. She received guidance once a week her first year as a new teacher. Participant 5 had the opportunity to ask questions and to talk about things she needed with a supervisor. In an ideal world she would have wished for an English teacher to cooperate with. She did not receive any help or support in grading other than from her colleagues.

Participant 5 explained that her pupils got a higher grade at their exam the two years she set overall achievement grades. She thought that there might be another reason than her grading being wrong or too strict. The reason being that she used a lot of time with the pupils who were going to have an exam in English. They all got individual counselling and they worked very hard with grammar the two days prior to the exam. Participant 5 believed that this made a difference. She looked over their exams and saw that her pupils were more conscious when writing. Furthermore, Participant 5 explained that this happened with pupils who especially were in-between the grades of 3 and 4. The pupils might also put in more effort in their exam answer just because it was an exam and the nerves that follows.

Participant 5 did not think that there should be considerable deviations. However, it was natural if there were some deviations. First of all because it was a very stressed situation. Secondly, because the exam tasks vary greatly from time to time. She remembered an exam from 3 years ago which was really “out there”. The preparation material was difficult, Participant 5 did not understand everything herself. However, the exam after that was very good. Participant 5 said that the help she provided her pupils with from earlier exams would not be the practice in the future. She explained that she did more than what was expected of a teacher to prepare her pupils before an exam. It will therefore be interesting for her to see in the future whether the work she did those two years had a considerable effect. Participant 5 believed that her pupils would have gotten lower grades if she had not helped them as much as she did when preparing for the exam.

Participant 5 liked the oral exam in English because it brought out the best in pupils. She did however not think it was strange if there were deviations between exam grades and overall achievement grades in written English. If there were deviations, Participant 5 did not think that it was the teachers’ fault. She said that those who made the exams did not always have a clear picture of the reality.

Before Participant 5 started to assess pupils and give them grades, she went through UDIR’s assessment examples. She remembered sitting with a colleague looking at how these examples were graded. She then found out that there were greater expectations of the pupils than she would have thought. Participant 5 explained that a pupil has to have a very good vocabulary and write very precise to achieve a 4. As a result, she explained that the exams had led to a different way of thinking but it has not led to any changes in her grading practice.

She felt that an exam was a quality assurance since she looked at those grades to see whether she was on the right track or not. This was something she automatically did, and something she believed all teachers automatically did. The question however was very difficult, and she did not know for certain what her position was in the matter.

Participant 5 was familiar with UDIR's examination guide for pupils and teachers. She explained that this was something she got alongside the exam. However, Participant 5 was not quite sure what the examination guide contained, seeing as she asked if the interviewer was referring to the competence aims. The interviewer then explained what the examination guide contained. After doing so, Participant 5 replied by saying she used this information and went through it in class. She also brought examples of exam answers in class. Participant 5 would not say that she actively used this examination guide in class, however she did show it to her pupils. Anyhow, she could probably have used it more than she does. Nevertheless, Participant 5 felt that she used a lot of time on the content from the examination guide in normal lessons at school as well, without using the actual examination guide.

APPENDIX 8: Interview summary – Participant 6

Participant 6 was a 61-year-old female who started working as a teacher in 1981 and has been an English teacher since 1995. This would imply that she has been working as a teacher for 39 years and as an English teacher for 25 years. She believed to have 10 years of experience with setting overall achievement grades in English.

When being asked whether she felt qualified to set grades after completing her teacher education she answered hesitantly no. She moved on to say that it was an enormous advantage to work in teams where they could be able to discuss assignments. There were many pupils who were in-between the grades of 4 to 5 and 3 to 4. Those pupils were very hard to decide on what to give them according to Participant 6. Having another teacher to help in those situations was very helpful and Participant 6 expressed that she has been very lucky with having colleagues around her to confide in. Furthermore, she stated that there were more competence aims and the curriculum was more detailed now than it used to be when she started out as a teacher. However, to be certain of a grade was difficult. Participant 6 asked a rhetorical question as to what we should emphasize when grading.

What Participant 6 has learned in the long run was that one should always try to read through an assignment without looking for flaws and try to understand the content of what a pupil has written, whether or not it communicates. This should get a separate grade, whereas grammar and language should be seen as a separate grade according to Participant 6.

