UiT The Arctic University of Norway

Faculty for humanities, social sciences, and teacher education

The Effect of Spell and Grammar Checkers on Pupils' English

A Study of Texts Written by Norwegian School Children Aged 13-14 Years

–

Nadia Sigvaldsen & Kathrine Kristiansen

Foreword

With this thesis, we end our five-year education with UiT- Norges Arktiske Universitet, thereby stepping into the roles as fully qualified teachers. As we submit our final work, we look back at five years filled with valuable memories with our peers. Thank you for the (very long) coffee- breaks, all the late-night social events, the trips to Juanitas and the light-hearted, sarcastic comments amidst the frustration.

To Kristin Killie, our supervisor. Thank you for the commitment to our work, thorough comments and for being present throughout this year. Your engagement to our project has elevated the task and motivated us throughout the entire process.

We would also like to thank Per Øystein Haavold for crucial guidance with statistics, and Vilde Jensen for taking the time to proofread the thesis last minute. Lastly, we would like to give our appreciation to the teachers and pupils who made our research possible to conduct.

Tromsø, May 2024
Nadia Sigvaldsen & Kathrine Kristiansen

Abstract

This thesis has investigated how the use of the spell and grammar checker in Microsoft Word affects spelling and grammar in English texts written by pupils aged 13-14 years in Norwegian schools. This topic was researched due to the increased use of digital aids in schools and the uncertainty regarding the effects of using of these tools. Assessment is a significant part of a teacher's professional responsibility in schools, and pupils often submit assignments digitally. Still, little research has been found on how spell and grammar checkers impact the evaluations pupils receive. Based on this, the study has investigated the extent to which we identify errors in texts written by a sample of pupils with and without the spell and grammar control program in Microsoft Word activated, as well as the extent to which these pupils claim to use, and understand how to use, this digital tool.

To research the impact of spell and grammar checkers on pupils' writing skills, quantitative data were collected by gathering 60 texts produced by 30 pupils under two separate conditions: one where they had deactivated the autocorrection function in Microsoft Word, and one where they had activated the feature. Additionally, pupils answered a questionnaire, providing insights into their knowledge and experience with the program. Various statistical tests were done to assess the significance of the study's results. To gain a more comprehensive understanding of the pupils' texts, the study includes a detailed analysis of eight chosen individual texts.

The findings of this study indicate that the spell check in Microsoft Word significantly contributes to reducing the number of spelling errors related to speech sound, incorrect capitalisation, and misuse of apostrophes among pupils. The study reviled that the grammar checker does **not** significantly help minimize pupils' verb-related errors, but instead proves effective in reducing other types of grammatical errors. However, while the use of the spell and grammar checker in Microsoft Word leads to a significant decrease in error rates, its effectiveness varies among individual pupils. Moreover, the findings of this study suggest that frequent use of spell and grammar checkers may have implications for summative and formative assessment, both for assessing each pupil's language skills fairly and for ensuring that assessment contributes to learning.

Sammendrag

Denne avhandlingen har undersøkt hvordan bruken av stave- og grammatikk-kontrollen i Microsoft Word påvirker rettskriving og grammatikk i tekster skrevet av elever i norsk skole i alderen 13-14 år. Årsaken til at dette ble forsket på, er den økte bruken av digitale verktøy i skolen og usikkerheten knyttet til effeken av å bruke disse. I skolen er vurdering en stor del av lærerens profesjonelle ansvar, og elever leverer ofte innleveringer digitalt. Det er likevel funnet lite forskning på hvordan stave- og grammatikk programmer påvirker vurderingen elevene får. Basert på dette har denne studien undersøkt i hvor stor grad vi identifiserer feil i et utvalg elevers tekster skrevet med og uten stave- og grammatikk programmet i Microsoft Word aktivert, samt i hvilken grad disse elevene hevder å bruke, og å forstå bruken av, dette digitale hjelpemidlet.

For å utforske påvirkningen av stave- og grammatikk-kontrollen på elevers skriveferdigheter ble det gjennomført en kvantitativ datainnsamling, hvor det ble samlet inn 60 tekster som er produsert av 30 elever under to separate betingelser: én hvor de ikke hadde tilgang til autokorreksjonsfunksjonen i Microsoft Word, og én hvor hadde slik tilgang. I tillegg har elevene svart på et spørreskjema, hvor vi stilte spørsmål som ga oss innsikt i deres kunnskap og erfaring med programmet. Ulike statistiske tester ble gjort for å teste om resultatene av studien er signifikante. For å få en mer omfattende innsikt i elevtekstene ble det også gjennomført en detaljert analyse av enkelttekster.

Resultatene fra denne studien indikerer at stavekontrollen i Microsoft Word bidrar betydelig til å redusere antall stavefeil relatert til ordlyd, feil bruk av store bokstaver, og feil bruk av apostrofer blant elever. I analysen av grammatikk-kontrollen avdekkes det at Microsoft Word **ikke** bidrar til å minimere elevenes verbrelaterte feil betydelig, men derimot viser programmet seg å være effektivt i å redusere andre typer grammatiske feil. Mens bruken av stave- og grammatikk-kontrollprogrammet i Microsoft Word fører til en betydelig reduksjon i antall feil, varierer effektiviteten blant individuelle elever. Videre viser funnene av denne studien at hyppig bruk av stave- og grammatikk- kontrollen kan ha implikasjoner for både summativ og formativ vurdering, både for å vurdere hver enkelt elevs språkferdigheter rettferdig og for å sikre at vurderingen bidrar til læring.

Table of Contents

List of Tables

List of Figures

1 Introduction

This master's thesis is a quantitative study where we have analysed the grammar and spelling in texts written by a sample of pupils in Norwegian schools, aged 13-14 years, with and without having the spell and grammar checker in Microsoft Word activated. The participants also answered a questionnaire which asked about their knowledge and experience with using the program. The study aims to reveal the extent to which the spell and grammar checker affects the spelling and grammar in pupils' written submissions. In this thesis, we will describe how the data was collected and analysed, followed by the results of the analysis. Finally, we will discuss the study's results within the framework of established theory and prior research done in this field.

1.1 Background and motivation for our research

During the course of our education at UiT, we have had a number of practice periods where we have visited multiple schools. We have also had experiences through working as substitute teachers, where we have encountered challenges relating to both assessment and the use of digital aids. Assessment practices vary amongst schools and teachers, and we observed that digital tools are used frequently in the classroom and that the pupils we met submitted the majority of assignments digitally. In light of this, we became interested in investigating how the spell and grammar checker in Microsoft Word influences the extent to which the pupils' submissions are suitable to use to assess language, based on the Norwegian curriculum for the English subject.

According to Utdanningsdirektoratet (2019, p. 2), English is an important subject for cultural understanding, communication, all-around education, and identity development. Furthermore, the subject shall equip students with English language proficiency in reading, writing, and oral communication, which is essential for their educational, social, and working lives. This proficiency stands as a core value in English education. Utdanningsdirektoratet (2019, p. 2) states that pupils are expected to be able to use English spell and grammar rules in their language production and to develop explicit knowledge of English grammar and other parts of the language system. During all our practice periods, we observed that none of the schools used handwritten submissions. All subject-related submissions were submitted through Microsoft Word, with access to spell and grammar checkers. This made us reflect on the competence aims for year 10 concerning linguistic and digital competence and skills. We reflected on whether the aims relating to spelling and grammar conflict with those relating to

digital tools. In other words, we questioned whether the use of spell and grammar checkers enables pupils to learn English or if it prevents them from doing this.

The competence aims in the curriculum for the English subject address, amongst others, linguistic competence and digital competence. We will discuss the aims after year 10, as they build on the aims after year seven, and they are most relevant for the age group we wanted to conduct our research on in this study. After year 10, the pupils should be able to "follow rules for spelling, word inflection, syntax, and text structure" and "use knowledge of word classes and syntax in working on one's own oral and written texts" (Utdanningsdirektoratet, 2019). Furthermore, it is stated that they should be able to "use different digital resources and other aids in language learning, text creation and interaction" (Utdanningsdirektoratet, 2019, p. 8). These aims are separate, and we question whether the teachers can identify the pupils' learning outcomes relating to linguistic competence simultaneously as they assess their competence in using digital resources in language learning. We therefore decided that the focus would be on the pupils' submissions rather than the teachers' assessment practice to identify whether these submissions are representative of their grammatical and spelling skills.

In addition to the competence aims, the English subject curriculum has three core elements that should be integrated with the subject (Utdanningsdirektoratet, 2019). These are communication, language learning and working with texts in English. Although all core elements can be related to text production, our research is based on the effects of spell and grammar checkers when assessing the pupils. Our study will focus on language learning, which is one of the core elements. In relation to language learning, it is described that "learning vocabulary, word structure, syntax and text composition gives the pupils choices and possibilities in their communication and interaction" (Utdanningsdirektoratet, 2019). Although the world is becoming more digitalised, giving pupils possibilities and choices in their communication might mean that they learn the language well enough to not rely on digital resources whenever they communicate, both written and orally. Moreover, when addressing writing as a basic skill, Utdanningsdirektoratet (2019) states that writing requires formulating texts that communicate. One can argue that if pupils are dependent on using the spell and grammar checker to communicate, their communication and interaction could be limited when they do not have access to this tool.

When deciding on what to research we wanted to ensure that our work would be useful for teachers of English as a second language. Since assessment is a significant part of teachers'

work, we wanted to contribute to more insight into assessing for learning. As digital tools are used more often in classrooms, and little research has been found on the effects of using these tools, we wanted to contribute to filling in a knowledge gap in this field. In addition to specifying the three core elements of English, the curriculum also specifies four basic skills pupils should acquire. These are reading skills, writing skills, oral skills and digital skills (Utdanningsdirektoratet, 2019). According to Utdanningsdirektoratet "digital skills in English involve being able to use digital media and resources to strengthen language learning" (2019). Therefore, one of our main motivations for conducting this research was to explore whether spell and grammar checkers can contribute to language learning and be used as a part of formative assessment, which we will explain further in the next paragraph.

According to Kunnskapsdepartementet (2019), formative assessment is the process of gathering and using evidence from pupils' learning to improve instruction. It occurs throughout a period, and it is formative in the sense that it is ongoing and provides information that is used to improve teaching and learning. As noted by Munden et al. (2017, p. 122), language teachers need to identify what their pupils know in order to help them learn more. This could mean that if the teacher does not have an authentic representation of the pupil's grammatical knowledge and skills, the formative assessment might not be valuable for the pupil's learning. According to Kunnskapsdepartementet (2019), teachers need to use formative assessment to promote learning and develop competence in the subject. This allows the teacher and the pupil to monitor progress towards achieving learning objectives that can be approached in various ways. In order to create a learning environment for the pupil, the teacher shall facilitate pupil participation and stimulate the desire to learn by using a variety of strategies and learning resources to develop competence in understanding and being understood in English, both orally and in writing (Kunnskapsdepartementet, 2019). We question whether teachers get access to the information needed to give formative feedback when the pupils use spell and grammar checkers. Our research aims to get insight into how the pupils' grammatical and spelling skills are represented when they use spell and grammar checkers and whether their learning needs can be identified when using this feature.

1.2 Research questions

The aim of this study is to reveal the extent to which spell and grammar checkers affect pupils' spelling and grammar in written submissions, thus gaining insight into how representative their digital submissions are of their language skills. To operationalise our

vision, we have made a main research question and two subquestions that will be addressed in this study. These are the following:

To what extent, and how, does the use of the spell and grammar checker in Microsoft Word affect spelling and grammar in texts written by pupils in Norwegian schools aged 13-14 years?

RQ1) To what extent do we identify differences in error rates in texts written by a sample of pupils with and without the spell and grammar checker activated?

RQ2) To what extent do the participants in our study claim to use, and understand how to use, the spell and grammar checking software in Microsoft Word?

The main research question and research question one (RQ1) will be answered through analysing the spelling and grammar in texts written by pupils both with and without the spell and grammar checker activated. The goal of including research question two (RQ2) is to gain insight into the pupils' knowledge and experience with using grammar control programs, as we identified this as a possible variable. This question will be answered using the data from the questionnaire. To answer the main research question, RQ1 and RQ2 we will do a variety of statistical tests, in addition to a qualitative analysis of selected pupils' submissions.

1.3 The outline of the thesis

In this thesis, our research is delineated across seven chapters. Chapter 2 addresses relevant theory and prior research done in this field. In chapter 3, we describe the method we used for collecting data and for analysing our dataset. In chapter 4, we present the results of the study, which we discuss further in connection with theory and relevant research in chapter 5. In chapter 6, a summary will be presented where the main findings of the study are highlighted. Finally, reflections regarding the limitations of our study and suggestions for further research are included as our final remarks in chapter 7.

2 Theory and relevant research

In this chapter, we describe theory and prior research which we have found relevant in our study. As we found a limited amount of research done with our research design and the effects of spell and grammar checkers, we have merged theory and prior research into one chapter.

2.1 Explicit to implicit knowledge conversion

According to Flognfeldt and Lund (2016, p. 76), explicit language learning refers to the conscious and deliberate acquisition of language through formal instruction, while implicit language learning involves acquiring language subconsciously through exposure to it in a natural context (Flognfeldt & Lund, 2016, p. 35). Explicit language learning may be more effective in mastering grammar rules and syntax, while implicit language learning may be more effective in developing fluency and natural communication skills. Ultimately, a combination of both approaches can lead to a well-rounded language competence (Flognfeldt & Lund, 2016, p. 76).

According to DeKeyser (1998), explicit knowledge becomes implicit when the learner's communicative practices are frequent. Additionally, Ellis (2006, p. 97) states that the conversion from implicit knowledge to explicit knowledge is possible when the learner is ready to acquire the explicit knowledge themselves. In grammar learning, this could mean that the knowledge of grammatical structures can become a part of the learner's communicative repertoire when the learner knows the differences between the implicit input and their own output. One can, however, argue that learners often need help from a teacher with identifying these differences, making it essential to map their knowledge throughout their learning.

For the pupils to develop grammatical competence, they need both knowledge about grammatical structures and skills to use this knowledge in their own production. By internalising, confirming, and consolidating a rule, the rule will, according to Fenner and Skulstad (2020, p. 225), become anchored into the pupils' long-term memory. This may create declarative knowledge, developing an explicit knowledge about grammar. However, as a competence also requires skill, fostering what Fenner and Skulstad (2020, p. 225) refers to as procedural knowledge should be beneficial. The grammatical rules would then become automatised so that the pupils can use their grammatical knowledge to produce their own

original work. Activities that can foster procedural knowledge must be more open-ended, requiring pupils to rely on their declarative knowledge about grammar to produce language. When accessing a spell and grammar checker, the pupils no longer rely on their declarative knowledge alone, and the pupils are not as dependent on their knowledge about grammatical structures when they produce language. This could help prevent the process of making the grammatical rules automatised. However, receiving guidance and assistance in their learning has been argued to help pupils in their language learning. This will be addressed next.

2.2 Socio-cultural theory

According to Imsen (2014, p. 183), learning can be viewed as an individual phenomenon happening inside the learner's head or as a social process, as learning does not happen without interaction. However, one of the leading theoretical perspectives is the socio-cultural theory, which views learning as a result of social interaction (Imsen, 2014, p. 195). According to Vygotskiĭ and Cole (1978, p. 85), the concept of a zone of proximal development is exceptionally important in school learning. This concept is based on there being an actual developmental level, which can be described as "the level of development of a child's mental functions that has been established as a result of already completed developmental cycles" (Vygotskiĭ & Cole, 1978, p. 85). In other words, the actual developmental level refers to what the child can do independently, not based on their age but on their demonstrated skill level. According to Vygotsky, the zone of proximal development is "the distance between the actual development level as determined by individual problem solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" (Vygotskiĭ & Cole, 1978, p. 86). This means that if a child is in their zone of proximal development in a certain matter, they cannot yet master the task they have been given independently. They can, however, master it with assistance or guidance.

Scaffolding theory was introduced based on the hypothesis that the zone of proximal development is part of the learning process. According to Bakhurst and Shanker (2001, p. 50), scaffolding theory in language acquisition means that the child receives teacher feedback, which is monitored and adjusted as the child begins to master the language. Scaffolding in the context of learning can, according to Bakhurst and Shanker (2001, p. 50), be compared to scaffolding in the literal sense. There must be a building at the core for the layers of scaffolding to have any purpose. In a learning context, this involves the development being ongoing and involving the interaction between environmental and biological factors. It has

been widely assumed that corrective feedback is one important way for the teacher to provide scaffolding, as explained in the next section.

2.3 Written corrective feedback

Sheen and Ellis (2011, p. 593) describe corrective feedback as the feedback the language learner receives when they have made a linguistic error in their second language. They differentiate between oral and written corrective feedback, but as we look into pupils' written work, we will not discuss the feedback they receive on their oral production. In their description of written corrective feedback, Sheen and Ellis (2011, p. 593) also make a distinction between direct and indirect corrections. Direct corrections involve input-prompting feedback, meaning that the teacher or whoever is correcting the learner gives the learner the correct grammatical form or sentence structure, with or without an explanation (Sheen & Ellis, 2011, pp. 593-594). Indirect corrective feedback involves output-prompting feedback through, for example, error codes or indications, thereby prompting the learner to reflect on grammatical issues and make any necessary changes (Sheen & Ellis, 2011, pp. 593-594).

Corrective feedback can also be viewed in a sociocultural context (Sheen & Ellis, 2011, p. 597). As mentioned above, Sheen and Ellis (2011, p. 597) state that learners are in the zone of proximal development when they can produce linguistic features through scaffolding in interactions with others but cannot do so independently. When moving from the zone of proximal development to self-regulation, learners can use these features correctly without guidance. Sheen and Ellis (2011, p. 597) argue that with a sociocultural view of corrective feedback, one has to accept that the corrections must be based on the individual learner's needs, as it must challenge them sufficiently.

2.3.1 Effects of written corrective feedback

According to Sheen and Ellis (2011, p. 605), many studies have researched the effects of written corrective feedback, most of them focusing on the effect of the different types of written corrective feedback. This research was spurred by an article published by Truscott in 1996, where he argues that grammar correction should be excluded from all writing courses and that it is both ineffective and can have harmful effects (Truscott, 1996, p. 328). He argues that grammar correction does not encourage pupils to enjoy their learning, as people do not like to be told that they are wrong (Truscott, 1996, p. 354). Moreover, he states that pupils simplify their writing to avoid corrections, causing them to learn less (Truscott, 1996, p. 355). Truscott's controversial article spurred a long debate about the effects of corrective feedback

on grammatical accuracy, and many researchers have argued against Truscott's claims (for an overview, see Hyland & Hyland, 2006). According to Ellis (2009, p. 16), corrective feedback is effective, but the nature of the feedback is dependent on the learner for it to be useful. Direct feedback is, for instance, effective for beginners, whilst advanced learners would benefit more from indirect feedback (cf. also Spada, 2011, p. 233 and Lira-Gonzales & Nassaji, 2020). In the next paragraph we will address the study done by Ghandi and Maghsoudi (2014), as the study was done on pupils with English as their second language (L2) and it researched indirect and direct corrective feedback. Additionally, we address the research done by Bitchener (2008), as the study has a high number of participants and researches a variety of approaches to corrective feedback.

Ghandi and Maghsoudi (2014, p. 53) compared the effects of indirect and direct corrective feedback for improving learners' spelling using an experimental design with two groups receiving different types of feedback. 56 girls from a high school in Iran participated in the study. There is no information about the participants' first language, but as the research was done at a school in Iran, one can imagine that English was their second language and that most participants had, for instance, Persian or Arabic as their first language. One group received direct feedback, whilst the other received indirect feedback. Ghandi and Maghsoudi (2014, p. 53) did pre and post-tests to gather their data. In the pre-test, they found that both groups' mean scores were approximately equal when testing spelling. In contrast, the group who received indirect feedback had a higher mean score in the post-test than those who received direct feedback. Ghandi and Maghsoudi (2014, p. 59) therefore argue that indirect corrective feedback is most effective in correcting spelling errors and emphasise the importance of students' contribution in this process.

A study done by Bitchener (2008, p. 109) in two schools in Auckland, New Zealand, was designed to test whether the accuracy of the use of two functions of the English article system varied over time, and if so, if this was a result of various corrective feedback techniques. The functions chosen were the referential definite article "a" in first mentions and the referential definite article "the" in subsequent mentions. The participants' knowledge about the articles was measured through pre-tests and post-tests, as well as a delayed post-test after two months. The participants were divided into four groups consisting of 17-20 participants in each group. Group one received direct corrective feedback and both written and oral meta-linguistic explanations, group two received direct corrective feedback and written meta-linguistic explanations, group three only received direct corrective feedback and no meta-linguistic

explanations, and group four was the control group. None of the 75 participants had English as their first language, but 88% of the participants had received classroom instruction in English for an average of eight years. Bitchener (2008, p. 113) found a significant improvement in the accuracy in the post-test, which was retained in the delayed post-test after two months. Group one, who received direct corrective feedback and both written and oral meta-linguistic explanations, had an accuracy of 67.59% in the pre-test and 82.65% in the delayed post-test. Group two, who received direct corrective feedback and written meta-linguistic explanations, had an accuracy of 53.11% in the pre-test and 76.78% in the delayed post-test. Group tree, who only received direct corrective feedback and no meta-linguistic feedback, had an accuracy of 59.35% in the pre-test, and an accuracy of 80.05% on the delayed post-test. The control group, by contrast, had an accuracy of 51.19% on the pre-test and 63.90% on the delayed post- test. Bitchener (2008, p. 116) therefore concluded that the study demonstrated that the significant improvements in accuracy can result from the written corrective feedback of errors. In his discussion, Bitchener (2008, p. 114) states that the results of the delayed post-test may not necessarily be solely based on the treatment provided between the pre-test and the immediate post-test, as variables such as additional instruction and self-study were not controlled for. An important question to ask is whether spell and grammar checkers may replace the teacher as the provider of scaffolding. We address this issue in the next section.

2.4 Research on the effect of spell and grammar checkers

In this subchapter, we present research relating to the effect of spell and grammar checkers. First, we present studies done on the usefulness of the feedback of grammar checkers. Next, we address research on the effect of using spell and grammar checkers on pupils´ spelling and grammar. Lastly, we present research on handwritten versus typewritten texts, done on Norwegian learners of English.

2.4.1 Research on the usefulness of the feedback of grammar checkers

A study conducted by Hernández Puertas (2018) researched what type of feedback, from the spell and grammar checker or from a teacher, would be most effective in targeting English articles, and what type would have a lasting effect. The research was done on 27 adults with Spanish and/or Catalan as their first language. All participants reported that they had studied English for a minimum of six years (Hernández Puertas, 2018, p. 75). The participants were divided into three groups, where one group received teacher's feedback, one received

computer feedback and the last group was a control group who did not get any feedback on articles. The data were gathered through a tailor-made test, after having an eight-week process where the participants produced a number of texts that they received feedback on. Hernández Puertas (2018, p. 79) found that the teacher's feedback had a positive effect, as the learners' errors on articles were reduced. By contrast, the group receiving computer feedback and the control group did not show any significant reduction in errors on articles. Hernández Puertas (2018, p. 82) concluded that grammar-checkers can be used as a supplement when teaching writing rather than a replacement, as the results demonstrated that corrective feedback given by a teacher reduces the number of errors short-term more than feedback given by a computer.

In a study done at an Australian college, Cavaleri and Dianati (2016) gathered quantitative data through a survey. A total of 18 students participated in the survey that aimed to gather data about Grammarly's perceived usefulness and user-friendliness. Eight of the participants reported to have English as their L2. Although this study is on Grammarly, Cavaleri and Dianati (2016, p. 225) describe that if a grammar checker finds an error, it explains the grammar rule and also offers a solution. This is not an element found exclusively in Grammarly, but in most grammar checkers. As addressed in chapter 2.3, direct feedback is when the correct grammatical form or sentence structure is provided, with or without an explanation (Sheen & Ellis, 2011, pp. 593-594). One can therefore argue that the feedback provided by grammar-checkers is direct. In the study referred to by Cavaleri and Dianati (2016, p. 232), 22.2% of the students claimed that the feedback in Grammarly is not always helpful. Whether the feedback is perceived as not useful due to its direct nature or if there are other explanations is not accounted for. In contrast, 77.8% of students felt that Grammarly positively influenced their writing confidence, which can suggest that using it may have long-term benefits for the students (Cavaleri & Dianati, 2016, p. 233). However, Cavaleri and Dianati (2016, p. 233) also argue that the usefulness of Grammarly may be limited, as some of the students reported that the feedback they received was flawed or hard to understand. For instance, one of the participants claimed that "[it's] okay, they suggested something that made no sense" (Cavaleri & Dianati, 2016, p. 232), and another participant responded, "I found Grammarly is not very reliable, it gives feedback that doesn't make any sense and I become very confused after using it to review my work" (Cavaleri & Dianati, 2016, p. 232). The number of students who reported this is, however, not specified.