When being asked whether she had participated in any courses/training in assessment/grading she could at least recall one time, if not maybe two. She explained that they looked at papers written by pupils which they discussed to see whether or not teachers think alike and had similar mindsets. What Participant 6 also experienced was that a whole grade consisting of 140 pupils had numbers instead of names on a written assignment. The teachers would then divide these 140 pupils between themselves, whereas they got 20 assignments each to grade. By doing so one would get an insight if there were any deviations from what a pupil normally achieved. Participant 6 described this as a method used to control if there were deviations. She explained that teachers could tend to get stuck on a mindset. An example given was if a pupil normally got the grade 4, the teacher was used to giving that pupil the same grade over and over again without really looking into what the pupil actually has achieved. Teachers

might get a revelation by performing anonymous grading. Participant 6 have participated in this method on two occasions. The researcher asked if there were any deviations when performing this grading system. Participant 6 said there were a few deviations, and these were really interesting. She also felt that it was reassuring to know that the teachers were mostly in agreement. What she has learned from this is that she might be too strict when grading.

However, Participant 6 explained that as a teacher she provided something pedagogical along the way in the 8th and 9th grade. That was to say, a pedagogical grade was often given on the basis of progression and whether a pupil had worked on the issues being presented to him/her when receiving feedback from the teacher. In the 10th grade on the other hand, a teacher must be more aware of the competence aims and focus on them primarily. She also explained that this was difficult. Some pupils were neither one grade nor the other. Participant 6 explained that in such situations, one should let the doubt benefit the pupil.

Participant 6 had not been an external examiner in either written or oral English and had therefore not experienced any examiner training. When being asked how Participant 6 worked with setting overall achievement grades in oral English, she mentioned several situations that were important. She tried to get everyone to talk English and to read English to bring out the pupils' efforts in class. She observed their activeness, language and the content the pupils managed to produce in class. Other than that, she supported her grades with presentations which have clear competence aims regarding language and reflection. These presentations were set with clear guidelines and criteria to see if they were at a high or low level.

Participant 6 divided presentations and conversations because there was not time to have a presentation and discussion at the same time. She explained that there was time for that when having an oral exam and when they were having oral mock exams. They normally had an oral mock exam at school C to prepare the pupils of what to expect when they were up against the real exam at the end of 10th grade. Participant 6 said that she has tried to have subject conversations as an assessment this year. This was not something she was used to having. This was a good opportunity for some to express themselves in English and to show their knowledge in a given subject. However, some pupils were overrun in the conversation by their peers. She explained that when it was done once, she learned a lot and the pupils who did not cope well with others got the chance to have an assessment such as this alone the next time or with other classmates. These conversations were usually in groups consisting of 3-4 pupils. Furthermore, the pupils she had now were very good at expressing themselves in

English which was a huge part of the curriculum. Participant 6 also added that there was a huge difference in how well the pupils talked English now than when she first started out as a teacher.

When working on setting overall achievement grades in written English, Participant 6 tried to constantly have some small assessments. These small assessments were divided into tasks. An example given by Participant 6 was when working with earlier exams, the pupils were sometimes given the task of either writing a short answer or a long answer. She also explained to her pupils that homework was an important aspect of their grade. In their homework they got a chance to show that they have understood a text and that they managed to articulate themselves by answering questions regarding a text they have read. Accordingly, when Participant 6 was in doubt whether a pupil should get a 3 or a 4 for an overall achievement, she used homework as an extra bonus of documentation to either tilt their grade up or down. The last big test at the end of the semester counted the most of their overall achievement grade according to Participant 6. However, Participant 6 would gladly look over the pupils' previous achievement in general if they were extremely unfortunate on this particular test. She explained that it was important to look closely at the pupils' last test when grading because the pupils changed a lot through that year. The boys especially become more mature and reflected throughout that last year in lower secondary school according to Participant 6. She also used the national curriculum when setting overall achievement grades.