2.4.2 Research on the effect of spell checkers

While there appears to be little research on the effect of spell checkers on spelling competence, an empirical study conducted by Hewa Fouad et al. (2022) researched the effects of spell-checking features on English language learners. The study was conducted at Tishk International University in Iraq. The exact number of participants in the study is not mentioned. However, it is stated that the research is done on two groups from a second-year study in the Education Faculty, English Language Teaching department (Hewa Fouad et al., 2022, p. 106). The participants completed pre and post-tests developed by the researchers. Each student was asked to write one paragraph with a word count of 100 words about "someone they admired" for both the pre-test and the post-test. The same topic being chosen for both tests is justified by the researchers wanting the participants to write similar texts using similar vocabulary (Hewa Fouad et al., 2022, p. 106). The participants were divided into two groups, with one being the experimental group and one being the control group. Both groups were given assignments to complete between the pre-test and the post-test, where the control group were instructed to complete these in handwriting, whilst the experimental group used spell-checking software. Further descriptions of the assignments are not found in the article.

Hewa Fouad et al. (2022) counted the numbers of spelling errors made by the two groups in the pre-test and in the post-test. They then calculated the mean difference between the results of the pre-test and the post-test (Hewa Fouad et al., 2022, p. 108). The experimental group had a mean difference of -1.250 from the pre-test to the post test, meaning that they in average made 1.250 fewer spelling mistakes in the post-test compared to the pre-test. For the control group, the mean difference was -3.00. Hewa Fouad et al. (2022, p. 107) calculated the p-value of the research and found that p=0.0151 and p=0.0002 for the experimental group and the control group, respectively. As the p-value was less than 0.5, both groups improved their spelling, indicating a significant difference after the imposed intervention. It is unclear why the control group also improved their spelling in the post test. It is possible that they focused more, understanding that this test was somehow more important than the first one, or that they became more aware of their spelling when they completed the assignments using spell checking software. However, it is also possible that a sample of 100 words was not enough to map the spelling competence of these learners.

2.4.3 Research on the effect of grammar checkers

John and Woll (2020, p. 169) researched the effectiveness of two online grammar-checkers, Grammarly and Virtual Writing Tutor, as well as the grammar checker in Microsoft Word. According to John and Woll (2020, p. 170), written feedback is time-consuming, especially with multiple drafts. Therefore, using a grammar checker can be appealing to teachers as it helps reduce their workload. John and Woll (2020) tested the three programs on a wide range of grammatical errors collected from ESL compositions and on sentences created by the researchers. The main objective of this study was to assess the appropriateness of using automatic corrective feedback generated by grammar checkers for second language learning.

For this study, John and Woll (2020, p. 185) had 50 handwritten essays produced by 28 adults between the ages of 21 and 36. The participants were advanced learners in their second year of a four-year TESL program at a university in Canada. The students wrote two essays, each in a total of three hours. The texts were a response to questions based on novels the students had read during the course. During these hours, the participants had notes, dictionaries and grammar references available. The researchers also had 129 sentences that were not authentic samples from second language writers, which were developed and generated by John and Woll's prior and personal experiences collected over several years. These sentences were based on exams and other assignments the researchers had read throughout the years, and they were designed to replicate errors that French learners of English typically make (John & Woll, 2020, p. 185).

In their analysis, John and Woll (2020, p. 185) excluded spelling, punctuation and vocabulary errors to maintain a grammatical focus. In total, the 50 compositions from the students' essays consisted of 23 108 words. John and Woll (2020, p. 185) detected a total of 358 errors in the different categories among these 50 texts. This equates to an average of one error for every 65th word (John & Woll, 2020, p. 176). Errors related to the verb phrase were by far the most frequent in this material, with 164 tokens, i.e. 43.02% of the tokens. Out of the three grammar checking tools tested, Grammarly identified the highest number of errors in the compositions created by the students, with 106 errors detected. Virtual Writing Tutor managed to detect 71 errors in the compositions, and Microsoft Word only managed to detect 18 of the errors. In the number of errors detected in the 129 sentences made by John and Woll, which also had 129 grammatical errors, Grammarly also outperformed the other programs. Grammarly identified 63 errors, Virtual Writing Tutor identified 51 errors, and Microsoft Word identified

19 errors. These numbers show that Grammarly was better at detecting errors than Virtual Writing Tutor, and both these programs outperformed Microsoft Word. As for the types of errors spotted, Microsoft Word failed to detect seven error types, viz. errors involving subject-verb agreement, tense shift, possessive form of nouns, pronouns, missing prepositions, unnecessary prepositions, and determiners (John & Woll, 2020, p. 179 and 180, tables 4 and 5). By contrast, Virtual Writing Tutor failed to detect three error types, viz. errors to do with tense shift, pronouns, and relative clauses. As for Grammarly, the only error type it was unable to detect was tense shifts. It appears that tense shifts are particularly difficult for grammar checkers to detect, suggesting that problems with tense shifting can only be improved upon through the help from a teacher. Based on the results, John and Woll (2020, p. 186) argue that Grammarly can be considered the best grammar checker tool, as it detects more mistakes.

2.4.4 Errors in handwritten versus typewritten texts

There has been done little research on the effects of using spell and grammar checkers on Norwegian learners of English. As the learners' L1 can affect the acquisition of their L2, we wanted to include research done on pupils with Norwegian as their L1. To our knowledge, the only studies researching the effects of spell and grammar checkers on pupils with Norwegian as their L1, are master's theses. In this subchapter we will address the studies by Rynning-Tønnesen (2023, p. 2) and Wedge (2021), which compare handwritten and typewritten texts.

Rynning-Tønnesen (2023, p. 2) compared students' grammatical abilities when writing by hand versus using grammar-checking software in digital writing tools. In contrast to our study, Rynning-Tønnesen (2023, p. 2) focused on handwriting versus typewriting, and the research was described as a case study with participants from one class. However, the data collection had similarities to our data collection process as both studies designed tasks to give the pupils, and both collected samples from the same participants twice. We cannot find a description of the timeframe given to the participants when writing, or a description of the teacher's, the researcher's and the pupils' roles during the data collection process. Therefore, we cannot identify all similarities and differences between our research design and Rynning-Tønnesen's design.

Rynning-Tønnesen (2023) did a quantitative content analysis of the data using a meticulously designed coding manual. The study was conducted on 26 students in the English subject in 8th grade. The coding manual consisted of four grammatical and five spelling categories. The

grammatical categories were subject-verbal agreement, pronoun agreement, article usage, and irregular past tense. The spelling categories included homophones, letter omissions, letter additions and capitalisation (Rynning-Tønnesen, 2023, p. 2). Additionally, as part of the spelling category, one of the sub-categories is referred to as misspelled words, which is described as "[a]n error in which a word is spelled incorrectly (e.g., recieve vs. receive)" (Rynning-Tønnesen, 2023, p. 26).

Rynning-Tønnesen found 176 grammatical errors and 302 spelling errors in the handwritten texts (Rynning-Tønnesen (2023, p. 64). The typewritten texts had 108 grammatical errors and 113 spelling errors. The table below gives an overview of the results, with *M* representing the male participants and *F* representing the females. The gender- difference in error rates is, however, not relevant to our study.

Total errors detected	Handwritten		Typewritten	
Grammatical errors: 284	(M) 102	(F) 74	(M) 54	(F) 54
Spelling errors: 415	(M) 155	(F) 147	(M) 44	(F) 69

(Rynning-Tønnesen, 2023, p. 39)

Rynning-Tønnesen (2023, p. 79) found a significant difference between the occurrence of grammatical errors in hand-written versus typewritten texts. The category with the biggest difference between handwritten and typewritten texts was article usage errors. These errors occurred 1.52 times per hundred words in the handwritten texts, and 0.32 times per hundred words in the typewritten texts. Rynning-Tønnesen (2023, p. 79) found misspelled words to be frequent in both modes of writing, and the results displayed a bigger difference in errors per hundred words in this category than in the grammatical category. Here, the students made an average of 4.29 errors per hundred words in the handwritten texts, and 0.80 errors per hundred words in the type-written texts. According to Rynning-Tønnesen (2023, p. 79) almost 50% of the total number of errors identified in the spelling category in the handwritten texts, were made by six participants, with a total of 135 errors. In the typewritten texts only 26 spelling errors were counted in the same six students' submissions, reducing the number of errors by 80.74%.

When researching what types of spelling errors Norwegian learners of English do, Wedge (2021, p. 2) questions how the use of computers affect Norwegian pupils' English spelling. Wedge (2021, pp. 30-31) conducted corpus research using handwritten texts by pupils in the 9th and 10th grade from The Corpus of Young Learner Language (CORYL), and a control

corpus comprising English texts written by pupils in 10th and 11th grade. The number of pupils who contributed to the study is not mentioned. The CORYL corpus contains a total of 65 079 handwritten words and the pupils in the control corpus produced a total of 66 079 words in Microsoft Word with the spell checker activated. Wedge (2021, p. 54) found that the percentage of misspelled words was lower in the control corpus (0.05%) compared to CORYL (3.0%), meaning that the participants made a higher number of spelling errors when writing by hand than they did when using Microsoft Word. It is noteworthy that the statement that the participants used Microsoft Word in the control corpus is an assumption rather than a variable which has been controlled for (Wedge, 2021, p. 56).

2.5 Pedagogical use of spell and grammar checkers

Based on the findings in their research, John and Woll (2020, p. 186) state that Grammarly and Virtual Writing Tutors' feedback is well-suited for use in English second language acquisition. On the other hand, Microsoft Word's grammar checker is not particularly useful due to its low accuracy. Furthermore, John and Woll state that an important note to make is that no matter what system is used, learners must still understand that the grammar-checker will not catch every mistake. Therefore, John and Woll (2020, p. 185) argue that using Microsoft Word's grammar checker is not recommended, as the coverage is too low for the pupils to benefit from it. Instead, from a teacher's perspective, the low overall coverage of the grammar checkers means that they cannot be relied on to provide meaningful corrective feedback on second-language compositions. Consequently, it would benefit teachers to familiarise learners with the limitations of the technology and not entirely rely on the feedback provided by grammar checkers (John & Woll, 2020, p. 185). John and Woll (2020, p. 187) findings show that grammar checkers' limited feedback and low overall coverage cannot replace human feedback, although they can be beneficial if teachers teach their pupils how grammar checkers can be used as a tool to enhance their learning outcome.

John and Woll (2020)'s suggestions regarding pedagogical use of spell and grammar checkers in the classroom are in line with suggestions provided by Dypedahl and Hasselgård (2018) and Lin et al. (2017). With the technology today, all word-processing programs are equipped with a spell checker function that corrects errors or provides an alternative word the writer could use (Lin et al., 2017, p. 1501). According to Dypedahl and Hasselgård (2018, p. 167), Microsoft Word is a widely recognized word-processing software commonly used as a writing tool in Norwegian schools. However, one can question pupils' knowledge regarding

grammar checkers. Dypedahl and Hasselgård (2018, p. 170) describe a grammar checker as a powerful tool if pupils know how to use it. Grammar checkers can prove to be an invaluable asset for educators, allowing them to equip their pupils with the necessary tools to improve their grammatical skills. When implemented in a sensible manner, learners can benefit from a wide range of possibilities that enable them to assess their language proficiency and make the necessary improvements. From identifying common errors to learning new vocabulary and improving writing style, using a grammar checker can help pupils achieve their language goals effectively (Dypedahl & Hasselgård, 2018, pp. 167-173). According to Lin et al. (2017, p. 1501), the main function of a spell checker is to correct errors in text production, and spelling plays a significant role in how a reader perceives a text, as errors can prevent the writer's intended message or idea from being conveyed effectively. Therefore, learners should have an adequate amount of knowledge regarding spell and grammar checkers.

Dypedahl and Hasselgård (2018, p. 168) state that one can create practical awareness in the classroom regarding grammar checker possibilities and limitations. If learners become familiar with various writing aids, such as grammar or virtual writing tutors, they can compare the errors that different spell checkers detect. By exposing pupils to the fact that not all writing tools and aids available on the internet are reliable, they can become more aware of their use of such resources. This may lead to improved writing skills as pupils learn to differentiate between reliable and unreliable online sources. A grammar checker in the classroom can be a helpful resource if utilized properly; however, Dypedahl and Hasselgård (2018, p. 169) highlight the importance of human feedback and how a teacher's assessment and analysis can never be replaced. Our study aims to identify the effects of the current use of spell and grammar checkers, and to gain information regarding the pupils' knowledge about using these tools. The method used in our research is described in the next chapter.

3 Methods

To answer our research questions, we have gathered data quantitatively, collecting texts written by pupils both with and without access to spell and grammar checkers. The pupils also answered a questionnaire, where we asked questions that would give us insight into their experiences with using spell and grammar checkers. In this chapter we will describe how we collected data, and changes we had to make throughout the process that may have affected our data. We will also address the validity and reliability of our study, and ethical considerations we made.

3.1 Research design

We have chosen to use repeated measures to compare differences in the data material. According to Creswell et al. (2021, p. 358), repeated measures, also known as within-subject design, is when researchers compare related measures from the same participants between different conditions. In our study, we have compared the differences, if any, in spelling and grammar when the students use spell and grammar checkers to when they do not use it. We have done this by giving a group of pupils two different tasks, where one was answered without access to spell and grammar checkers, while task number two was answered with this feature activated. Due to the design of our research we had to be present in the classroom when the pupils made their contributions. This was a necessity to ensure that all pupils followed the guidelines given to them, and that everyone received the same information.

In our research, we have used a mixed method approach to gain a more comprehensive understanding of our research problem. According to Creswell and Plano Clark (2011, p. 4) mixed methods research combines elements from both qualitative and quantitative research to gain a richer understanding of the research problem, and the method can be applied in any stage of the study.

Our primary method for collecting data was to use a quantitative approach. According to Creswell et al. (2021, p. 37), quantitative research aims to determine the overall tendency of individual responses and how it varies among people, mainly dealing with variables and their correlation with each other. We used two different methods for our data collection to ensure that we identified possible variables. First, we collected texts written by pupils both with and without access to spell and grammar checkers. We then gave them a questionnaire (see

chapter 3.3.2 and appendix) with four questions that they answered. The collection of both the questionnaire and pupils' texts will be addressed with more detail in chapters 3.3 and 3.4.

When analysing our data, we used a mixed method approach. First, we analysed our data quantitatively by finding statistical figures based on the whole sample. Next, we analysed specific texts using qualitative elements. Hammersley (2013, p. 12) state that one of the goals of qualitative research is to take into account the subjectivity in the research process and to study cases in detail. As the texts written by the pupils in our sample are suitable for both quantitative and qualitative research, we chose to analyse four of the candidates' submissions qualitatively to gain insight into the extent to which our results vary amongst individuals, taking subjectivity into account.

3.2 Validity and reliability

According to Cohen et al. (2018, p. 245), researchers must evaluate their studies' strengths and weaknesses throughout the process. Recognising these aspects is essential to determine the validity and reliability of the research. In this chapter, we briefly explain the concepts of validity and reliability and general aspects of our study which strengthen and weaken these. We will not go into detail regarding specific measures we have taken to strengthen these aspects in our research and choices we have made that may weaken them, as this will be discussed throughout the thesis.

3.2.1 Validity

Creswell et al. define *validity* as "the development of sound evidence to demonstrate that the intended test interpretation (of the concept or construct that the test is assumed to measure) matches the proposed purpose of the test" (2021, p .684). The term relates to how coherent the different parts of the research design are, so that they work together to answer the research questions. Cohen et al. (2018, p. 245) define validity as the extent to which the chosen method measures what it claims to measure. Addressing this with regard to our research questions, our method ensures that we get insight into the written skills of our participants, both with and without access to grammar- checkers. However, we recognized that the participants' prior knowledge in using grammar- checkers could be a variable we had to consider. We therefore designed a questionnaire to ensure that we have data to discuss all noticed variables.

According to Creswell et al. (2021, p. 345), *internal validity* is the degree to which inferences about the constructs or variables within the study are valid. In our study, the strength of the internal validity is largely determined by the extent to which our subjectivity has affected the results, as the error rates are based on errors identified by us. The coding process is explained further in subchapter 3.5.1. To strengthen the internal validity, we reviewed the pupils' submissions three times, to ensure accuracy and to avoid overlooking errors and coding data incorrectly. We also excluded error types that can be subject to individual interpretations. Additionally, we chose to deactivate the spell and grammar checker on the computer rather than using handwritten texts, which will be described further in subchapter 3.3. This was a measure we took to ensure correct interpretation of all words, avoiding subjectivity of interpretation.

External validity refers to the generalisability of the study's findings to other populations, settings, and contexts (Creswell et al., 2021, p. 345). The strength of the external validity in a study is dependent on the extent to which the research is relevant and applicable beyond the specific group that was investigated. As we did convenience sampling, meaning that we contacted teachers based on acquaintance, rather than probability sampling, our study is not possible to generalise. This means that the external validity of our study is weak. However, we did not find any indications that our sample misrepresents the population of English L2 learners in Norway, or that the sample stands out from the subgroup.

Construct validity concerns operationalising constructs so that it is clear what is meant when working with them (Cohen et al., 2018, p. 256). In our study, we have explained what we mean by spelling and grammar, and what error types are included in each category, to operationalise the terms for our analysis. Additionally, according to Cohen et al. (2018, p. 337), self-reporting in questionnaires can introduce bias, as respondents can under-report or over-report. Therefore, to strengthen the construct validity of our study, RQ2 is about the pupils' reported knowledge and experience rather than their actual knowledge and experience regarding using spell and grammar checkers. Moreover, Cohen et al. (2018, p. 337) state that the researcher must ensure that the respondents understand the questions and that they understand them the same way as the researcher. We did this by explaining the words used in the questionnaire (see chapter 3.3.2), and by letting the pupils ask for clarification. We also used Norwegian both in the questionnaire and in the discussion of the questionnaire. However, the construct validity of our study is weakened by including only four questions

which were quite comprehensive, rather than contextualising them and including more questions.

According to Creswell et al. (2021, p. 345) *statistical conclusion validity* refers to the appropriate use of statistics and statistical tests. In our analysis we have chosen tests based on the data we have, as many tests could be considered inappropriate for our sample size and the type of data we have gathered. We have also set the level of significance (see chapter 3.5.3) as recommended by Cohen et al. (2018, pp. 739-740) and as done in previous research in this field. Additionally, the conclusions drawn from the statistical tests have been evaluated taking into account the number of errors identified, meaning that we have had a critical view when interpreting our results.

3.2.2 Reliability

Reliability means that the results from a study should be similar when the research is repeated using the same instrument or research design, and that the scores from the data should be consistent and not influenced by the method of measurement (Creswell et al., 2021, p. 681). According to Gleiss and Sæther (2021, p. 202), reliability can be described as the quality of a research process and whether the research can be trusted. Here, we address the reliability, as recommended by Gleiss and Sæther (2021, p. 202), by asking ourselves two questions.

The first question we want to address is how our data can be affected by our method when gathering it. When gathering data, the amount of explaining we did regarding the questionnaire was affected by how many questions the pupils asked. We strived to have identical procedures in all schools when collecting data, but due to the pupils' individual needs there were some differences. For instance, when the pupils worked on their written submissions (see chapter 3.3.1 for further information) some of them had to be fed ideas regarding what to write. Additionally, as mentioned in the subchapter above, our subjectivity might have affected the number of errors we were able to identify. However, by defining each spelling and grammatical category beforehand and by relying on a plan with a manuscript when collecting data, we believe that the probability of there being a bias has been minimised.

The second question to be addressed is whether the results of our research can be reproduced by other researchers. To strengthen the reliability of our research we have strived to be transparent in our description of how we gathered data and of the procedure when analysing

our data. We have explained our methods in as detailed a manner as possible, to ensure that other researchers can replicate our research design.

3.3 The pilot study

Before conducting our research, we wanted to test whether the plan for our data collection had considered possible challenges and whether changes to our plan were necessary. As one of us was a subject teacher and homeroom teacher in a school, we were fortunately able to conduct a pilot amongst some of her pupils. The aim of the pilot was not to test whether we received interesting data, but to test our research design and to identify possible issues that needed to be addressed. In total, 19 pupils in the 8th grade participated in the pilot, and the data is not part of our final sample. In this subchapter, we will describe how we conducted the pilot study, and issues that were identified which led to changes in the main study. Therefore, the majority of information regarding our data collection is included in this subchapter.

3.3.1 The pupil/learner texts

In our project outline, we initially planned for task one to be handwritten by the pupils and for task two to be done in Microsoft Word with spell and grammar checker activated. We were also uncertain about the amount of time the pupils would need to produce enough text for the data to be usable in the study and if the tasks we designed would motivate the participants to write. As many pupils are unconfident and unenthusiastic when it comes to writing, it is important to give them interesting and enjoyable tasks to help motivate them to write (Harmer, 2007, p. 329). When making the two tasks we were going to give to the pupils, it was essential to us that the tasks had the potential to be enjoyable, and that they motivated them to write as much as possible. We therefore chose to design tasks where the pupils would write from their own experiences, as they could find it enjoyable to reminisce about memories from their lives. Another reason for the tasks being designed so that both tasks would be written in the past tense, was that it increases the comparability of the texts, including the error types which they might potentially contain.

Using texts written in the past tense is interesting also because it has been demonstrated that both L1 and L2 learners of English struggle with the past tense. In a study done on 68 grade school children with English as their L1, the participants performed a grammaticality judgment task on ten different grammatical structures (McDonald, 2008, p. 252). In her research, McDonald (2008, p. 264) found that two of the tested structures caused trouble even for children as old as 9 ½ to 11-years old. These were third person singular agreement and

regular past tense (cf. also Brown's morphological stages, Brown 1973). This is consistent with Li's (2012, p. 62) findings when testing whether there is a developmental relationship between the use of tense morphology and related syntactic properties in L2 learners in the early stages of language acquisition. Six children between the ages six and nine, with Chinese as their L1 and English as their L2, participated in the study (Li, 2012, p. 63). Li (2012, p. 66) collected data by giving the participants a variety of tasks every month in a span of seven months. Li (2012, p. 123) found the children's learning progress to be slowed down by them struggling to learn verb tense and agreement, particularly the past tense -ed suffix. Additionally, the participants in the study showed a tendency to develop sentence structure skills before mastering verb inflections. Admittedly, Li's research was done on participants with Chinese as their L1, and it is important to note that results of such studies may be affected by the participants' first language; nevertheless, based on the available research, we decided that past tense morphology was an interesting aspect of grammatical knowledge for us to focus on.

Based on the reasons mentioned above, the wording in task one was "write a text in English about the best school trip you have had. Include as many details as possible". When conducting the pilot, the pupils wrote task one on paper with no utilities accessible besides a pen and paper. The wording in task two was "write a text in English about your best memory from this year. Include as many details as possible". For this task, the pupils used their computers, and we orally explained how they would check that the spell and grammar checker was activated. We also checked that all pupils had activated the spell and grammar checker by checking each pupil's computer settings before they started writing. The wording of the tasks remained unchanged in the main study, as the pilot demonstrated that the pupils produced an expected amount of text when writing on the computer based on the time they were given.

One of the features of our design that we chose to change after doing the pilot, was that we no longer wanted the pupils to write task one by hand. The reason for this was that many of the pupils' handwriting was challenging to read, resulting in us having to interpret many words. We therefore viewed it as something that could weaken the validity of our study as, due to anonymity, we could not rely on member-checking to ensure that our interpretations were correct. While the pupils produced enough text when they wrote on the computer, we observed that the pupils did not produce enough text for it to be usable in our study when writing by hand for the first task. We reflected on the probability that this could be a result of

them writing in a slower pace when writing by hand, as the submissions written on the computer were significantly longer. We therefore decided that both tasks were to be written in Microsoft Word, with the spell and grammar checker only activated in the second task.

In the pilot study, the pupils were given 40 minutes to write each task, with no breaks after the first task. However, we observed that most of them had finished after 25-30 minutes, and started disturbing the pupils who were still writing. For this reason, we chose to change the timeframe the pupils were given for each task, and we decided that 30 minutes would be enough for each task in the main study. Another observation that we made, was that many of the pupils were tired and unfocused when they started writing the second task as they made some complaints about it. We were worried that they would produce less text than they had the potential to, and as a result we decided that we would add a 10-minute break after the first task in the main study.