When Participant 6 was asked what she emphasized when determining a grade in oral English, she answered by saying that a pupil must have good content, reflection and a good language in order to achieve a high grade. However, there was certainly room for incorrect grammar, a few stops and some occasions of stuttering according to Participant 6. The pupils did not need to be perfect when speaking. To communicate well was important, but at a high level the pupil must also have good content. Content, reflection and a slightly richer vocabulary than the pupils who achieved the grades 3 and 4 was necessary to reach a high level orally according to Participant 6. She recommended her pupils to find some keywords and synonyms to those keywords, which they could use in their oral evaluations. She did this so the pupils could get a richer vocabulary of words that conveys the same meaning. Participant 6 felt that many pupils have a very unformal but good language in English, because they watched a lot of English sitcoms. However, they lacked a formal language and a high vocabulary, because they used very simple words.

When being asked what Participant 6 emphasized when determining a grade in written English, she explained that she first tried to read the content and whether or not they had answered to what the task asks of them. Participant 6 set a character for herself in terms of the content, without looking directly at their language and grammar. She said that she sometimes set a separate grade for the content if there were huge deviations between the language and content. After doing so she looked into their language, and for the weaker pupils she also has to look at punctuation. Participant 6 explained that one might easily get caught on a pupils' punctuation mistakes if there were some, because it communicated poorly. In such cases, she focused on punctuation, initial letters, paragraphs, saying that they should have learned this prior to 10th grade.

What Participant 6 emphasized when determining a grade was therefore depending on what level the pupils were at. For low-achieving pupils, she aimed at them trying to convey meaning and being able to communicate. However, for high-achieving pupils she liked to use a division of content and language. There were pupils who have reading and writing difficulties which resulted in a lot of concord errors, but some of them could have a good language, good sentences and a nicely written text. Participant 6 tried to look past those concord errors such as is/are when reading through the text at first to understand the content. This was something she had learned at a course, because it was important to look at the pupils' communication skills first.

Participant 6 was asked whether she emphasized something more or less if a pupil was in-between two grades. She then answered by saying that the quality of a pupil's homework would be emphasized when in doubt. Progression could also be emphasized to some extent. However, attendance was not something Participant 6 could emphasize if her pupils showed effort in class. Attendance was not something one could grade according to Participant 6. Given the above, if a pupil showed good competence at reading his/her homework, it was very different from just showing up in class and achieve a better grade. Participant 6 had an example regarding progression where if she told a pupil to work on a clearer introduction, better structure with paragraphs and more reflection and the pupil managed to do so over time, that would be emphasized when determining the pupil's grade at the end of the 10th grade.

Participant 6 was asked whether she used the whole grade scale. She answered by saying that it happened. However, she felt that a pupil's achievement should be very low to achieve the lowest grade. Pupils often got an IV (not assessed) instead of the grade 1. On the contrary, Participant 6 said that she did not hold back regarding giving the highest grade if a pupil was at that level. She did not hold back on the half year evaluations before Christmas either, saying that some teachers did. She believed this to be the right way of setting grades.

Whether or not it was possible to get a fair assessment was a difficult question to answer. Participant 6 said that the criteria determine what she as a teacher should consider when assessing. She has experienced a pupil who struggled with writing long texts. That was not something he liked to do. This pupil was more of a mathematician. He had a very good language in English, but to write long texts and stories was not something he mastered. This particular pupil did not get to show his strengths which could be seen as unfair. Participant 6 felt that the grade pupils got was often not fair in terms of how they would work with the English language outside of school and after having completed school. It was often the kind, good girls who did their homework that did well in the subject and achieved the higher grades.

When talking about cooperation with colleagues on grading, Participant 6 explained that she was in a team with two other English teachers at the moment. Participant 6 also had a friend which she consulted, because she has learned the most from her. She used to work with that friend a long time ago. When Participant 6 was in doubt, she got her friend to read through the paper. Participant 6 was also asked whether the support at her school was good in terms of grading. Accordingly, she replied by saying that the limit was high before one dared to ask here in Norway, because we did not want to nag others. When in doubt, one should ask said Participant 6, and she got asked sometimes for help herself. The team she was working with now was very experienced, and Participant 6 thought it was a huge advantage when there were several English teachers working together at the same grade.