We also made changes to the questionnaire from the pilot study to the main study. These will be addressed in the following subchapter.

3.3.2 The questionnaire

In addition to collecting pupils' texts, we also wanted to establish whether their knowledge regarding spell and grammar checkers would affect the result of our research. According to Gleiss and Sæther (2021, p. 142), questionnaires are valuable tools in research as they allow for quantitative data collection from a larger sample size than one would get from a qualitative interview. The primary objective of a questionnaire is usually to provide researchers with the necessary information to answer their research questions or thesis (Gleiss and Sæther (2021, p. 142). For instance, in our case, we wanted insight into whether the pupils knew that possible spelling errors are marked with a red line or that possible grammatical errors are marked with a blue line in Microsoft Word. Therefore, we decided to create a questionnaire for the pupils to answer after they had completed the writing tasks.

After completing the second task, the pupils received a questionnaire about their habits regarding the use of spell and grammar checkers, instructing them to respond with a simple yes or no to each question. According to Creswell et al. (2021, p. 439), the alternatives when answering a questionnaire are usually closed, where the participants have limited options. An advantage of using closed questions is that it increases the possibility of comparing answers in the analysis. Gleiss and Sæther (2021, p. 150) state that such alternatives help the

respondents understand the questions as the researchers can be more precise. However, as we wanted to allow the pupils to elaborate, we did not have boxes for them to tick off. According to Creswell et al. (2021, p. 440), open questions are suitable to gather information from the participants that the researcher has not considered. We wanted to combine the two to ensure that all pupils answered the questions they were given and made comments where they wished to do so. Additionally, we wanted to ensure that we gathered as much information as possible for our analysis. Therefore, we clarified that the pupils could provide additional information if they found it relevant. As a result, the questions in our questionnaire were the following:

1) Bruker du grammatikk-kontrollprogrammet (autokorrekturen) som ligger inne i tekstbehandlingsprogrammet (Word)?
   (Do you normally use the grammar-control programme (autocorrect) that is accessible in the word processor (Word?))

2) Bruker du et annet grammatikk-kontrollprogram (autokorrektur) enn det som er i Word (f.eks. Grammarly)?
   (Do you use another grammar-control program (autocorrect) than the one that is accessible in Word, for instance Grammarly?)

3. Skjønner du hvordan man bruker grammatikk-kontrollprogrammene?
   (Do you understand how to use the grammar-control programmes?)

4. Leser du beskrivelsene for hvorfor programmet anbefaler endringer (beskrivelsene som kommer før forslag til løsning ved rød eller blå linje under ord eller setninger?)
   (Do you read the descriptions of why the programme has recommended changes (the descriptions that appear with the changes that are suggested through a red or a blue line underneath words or sentences)?

When conducting the pilot, we gave no explanations of the questions, other than when pupils explicitly asked for clarification. As we received some questions regarding some terms in the questionnaire during the pilot, we decided to go through the questionnaire orally before handing it out to the pupils in the main study. This way, we reduced the likelihood of pupils answering the questions without understanding all the terms in the questionnaire, thus reducing the chance of misunderstanding the questions.

As we knew that some of the vocabulary used in the questionnaire could be unfamiliar to the pupils and also that pupils' English competence generally varies, we conducted the survey in

Norwegian. This was done to ensure comprehension and reliable answers and remained unchanged in the main study. In the next subchapter, the main study will be addressed.

3.4 The main study

After conducting the pilot study, we made adjustments as described in the previous subchapter and proceeded to collect data for our main study. In this subchapter, we will give a description of the sample who participated in our research, and the procedure for collecting data for our main study.

3.4.1 Description of sample

When collecting our data, we visited three classes in three different lower secondary schools in Norway. Two of the classes we visited were in the 9th grade and one was in the 8th grade. Our sample consists of 30 pupils, aged between 13 and 14 years old. As we did not include personal questions in the questionnaire, we cannot guarantee that we know all pupils' L1 and L2. However, as we gathered our data in an area with predominantly L1 Norwegian speakers, we can assume that most of the pupils had Norwegian as their L1 and English as their L2. This question is not crucial to our study as our focus is not on the participants' L1, but on the effect of using spell and grammar checkers. Initially, we wanted our sample to include about 60 pupils, but many of the pupils chose not to participate. In our sample, we chose to exclude the learners with an individual education program (IEP), which allows them to use digital writing tools in all written work. We will describe this further in the chapter addressing ethical considerations, viz chapter 3.6.

When collecting our data, it was essential that we recruited a high number of participants, as we aimed to do a quantitative study. When choosing what schools to visit, we used a nonprobability sampling method referred to as convenience sampling (Creswell et al., 2021, p. 170). Convenience sampling is described as a method where units are selected for inclusion because they are the easiest for the researcher to access (Creswell et al., 2021, p. 173). We chose to use our prior connections when gathering data for our research. As such, we contacted previous practice teachers and acquaintances to simplify our data collection process. Although convenience sampling can weaken the external validity of our study, our method for collecting data requires us to be present in the classroom. Therefore, the given timeframe and geographical logistics limited our probability sampling possibilities.

3.4.2 Collecting data

When collecting our data, we had planned the procedure based on the pilot described in chapter 3.3. According to Creswell et al. (2021, p. 652), it is important to keep an accurate record of the information collected, to organize it into data files for quantitative or theme analysis, and to examine the quality of the information. As all three data sets were collected separately in one session per class, we aimed for every school class participating in the study to receive the same experience, with no room for deviation or inconsistency. Therefore, we relied on a manuscript when introducing the tasks and the questionnaire to ensure all participants received the same information. This was done to prohibit factors within our control, such as misinterpretations of terms used in the questionnaire, from affecting the data. There were, however, factors out of our control which can have affected our data, thereby weakening the reliability of our study, such as factors relating to the day of the week and the time of the day we visited the schools.

Each pupil in our study was assigned a unique candidate number that they used for both writing tasks and the questionnaire. Before starting the first writing task, we helped the teachers in the classroom to assist the pupils in turning off the spell and grammar checker on their computers. As we had to go through the process of activating the spell and grammar checker multiple times, we had made a PowerPoint presentation explaining each step thoroughly. We did not proceed from one step to another until all pupils had finished the previous step. For instance, the first step was to open a document in Microsoft Word. We did not tell the pupils to write their candidate number until all pupils had opened a document.

In all groups, the pupils used Microsoft Word from Office 365. The first task was given to the pupils and, as mentioned in chapter 3.3.1, they were given 30 minutes to complete it. As mentioned in our description of the pilot, the tasks remained unchanged. Therefore, the first task was for the pupils to write a text in English about the best school trip they have had. To ensure that the guidelines for the use of spell and grammar checkers were followed, all teachers walked around the classroom for the entire 30 minutes. The pupils were then given a 10-minute break. For the second task, all the teachers ensured that the spell and grammar checkers were turned on, and the pupils were given another 30 minutes to complete the task. In this task the pupils were going to write a text in English about their best memory from the year.

During the writing sessions we observed that some of the pupils struggled with coming up with ideas on what to write about. In such cases we decided to help them come up with ideas by asking them questions that could help them recollect or that could give them ideas. According to Harmer (2007, p. 329), it is essential that the teacher has enough suggestions and ideas to feed the pupils with if they cannot think of anything to write about. However, we were conscious of asking them questions rather than feeding them concrete ideas, to ensure that we did not affect their submissions more than necessary. After the pupils had submitted their texts for task two, they were given 10 minutes to complete the questionnaire.

As explained in chapter 3.3.1, we had originally decided to focus on the acquisition of the past tense among our participants. Our original research question was therefore more specific, asking how the use of the grammar checker in Word affected the pupils' past tense forms. However, when analysing our data, it became apparent that the texts we had collected contained few past-tense errors, while they did contain a high number of spelling errors. As a result, we decided to broaden our scope to include both grammar and spelling errors. Unfortunately, as we collected data from the questionnaires at the same time as we collected written tasks, our questionnaire did not include any questions about spell checkers. One may suspect that people generally do not distinguish between the spell and grammar checker in Word (and other word processor programs), and the answers to the questions in the questionnaire may therefore be relevant for both spell and grammar checkers as they are normally both activated by default and have to be manually switched off. However, we cannot be sure that this is the case.

3.5 Procedure for data analysis

In this subchapter we describe the procedure for our analysis. We will give a description of how we coded our data, as well as a description of how we analysed the answers to the questionnaire. The coding program, the statistical software program we used and the statistical tests we ran will also be addressed here.

3.5.1 Coding procedure

Our text base is smaller than we had hoped for, as many of the participants produced less text in 30 minutes than we expected them to do. We therefore have what Creswell et al. (2021, p. 212) refers to as missing data, where data are missing from the database because participants did not supply them when we collected data. We counted submissions comprising less than

100 words per task as missing data. This is because we found it difficult to generalise about someone's language based on a few words. We could have set the limit higher, e.g. to 200 words, but we found that 100 words revealed a lot about these learners' language, particularly concerning spelling. Once we had examined our database for missing data, we needed to decide how to handle the missing data. We chose to eliminate the candidates whose submissions fell under the word limit from our database, although it yielded fewer participants to our study. Initially, we had 35 candidates. After removing the participants who produced less than 100 words, we were left with a sample of 30 candidates.

We used colour coding when analysing to differentiate between five codes that we had decided on beforehand. We differentiated between spelling errors related to speech sounds, spelling errors related to capital letters and apostrophes, verb-related grammatical errors and other grammatical errors. Our main reason for keeping verb-related grammatical errors apart from other grammatical errors was that verb-related errors were frequent in our data. We used NVivo when coding our data, which is a qualitative data analysis computer software package developed by Lumivero (NVivo, 1997). The program allowed us to manage, analyse, and visualise qualitative data and documents systematically and individually.

To ensure consistent analysis of the data, and to enhance the internal validity of our study, we coded each text three times. First, one of us read through the texts and colour coded them by hand. Then, the other person inserted the codes into NVivo whilst checking for errors that might have been overlooked or passages that might have been mistakenly marked as errors. Finally, we read through every submission aloud using a smartboard, to check our work a third time.

Some of the participants in our study wrote words that involved multiple errors. If a word has been spelled incorrectly, we have counted it as one spelling error, no matter the number of incorrect letters. If a word has errors that could fall into more than one of the above-mentioned categories, one single word could be counted as involving multiple errors. For instance, candidate number 35 wrote "we had a task to write down everything we se around us". The word "se" has been coded as both a verb-related grammatical error and a spelling error. The candidate has written the word "see" incorrectly, and the context requires the verb to be conjugated into past tense.

3.5.1.1 Verb-related grammatical errors and other grammatical errors

Before we started coding, we agreed on how to categorise the errors we identified. As mentioned above, we noticed that verb-related errors were frequent in our data, and we decided to separate verb-related errors from other grammatical errors. In this section, we describe the errors included in what we have coded as verb-related grammatical errors and other grammatical errors.

Amongst the verb-related grammatical errors, there are subject-verb agreement errors, i.e. lack of number and person agreement between the subject and the verb. An example of this is when candidate number 23 wrote "when they was done eating". Here, "was" is marked as a verb-related grammatical error, as "they" is plural and the verb "to be" should be in the plural past tense form "were". We have also included wrong use of tense and aspect based on context. For instance, candidate number 26 wrote "after we were done cleaning we got ready on the bathroom and then we walked to the main room and eat". In this sentence we have marked "eat" as a verb-related grammatical error, as the candidate started the sentence in past tense and the verb "eat" therefore also requires the past tense "ate". Additionally, our dataset includes many conjugation errors, where for instance irregular verbs are conjugated as if they were regular. As an example, candidate number 20 wrote that they "singed" their chants. As "to sing" is an irregular verb, "sang" would have been the correct form. Therefore, "singed" is marked as a verb-related grammatical error. We also included the use of progressive forms where they do not belong, for instance candidate number 20's incorrect use of -ing in "we were staying" rather than "we stayed". Lastly, the absence of infinitive markers is included in this category. For instance, in candidate number 11's submission, "going be with" is marked as a verb-related error as the infinitive marker "to" in "to be" is missing. Other grammatical errors include all errors that are not verb-related, but where grammatical rules are broken. In this category we have included the absence of articles and the incorrect use of articles. For instance, candidate 24 has written the noun phrase "an podcast". This was marked as an error as the form which is required is "a". We have also included the incorrect use of nouns. For example, candidate 6 wrote "he broke five tooth". We marked "tooth" as a grammatical error as the context requires the noun to be plural, viz "teeth". Incorrect use of pronouns and determiners are also marked as errors. This includes pronouns being used as determiners (for instance "they're" rather than "their"), lack of plural marking on a pronoun ("he like" rather than "he likes") and incorrect use of relative pronouns ("the team which" rather than "the team that"). We have also marked incorrect uses of adjectives and adverbs as grammatical

errors. For instance, candidate number 2 wrote "the trip there went smooth". We marked "smooth" as an error, as a Standard English text requires "smoothly" in a context such as this one. Lastly, we have categorised incorrect uses of existential "there" as grammatical errors (for instance, "it was a lot of noise" rather than "there was a lot of noise").

3.5.1.2 Spelling, capitalisation, and punctuation errors

When coding our data we relied on some of the categories from the University of Oxford (2014) style guide to determine whether spelling and punctuation could be considered correct. We chose specific categories from the guide which we used for our coding, and we chose to exclude certain categories based on the limitations of our study. For instance, we did not count writing numbers under ten using figures as an error, although the style guide clearly states that it is (Oxford, 2014, p. 7). This is an aspect of spelling that, to our knowledge, is not systematically taught in primary school; hence, Norwegian pupils are not likely to have knowledge about this, which makes it sensible to exclude this point of spelling from our data. For the same reason, we have excluded errors to do with the use of cardinal numbers in contexts where ordinal numbers are required, as in "on the 2 day" (candidate 25, task 1).

Additionally, we chose to exclude preposition errors among our grammatical errors as we consider such errors to be vocabulary errors. Additionally, we excluded spelling errors where it is clear that the pupil has pressed the wrong letter on the keyboard by mistake, as they have spelled the word correctly in other occasions. For example, candidate number 17 wrote "whwn" in one instance but has written "when" in other sentences. We therefore assume that the candidate pressed the wrong key rather than not knowing how to spell the word, as "W" and "E" are next to each other on the keyboard. Lastly, although the style guide from Oxford (2014, p. 23) regards it as an error, we did not count the American use of -ize/-yze/-ization as an error. As it is not stated in the curriculum whether pupils should write American or British English, we do not know what instructions their teachers have given the pupils regarding what style to write in.

When coding our data, we chose to have capitalisation errors as a separate category due to the frequent occurrence of these errors. The standard rules we relied on were for the pronoun "I", proper nouns and the first word in a sentence to be capitalised. As the Oxford Style Guide (2014, p.5) states that titles and subtitles also should be capitalised, we regarded the absence of this as an error. For instance, candidate number 27 has written both "norwegian" and the title of Roald Dahl's book "heksene" without capitalisation. These are both regarded as errors.

We disregarded most punctuation errors in our coding, as the placement of commas and full stops can be viewed as subjective in many instances. As an example, candidate number 25 wrote 453 words in task number two and included only two full stops. For us to count the number of errors in punctuation in this submission, we would have had to count the number of occasions where we would have added a full stop to improve the sentence. We considered this to be too dependent on our subjective assessments and therefore excluded it to enhance the study's validity. However, we chose to follow the Oxford Style Guide (2014, p. 9) for apostrophes, as mistakes made in that category can be related to spelling and can potentially lead to miscommunication. We therefore viewed the correct use of apostrophes to be *s* to indicate possession after singular nouns, plural nouns which do not end with *s* and indefinite pronouns. The use of apostrophe after plural nouns ending in *s* was also considered correct. Using *it's* to indicate possession was considered incorrect, as *it's* is a contraction of *it is* (Oxford, 2014, p. 9). We also counted it as an error if a candidate chose not to include an apostrophe when using contractions. For instance, "wasent" and "dident" have been marked as both spelling errors related to speech sound and apostrophe errors in candidate number 25's submission in task number one.

When coding our data we did not include the spelling of lexical compounds, as many compounds have variable spellings (e.g. chat room, chat-room, chatroom). However, pronouns and adverbs with only one spelling was included, such as "everybody" and "themselves".

3.5.2 Method for analysing the questionnaire

As mentioned earlier, we developed our questionnaire with closed questions. We asked the students to respond to our questions with either "yes" or "no", but we also gave them the opportunity to elaborate on their answers. Based on this we decided to categorise the pupils' responses into three different categories - "yes", "no", or "sometimes". We found that having more than three categories would make reading and understanding our table challenging. As we reviewed the responses to our questionnaire, we saw it was necessary to interpret some of the answers. Some of the questions had unclear answers which were difficult to place within a simple yes-no format. For example, when the pupils were asked whether they understand how to use grammar control programs, candidate number 25 answered "litt" (a little). This is one of the answers categorised as "sometimes". Based on this, all answers that indicate that the

candidates' knowledge and experience with grammar controls are limited, yet existent, are categorised as "sometimes".

3.5.3 Testing significance

Although the statistics from our research is not possible to generalise, as we did not do random sampling, it can be viewed as meaningful to do hypothesis testing to strengthen the validity of our results. To discover whether there is a statistically significant difference between the means of our two samples of data, we have conducted a t-test. The t-test is used to demonstrate whether the difference between the samples can be considered random or if it can be viewed as significant (Cohen et al., 2018, p. 777). Our samples represent the same group under two different conditions. The sample from task number one illustrates the participants' number of grammatical errors and spelling errors when the spell and grammar checker is deactivated, while the sample from task number two shows their number of errors in the same categories with the spell and grammar checker activated. This means that we are testing the same, single group under two conditions, which qualifies our data for a paired t-test.

According to Cohen et al. (2018, p. 780), a paired t-test can be used to test whether the difference between two samples of data is statistically significant when the same group is measured under two conditions. This is a type of hypothesis testing where the null hypothesis is tested. The null hypothesis must state that there, for instance, has been no change in the two groups after an intervention (Cohen et al., 2018, p.730), meaning that the researcher's work is to prove the null hypothesis to be untrue. In our testing, the null hypothesis is that there is no change in the number of errors counted per hundred words in texts where the pupils activated the spell and grammar checker in Word, from when they deactivated it. The t-test is done to determine whether our null hypothesis is supported or not supported by our data. When we had formulated our null hypothesis, we had to set a significance level. The significance level demonstrates the maximum risk one takes that observed differences are random (Creswell et al., 2021, p. 220). We chose to set the significance level to 5%, or 0.05, as a large proportion of previous research in this field has done so (Mackey & Gass, 2015, p. 308). This level of significance makes up the critical region in a curve, meaning that if the sample of data falls into that curve the null hypothesis can be rejected.

To determine whether our sample of data falls into the critical region, we had to use the statistical software SPSS to examine the p-value. The p-value represents the probability that

the variation in a sample is due to chance, as would be the case if the null hypothesis was true (Creswell et al., 2021, p. 220). Once the p-value has been calculated, we know that our results are statistically significant if we find that p=<.05, and we can reject the null hypothesis that there is no difference between the two samples. When analysing our data we have conducted multiple paired t-tests, to check the significance of the total error rates and the differences in error rates in each category.

3.5.4 Testing for correlations

To test for correlations between the results of the text analysis and the results of question one and question three in the questionnaire, we used a point-biserial correlation test. This is used to measure the correlation between a continuous variable and a dichotomous variable (Lærd statistics, 2018). According to Cohen et al. (2018, p. 766), a continuous variable is one that can be measured as any value between two points on a scale, such as weight, height, or in our study; difference in error rates per hundred words from task one to task two. A dichotomous variable can be described as a variable that can only take two values, such as male or female or pass or fail (Cohen et al. (2018, p. 766). As the questions in our questionnaire were semi-open, and some of the pupils had answers that could be categorised as "sometimes", we decided to include only the yes and no answers when testing for correlations. This is because the point-biserial correlation test is the only correlation test we found suitable for the type of data we gathered, and the test requires there to be only two values. As a result, only 28 of the candidates are included in the data used for this test, thereby weakening the validity of our findings.

When doing the point-biserial correlation test we chose two questions which we found to be relevant to test for correlations with the results of the analysis of the texts. These were, as mentioned above, questions one and three. As shown in section 3.3.2, question one asked whether the pupils normally use the grammar control program accessible in Microsoft Word, and question three asked whether the pupil understands how to use grammar control programs. Next, we scored their answers to make the variable measurable, where no=0 and yes=1. We tested the correlations with these responses and the individual differences in error rates from task one to task two, using statistics per hundred words.

The test provides a degree and strength of association between the two variables, ranging from -1.00 to +1.00 where 0.00 indicates that there is no association at all (Creswell et al., 2021, p. 392). This value is hereby referred to as rpb (Degree and strength of association

between two variables). Additionally, the test provides a p-value which determines whether the degree of association that we were given is statistically significant, suggesting whether we can reject the null hypothesis of there being no correlation between the two variables. For instance, if rpb = -.358 and p=.023 the degrees of association indicate a negative correlation and the p-value suggests that the correlation is statistically significant when the level of significance is set to 0.05 (Lærd statistics, 2018).

3.6 Ethical considerations

According to Gleiss and Sæther (2021, p. 43), researchers are responsible for all individuals involved or participating in research. Researchers in Norway must follow the national research ethics guidelines developed by the National Committee for Research Ethics in the Social Sciences and the Humanities (NESH). Before collecting data, we were required to apply to the Norwegian Agency for Shared Services in Education and Research (SIKT). This involved obtaining approved consent forms and confirming that our management of personal data complied with privacy policies. This was a necessary step to proceed with the data collection process.

As we are researching people, more specifically children, a thorough review of our research design was crucial. This included every aspect of our chosen methods, such as the questionnaires and the gathering of pupils' texts. According to Gleiss and Sæther (2021, p. 44), informed consent is fundamental in all research to avoid misunderstandings or consequences. As the participants were under 15 years old, the pupils' parents had to consent on their behalf. Preserving the participants' anonymity is essential; no one must suffer harm while participating in research activities. All necessary measures must be taken to ensure the safety and well-being of all participants involved (Gleiss & Sæther, 2021, p. 45). To ensure the participants' anonymity, we gave them a candidate number for both the texts and the questionnaires. The pupils then submitted their work to their teacher, who forwarded their submissions to us with only their candidate numbers on them. All tasks were uploaded using OneDrive and the teachers ensured that the pupils' names were excluded from the documents given to us. The teachers did this by converting all the tasks into portable document format (pdf) to maintain anonymity and then shared them with us through email.

Another ethical issue considered was each pupil's individual rights. According to Dysleksi Norge (2021), pupils diagnosed with dyslexia have the right to access a computer or tablet with reading- and- writing-supportive software for all writing tasks. Based on this, we found

it to be unethical to include pupils with an individual education plan (IEP), which gives the pupils access to writing assistance software, as they may not be comfortable with producing text without the mentioned support. As we did not know the pupils beforehand it was not possible for us to know how their participation would affect them, and we did not consider ourselves to be in a position to make a decision where the pupils would give up their rights in order to participate. However, in one of the groups we visited, the teacher chose to have all pupils participate in the study. The teacher then removed their texts from the sample given to us.

4 Results

In this chapter, we present the findings from our analysis from both the pupils' texts and the questionnaire. We present descriptive statistics, which can be described as presenting data without attempting to predict population parameters (Cohen et al., 2018, p. 727). This is done by using various tables and statistical measures. We also present inferential statistics to gain a more comprehensive understanding of the correlation of different variables. This means that we will test for possible correlations and statistical significance to try to check whether the data are internally consistent. First, we analyse the data quantitatively, meaning that we provide statistics based on the whole dataset. Then, we provide a more detailed analysis of some of the pupils' texts, where we have chosen four candidates' submissions that we found interesting.