When moving on to talking about the relationship between overall achievement grade and exam grades, Participant 6 thought that the average should be higher for overall achievement than exam grades and that this situation would be perfectly normal. It was a very special situation with an exam, and the pupils could have a good or a bad day. There was more time to figure out the level of your pupils if a teacher had followed them for 3 years she said. If

they have one bad assessment at school, they got a new opportunity to show what they were capable of, because teachers were after all looking to see what the pupils knew. By doing so, it was natural that the overall achievement grades were higher than an exam according to Participant 6.

Participant 6 answered yes straight away with no hesitation as to whether exam grades were a quality assurance of how she and her colleagues had set the overall achievement grades. She thought that it was good to see where these exam papers end up on the grade scale. Because there were several teachers who had set this grade together and not only one teacher, Participant 6 explained that she got a clue in terms of her own thoughts. However, it had not led to any big changes in her own grading practice. She did include this in her teaching by preparing the pupils of what the criteria are. Participant 6 was familiar with UDIR's examination guide for pupils and teachers. She did not use this in the classroom, saying that some things went automatically when being a teacher for so long.

APPENDIX 9: Consent form

Do you want to participate in the research project:

“The relationship between overall achievement- and exam grades in the English subject in lower secondary schools of Norway”?

This is a question for you to participate in a research project where the purpose is to analyse the correlation between overall achievement- and exam grades in the English subject at lower secondary schools in Norway. This form will give you information about the aims and purpose for this research project is, and also what it means for you to participate in this research project.

Purpose

This research project is a part of a master’s thesis and will only be used for this purpose. The purpose of this master thesis is to analyse how teachers in lower secondary schools in Norway view assessment and grading practices. The hope is to interview English teachers about assessment and grading to hear their opinions and experiences around this.

Prior to the interview you will also be asked to correct an exam paper. I will provide an attachment of an exam paper from spring 2018, in addition to the exam tasks. I want you to rate this exam paper as well as giving a comment on what has been emphasized. The researcher would prefer if this could be clear prior to the interview. This research project is a part of an MA-thesis and will only be used for this purpose.

Who is responsible for the research project?

The University of Stavanger in Norway is responsible for this research project.

Why are you getting asked to participate?

You are being asked to participate since you are an English teacher at a lower secondary school in Norway. Contact information has been forwarded or suggested to me by the principals at your schools.

What does it mean for you to participate?

If you choose to participate in the project, it means that you are agreeing to grade and comment on an exam paper as well as participating in an interview that will last for approximately 30 minutes. The researcher will audio-record the interview and transcribe it. The transcription will be included in the research study.

Participation is voluntary

Your participation in this research project is voluntary. If you choose to participate, you may withdraw your consent at any time without giving any reason. All information about you will be anonymized. It will not have any negative consequences for you if you do not want to participate or later choose to withdraw.

Privacy – how we store and use your information

The information you give during the interview will only be used in this research project. Everything is confidential and anonymous. The information will be treated confidentially and in accordance with the privacy policy.

Only the researcher will have access to the audio-recording from the interview. After the audio-recording has been transcribed and anonymized, it will be deleted. You will not be able to be recognized in the publication of this MA-thesis. You will be anonymous.

What happens to the information after the research project is finished?

When the research project is finished (May 11th, 2020), the collected data will be used in research. The information will be presented as a summarized transcription in the research study.

Where can I find out more?

The assignment from the University of Stavanger has been approved by the NSD (Norwegian centre for research data). NSD states that the personal information given in the research project is in regulation with the Norwegian privacy laws.

If you have questions about the study, or wish to exercise your rights, please contact: University of Stavanger, Kjetil Vikhamar Thengs Email: kjetil.v.thengs@uis.no tlf:92846947 NSD - Norwegian centre for Research Data AS, by e-mail (personverntjenester@nsd.no) or phone: 55 58 21 17.

Best regard

Jane Hodne
(researcher)

Kjetil Vikhamar Thengs
(supervisor/associate professor of English linguistics)

Declaration of consent

I have received and understood the information about this research project: “English teachers’ grading and assessment practices in Norwegian Lower Secondary Schools”. I hereby consent to:

Participate in a semi-structured interview.

To give an objective assessment and grade of a written exam in English from the schoolyear 2018/2019.