4.1 A high reported knowledge and experience with grammar checkers

To answer research question 1, whether the pupils understand how to use the spell and grammar checking software in Microsoft Word, we analysed the data from the questionnaire. The bar chart below offers a visual presentation of the pupils' responses to the questionnaire.
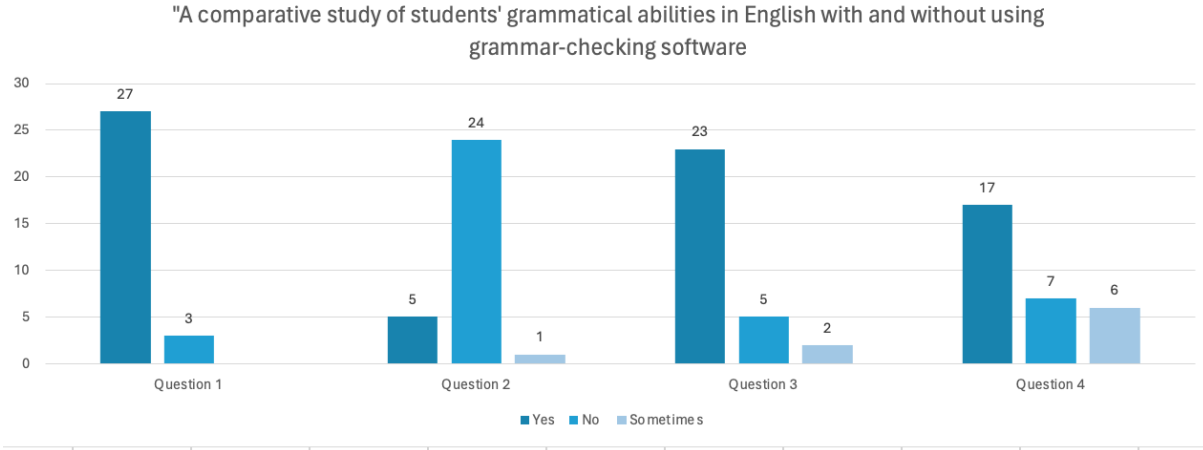
Figure 1: Responses to the questionnaire

As a response to question one: "Do you normally use the grammar-control program (autocorrect) that is accessible in the word processor (Word)?», 27 pupils responded with "yes" and only three pupils answered "no". This means that a total of 90% of the pupils normally use the grammar checker in Microsoft Word. In response to question two, "do you use another grammar-control program (autocorrect) than the one that is accessible in Word, for instance Grammarly?", there were five pupils who answered "yes", 24 who answered

"no", and one pupil who answered "sometimes". This adds up to a percentage of 80% who responded that they do not use other grammar checkers than the one in Microsoft Word. The answers to questions one and two suggest that the spell and grammar checker in Microsoft Word is used by many pupils and that it is used more frequently than other spell and grammar checkers.

In question three, we asked the pupils if they understood how to use grammar-control programs. 23 pupils answered "yes", five pupils answered "no" and two pupils answered "sometimes". Thus, while 77% of the pupils in the sample claim to understand how to use grammar checkers, 23% reported that they do not normally understand how to use it. In question four of the questionnaire, we asked if the pupils read the descriptions of why the program has recommended changes. In response to this question, 17 pupils answered "yes", seven answered "no", and six pupils answered "sometimes". Thus, only 56.66% of the pupils reported that they routinely read the descriptions of the recommended changes.

As 90% of the pupils reported that they use the grammar checker in Microsoft Word, and 77% claimed that they understand how to use the grammar checker, the figures suggest that 13% (or a total of four) of the pupils normally use the grammar checker without understanding how to do so. However, when looking at each individual questionnaire, three of the pupils reported that they do not normally use the grammar checker, but they do understand how to use it. Six of the pupils reported that they normally use the grammar checker in Microsoft Word yet responded "no" or reported that they were unsure about how to use it. This suggests that in reality, 22.22% of the pupils who normally use the grammar checker (six pupils), do not claim that they understand how to use the program.

When comparing the responses from question number 3 and question number 4, the figures indicate that only 20.34% of the pupils who claim to understand how to use the grammar checker read the descriptions of the recommended changes. However, while four of the candidates claimed that they did not understand how to use the grammar checker, the same candidates also reported that they read the descriptions that follow the suggestions. Eight of the pupils who claimed that they know how to use the grammar checker also responded "no" to question four. This means that 34.78% of the pupils who claim to understand the instructions from the grammar checker do not in fact use it.

4.2 Decreased error rates with spell and grammar checker activated

Table 1 displays an overview of the number of errors each participant made in tasks one and two. The errors have been classified as explained in section 3.5.1. The green section categorised as "total" is the total number of errors counted in each task.

Candidate number	Task 1							Task 1						
	No, Of words	Spelling errors	Verb errors	Capital letter	Apostrophe	Other grammatical errors	Total	No, Of words	Spelling errors	Verb errors	Capital letter	Apostrophe	Other grammatical errors	Total
1	173	11	6	3	2	7	29	118	0	1	0	0	1	2
2	188	5	1	0	0	3	9	178	3	0	0	0	2	5
3	225	4	16	5	0	1	26	124	0	4	0	0	0	4
4	202	5	2	2	0	2	11	211	1	3	0	0	4	8
6	323	24	7	14	0	2	47	348	8	0	3	0	1	12
7	214	6	0	2	0	1	9	245	2	2	1	0	2	7
8	198	5	1	2	0	1	9	251	9	0	3	0	0	12
9	446	4	1	0	0	1	6	375	1	1	2	0	0	4
10	307	2	1	0	0	7	10	355	3	6	0	1	3	13
11	358	2	3	6	3	2	16	365	0	3	5	1	3	12
12	201	34	10	4	0	1	49	231	44	7	6	1	2	60
13	165	5	0	0	1	0	6	184	1	1	0	0	1	3
14	349	8	4	0	1	3	16	468	4	5	2	2	2	15
15	172	19	3	4	1	7	34	206	5	2	6	0	3	16
16	274	6	2	15	4	4	31	117	0	0	1	0	0	1
17	202	31	9	7	0	3	50	102	4	4	2	0	1	11
18	317	9	7	0	0	1	17	272	7	10	0	0	0	17
19	210	3	0	3	0	2	8	175	1	1	2	0	0	4
20	180	15	3	8	0	1	27	135	1	3	5	0	0	9
23	357	16	4	4	0	2	26	218	2	1	0	1	2	6
24	115	2	4	2	1	2	11	298	1	6	0	0	4	11
25	364	27	11	4	8	1	51	453	44	18	14	0	4	80
26	417	11	9	2	0	0	22	150	3	4	2	0	0	9
27	551	6	4	5	1	2	18	622	6	5	2	1	2	16
28	296	2	0	2	0	0	4	193	0	2	1	0	0	3
31	298	10	4	3	1	2	20	102	0	2	0	1	0	3
33	146	9	1	3	1	2	16	215	5	5	2	0	3	15
34	135	3	0	1	0	4	8	205	4	1	1	0	0	6
35	192	15	1	4	0	2	22	289	2	4	1	0	0	7
37	263	3	1	8	1	0	13	192	0	3	3	0	1	7

Table 1: Number of errors counted in each category

When studying the data, we found that candidate number 17 is the candidate who had the biggest difference in error rates from task one to task two, counting a total of 50 errors in task one and 11 errors in task two. This is equivalent to a 78% decrease. However, our analysis will not be based on the figures in this table, as the number of words each candidate produced varies. For instance, candidate number 18 made six spelling mistakes in task two, while candidate number 26 only made two spelling mistakes in task two. However, candidate number 18 produced 622 words in the given task, while candidate number 26 produced 218 words in the same task. This means that while candidate number 18 had produced a higher number of errors compared to candidate number 26, the numbers of errors per hundred words are similar. Candidate number 18 had 5.4 errors per hundred words and candidate number 26 had 5.3 error per hundred words. Therefore, to give a more accurate representation of the number of errors counted per candidate, we have made a table demonstrating the figures per hundred words. This table will be introduced later in the subchapter.

Figure 2 demonstrates the number of words each participant produced per task. The dark blue bars represent the number of words produced in task one, where the participants did not access the spell and grammar checker. The light blue bars represent the number of words produced in task two, where all participants had activated the spell and grammar checker.
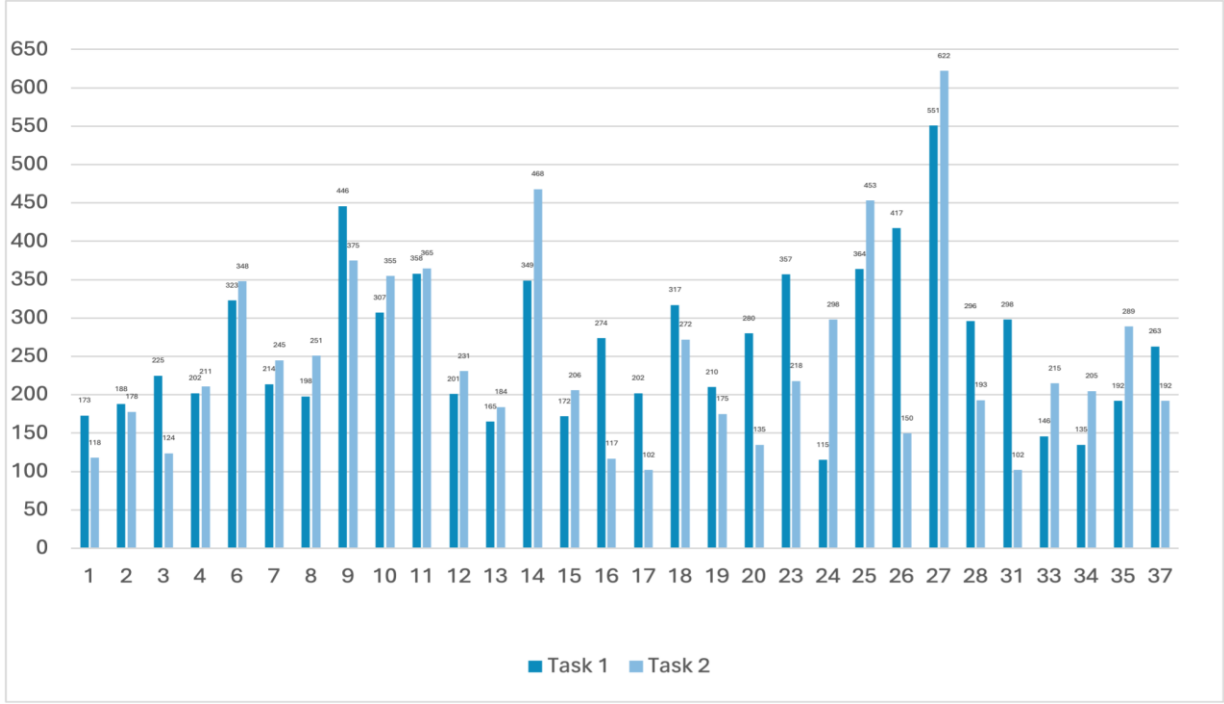
Figure 2: Number of words produced per task, per candidate

The mean number of words produced in task number one is 261.2 words and 246.5 in task number two, meaning that the participants, on average, produced 14.7 words more in task one than in task two. However, while 14 of the participants produced a higher number of words in task one than in task two, as many as 16 participants had a higher word count in task two compared to task one. On average, the participants with a higher word count in task one produced 99.6 more words in task one than in task two. The 16 participants with a higher word count in task two than in task one produced an average of 59.6 more words in the second task than in the first task. These figures could be a result other factors than singularly the use and non-use of the spell and grammar checker. For instance, some of the participants could have found the topic in task number one most inspiring, while others were more inspired by the topic in task number two. Additionally, some of the participants could have been tired when writing the second task, as they had already written for 30 minutes when it was introduced, causing their text production to become reduced.

Due to the differences in the number of words produced, we do not view it as meaningful to use the raw error rates when calculating the average number of errors made in each category. To give more comparable numbers, Table 2 shows the errors counted per hundred words (phw) per page, viz normalised numbers.