I give my consent that the information I give in this research study can be used in this MA-thesis until the project is finished on May 11th, 2020.

Signature:

APPENDIX 10: Exam paper

1a

In this preparation material, there are many different things that people can have a sense of belonging to including language, religion, sport and others. Some of them are relevant to my sense of belonging, while some of them are not. The text "To belong or not to belong" is very relevant to my sense of belonging, because I do have a strong sense of belonging to my mother tongue. Even though I can speak many languages, but only my mother tongue gives me that strong feeling of belonging. It's the language that always reminds me about who I am and where I came from.

In the other way, the text "The beautiful game" is not relevant to my sense of belonging. Even though I'm not an very sporty person and I'm not a fan of sport, but I still don't think that sport gives you that strong sense of belonging as your society, country or language do. These things are constants, while can always change your sport over time. For example, you will not stay in your football team for your whole life and you will change it one day. That way you will lose that sense of belonging to your team one day.

1B

I think that the main message of this appendix is to show the benefits of living in a multicultural society, which means a diverse society with cultures from all over the world. It shows also how Britons are open for other people and how they aspect differences between people. I think that that these are the most important reasons for why Britain has become a multicultural society. I mean, 70% of Britons according to a research say that multiculturalism is a good thing, and this is a very big number. This is maybe as a result of colonial times, because then they had to travel to these countries where they had colonies, and so become more open for different people. As it seems that this is the only benefit of colonies. The new generations have also managed to be open for other people and remove the old racist thoughts.

To have a sense of belonging

A sense of belonging is a person's sense of affiliation and real attachment to something. It is a basic human need, and it's very important for your well-being just like food and shelter. People can have a sense of belonging to many things including their country, society, religion, language, a particular group, sport and among many others. People are different and every person has his own sense of belonging, and in this article, I will talk about two of them and explain why they are wrong by using to characters from the preparation material.

Most of people think that where you grow up is where you belong to, however, that is not truth. Being born in a particular place or being in a particular group doesn't necessary mean that you belong to them. Where you belong is where you find security, confidence and mutual respect. If you didn't find these, you will not have that sense of belonging.

A good example on this is Connor Clarke, the 18 years old boy, from the text "Island life on Outer Hebrides in words pictures". Conor has lived his whole life at the old-fashioned and conservative Isle of Lewis. He came as a gay for two years ago, and wants to leave the Island. Even though Conor is from this Island and he has spent his whole life here, but he doesn't feel that he belongs to this island. It could be because he didn't find the acceptance for being gay, since it's a conservative society where he lives and they don't accept homo sexuality. He have many dreams and the only thing he wants to do now is to leave the Island. That why we can say that it's very important to find the acceptance for who you are in where you grow up, to have the sense of belonging.

The second wrong thought among people I want to talk about is that a person can only have a sense of belonging between people that are similar to him. For example, people who belong to your culture, speak your language, or share your interests. However, that is also not completely true. It may be easier to have a sense of

belonging to people that are similar to you, but it's not possible to have a sense of belonging to totally different people.

A good example on this are the Canadian couple from the text "Britain's most diverse high street revealed and it's home to shopkeepers from 23 different countries". This street is very diverse and it's home to people from all parts of the world. In this street, all people help and collaborate with each other when they face troubles, including the Canadian couple who run a bookshop in this street.

Even though the people in this street are very different from the couple, but they still live together and have mutual respect for each other. The couples feel that they belong to this street and these people, even if they have different cultures, speak different languages and maybe don't share the interests at all. They don't care where people are from, what religion do they have, what colour, what race, and above all, they. Have a very strong sense of belonging to this people and this street.

Sources

To belong or not to belong:

https://pgsf.udir.no/Year2018/PreparationRoom4/eng_02.aspx

The beautiful game: https://pgsf.udir.no/Year2018/preparationRoom4/eng_03.aspx

Britain's most diverse high street revealed- and it's home to shopkeepers from 23 different countries: http://pgsf.udir.no/Year2018/preparationRoom4/eng_04.aspx

Island life on Outer Hebrides in words and pictures:

http://pgsf.udir.no/Year2018/preparationRoom4/eng_05.aspx