Candidate number	Task 1							Task 2						
	No, of words	Spelling errors phw	Verb errors phw	Capital letter phw	Apostrophe phw	Other grammatical errors phw	Total	No, Of words	Spelling errors phw	Verb errors phw	Capital letter phw	Apostrophe phw	Other grammatical errors phw	Total
1	173	6,36	3,5	1,7	1,2	4,0	16,8	118	0,0	0,8	0,0	0,0	0,8	1,7
2	188	2,66	0,5	0,0	0,0	1,6	4,8	178	1,7	0,0	0,0	0,0	1,1	2,8
3	225	1,78	7,1	2,2	0,0	0,4	11,6	124	0,0	3,2	0,0	0,0	0,0	3,2
4	202	2,48	1,0	1,0	0,0	1,0	5,4	211	0,5	1,4	0,0	0,0	1,9	3,8
6	323	7,43	2,2	4,3	0,0	0,6	14,6	348	2,3	0,0	0,9	0,0	0,3	3,4
7	214	2,80	0,0	0,9	0,0	0,5	4,2	245	0,8	0,8	0,4	0,0	0,8	2,9
8	198	2,53	0,5	1,0	0,0	0,5	4,5	251	3,6	0,0	1,2	0,0	0,0	4,8
9	446	0,90	0,2	0,0	0,0	0,2	1,3	375	0,3	0,3	0,5	0,0	0,0	1,1
10	307	0,65	0,3	0,0	0,0	2,3	3,3	355	0,8	1,7	0,0	0,3	0,8	3,7
11	358	0,56	0,8	1,7	0,8	0,6	4,5	365	0,0	0,8	1,4	0,3	0,8	3,3
12	201	16,92	5,0	2,0	0,0	0,5	24,4	231	19,0	3,0	2,6	0,4	0,9	26,0
13	165	3,03	0,0	0,0	0,6	0,0	3,6	184	0,5	0,5	0,0	0,0	0,5	1,6
14	349	2,29	1,1	0,0	0,3	0,9	4,6	468	0,9	1,1	0,4	0,4	0,4	3,2
15	172	11,05	1,7	2,3	0,6	4,1	19,8	206	2,4	1,0	2,9	0,0	1,5	7,8
16	274	2,19	0,7	5,5	1,5	1,5	11,3	117	0,0	0,0	0,9	0,0	0,0	0,9
17	202	15,35	4,5	3,5	0,0	1,5	24,8	102	3,9	3,9	2,0	0,0	1,0	10,8
18	317	2,84	2,2	0,0	0,0	0,3	5,4	272	2,6	3,7	0,0	0,0	0,0	6,3
19	210	1,43	0,0	1,4	0,0	1,0	3,8	175	0,6	0,6	1,1	0,0	0,0	2,3
20	180	8,33	1,7	4,4	0,0	0,6	15,0	135	0,7	2,2	3,7	0,0	0,0	6,7
23	357	4,48	1,1	1,1	0,0	0,6	7,3	218	0,9	0,5	0,0	0,5	0,9	2,8
24	115	1,74	3,5	1,7	0,9	1,7	9,6	298	0,3	2,0	0,0	0,0	1,3	3,7
25	364	7,42	3,0	1,1	2,2	0,3	14,0	453	9,7	4,0	3,1	0,0	0,9	17,7
26	417	2,64	2,2	0,5	0,0	0,0	5,3	150	2,0	2,7	1,3	0,0	0,0	6,0
27	551	1,09	0,7	0,9	0,2	0,4	3,3	622	1,0	0,8	0,3	0,2	0,3	2,6
28	296	0,68	0,0	0,7	0,0	0,0	1,4	193	0,0	1,0	0,5	0,0	0,0	1,6
31	298	3,36	1,3	1,0	0,3	0,7	6,7	102	0,0	2,0	0,0	1,0	0,0	2,9
33	146	6,16	0,7	2,1	0,7	1,4	11,0	215	2,3	2,3	0,9	0,0	1,4	7,0
34	135	2,22	0,0	0,7	0,0	3,0	5,9	205	2,0	0,5	0,5	0,0	0,0	2,9
35	192	7,81	0,5	2,1	0,0	1,0	11,5	289	0,7	1,4	0,3	0,0	0,0	2,4
37	263	1,14	0,4	3,0	0,4	0,0	4,9	192	0,0	1,6	1,6	0,0	0,5	3,6

Table 2: Errors per hundred words in each category

When calculating the mean of the total number of errors counted per hundred words, we found that 8.8 is the average number of errors per hundred words in task number one. In task number two, the mean is 5.0. This means there is, on average, a decline of 3.8, or 43.18%, in the number of errors per hundred words when the participants submitted texts with the spell and grammar checker activated. However, some pupils produced more errors in the second task than in the first task. Candidates 8, 10, 12, 18, 25, 26 and 28 all have more errors in task two than in task one. This means that the candidates in question made a higher number of errors in our pre-made categories when the grammar checker was activated than they made when it was deactivated.

When analysing the number of errors made per 100 words, candidate number 17 is no longer the participant who had the biggest difference in the number of errors counted in task one versus task two, as suggested by the raw counts (see above). In task number one, we counted 24.8 errors per hundred words in candidate number 17's submission. Candidate number 17's submission is therefore the submission where we counted the highest number of mistakes in task one. In task two, we counted 10.8 errors per hundred words in the same candidate's

submission. The candidate who had the biggest difference in number of errors between task one to task two, is candidate number 1. We counted 17.3 errors per hundred words in task number one, and 1.7 errors per hundred words in task number two. This means that candidate number 1 had 15.6 errors less per hundred words in task number two than in task number one, while candidate number 17 had 14 errors less per hundred words in task number two than in task number one.

To analyse the total number of grammatical errors versus spelling errors, we have made a table where spelling errors related to speech sound, capitalisation errors and apostrophe errors are merged into one category, viz spelling errors. Verb-related grammatical errors and other grammatical errors are merged into a category of "grammatical errors". We have included both statistics per hundred words and the actual number of errors we counted.

Candidate number	Task 1					Task 2				
	No, Of words	Spelling errors	Spelling error phw	Grammatical errors	Grammatical errors phw	No, Of words	Spelling errors	Spelling errors phw	Grammatical errors	Grammatical errors phw
1	173	16	9,3	14	8,1	118	0	0,0	2	1,7
2	188	5	2,7	4	2,1	178	3	1,7	2	1,1
3	225	9	4,0	17	7,6	124	0	0,0	4	3,2
4	202	7	3,5	4	2	211	1	0,5	7	3,3
6	323	38	11,8	9	2,8	348	11	3,2	1	0,3
7	214	8	3,7	1	0,5	245	3	1,2	4	1,6
8	198	7	3,5	2	1	251	12	4,8	0	0,0
9	446	4	0,9	2	0,4	375	3	0,8	1	0,3
10	307	2	0,7	8	2,6	355	4	1,1	9	2,5
11	358	11	3,1	5	1,4	365	6	1,6	6	1,6
12	201	38	18,9	11	5,5	231	51	22,1	9	3,9
13	165	6	3,6	0	0	184	1	0,5	2	1,1
14	349	9	2,6	7	2	468	8	1,7	7	1,5
15	172	24	14,0	10	5,8	206	11	5,3	5	2,4
16	274	25	9,1	6	2,2	117	1	0,9	0	0,0
17	202	38	18,8	12	5,9	102	6	5,9	5	4,9
18	317	9	2,8	8	2,5	272	7	2,6	10	3,7
19	210	6	2,9	2	1	175	3	1,7	1	0,6
20	180	23	12,8	4	2,2	135	6	4,4	3	2,2
23	357	20	5,6	6	1,7	218	3	1,4	3	1,4
24	115	5	4,4	6	5,2	298	1	0,3	10	3,4
25	364	39	10,7	12	3,3	453	58	12,8	22	4,9
26	417	13	3,1	9	2,2	150	5	3,3	4	2,7
27	551	12	2,2	6	1,1	622	9	1,5	7	1,1
28	296	4	1,4	0	0	193	1	0,5	2	1,0
31	298	14	4,7	6	2	102	1	1,0	2	2,0
33	146	13	8,9	3	2,1	215	7	3,3	8	3,7
34	135	4	3,0	4	3	205	5	2,4	1	0,5
35	192	19	9,9	3	1,6	289	3	1,0	4	1,4
37	263	12	4,6	1	0,4	192	3	1,6	4	2,1

Table 3: Spelling and grammatical errors per hundred words

The mean number of spelling errors per hundred words is 6.2 in task number one and 3.0 in task number two, which is a decline of 51.6%. The mean number of grammatical errors counted per hundred words is 2.6 in task number one and 2.0 in task number two, resulting in a decline of 23%. These figures could suggest that the spell and grammar checker contributed to the pupils having a greater decline in spelling errors compared to grammatical errors. However, while the percentage declines in errors seem substantial, we cannot draw

conclusions solely based on these figures as they do not include measures of the variability around the mean, making them less meaningful to compare. To check the variability, thereby testing whether the two means are comparable, we have calculated the standard deviation of the two categories in each task. The standard deviation for spelling errors per hundred words is 5.0 in task number one and 4.4 in task number two, indicating that there is more variability in task number one than in task number two. Similarly, the standard deviation for grammatical errors is higher in task number one, viz 2.1. This indicates more variation than in task number two where the standard deviation is 1.4. These figures indicate that the difference amongst the pupils in both spelling and grammatical errors are greater in task one than in task two, suggesting that the spell and grammar checker makes the pupils' individual competence less visible.

To test the significance of the figures in the table above we did a One-Sided t-test based on the above-mentioned means and standard deviations. The test showed that p= <.001 regarding spelling errors per hundred words, and p= .037 with regard to grammatical errors per hundred words. As we set the significance level to 0.05, the results of the t-test indicate that there are statistically significant differences between both grammatical and spelling errors counted per hundred words in task one versus task two. When comparing the two p-values it is notable that the difference in grammatical errors were less statistically significant than the difference in spelling errors, as the p-value is higher, increasing the probability of there being a false positive in this category.

4.2.1 Grammatical errors

In our sample, grammatical errors accounted for 32.7% of all errors counted, when calculating based on the raw numbers. The pie charts below illustrate the distribution of errors related to verb and other grammatical errors counted per hundred words. In the charts, the total number of grammatical errors equals 100%. The left pie chart represents the types of errors made in task one and the pie chart to the right represents errors made in task two.
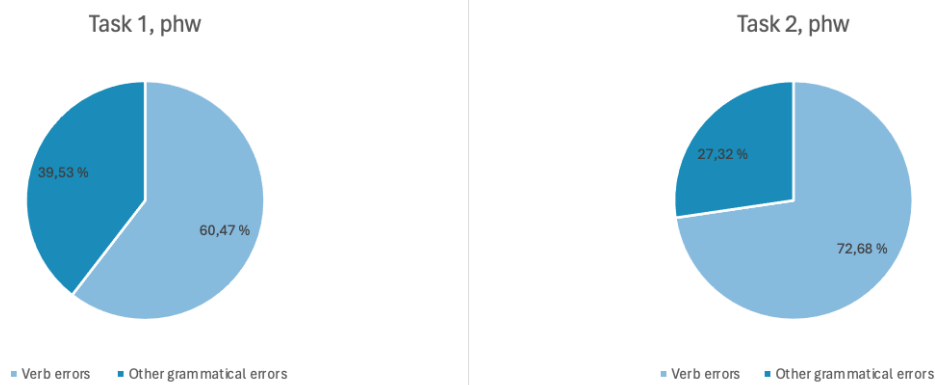
Figure 3: Distribution of the types of grammatical errors made in task one and task two

In task one, the grammatical errors the pupils made most frequently were verb-related errors, as 60.47% of the errors counted per hundred words fell into this category. Other grammatical errors accounted for 39.53% of grammatical errors. Based on the pie chart for task two, 27.32% of the errors counted are classified as verb-related grammatical errors, while other grammatical errors make up 72.68% of the total errors per hundred words. This change in distribution does not suggest that the pupils made, for instance, a higher number of verb-related errors in task number two, but that verb-related errors counted for a higher percentage of the total error count in task two than in task one. Although it is not possible to make assumptions based on these figures, they could indicate that the spell and grammar checker was more helpful with other grammatical errors, or that task two required the pupils to use a higher number of verbs than they found necessary in task one. In the following chapters, the two grammatical categories will be addressed in more detail.

4.2.1.1 Verb-related grammatical errors

Table 4 displays descriptive statistics for the verb-related grammatical errors counted in the two writing tasks. The figures are based on the number of errors counted per hundred words.

	Mean	**Standard deviation**	**P-value**
Task 1	1.6	1.7	
Task 2	1.5	1.2	P=.32

Table 44: Descriptive statistics for verb-related grammatical errors counted in task one and task two

When comparing the error rates in the two different tasks, we found that task one has a mean of 1.6 while task two has a mean of 1.5. This indicates that the pupils made a higher number of errors per hundred words in task one than in task two. The standard deviation for task one

is 1.7 while task two has a standard deviation of 1.2, suggesting a greater variability around the mean in task one compared to task two. This could indicate greater consistency in the pupils' verb-related error rates in task two than in task one.

Although the figures demonstrate a decrease in verb-related error rates per hundred words from task one to task two, we found that p=.32, suggesting that the decrease is not statistically significant. This could mean that the difference in error rates is caused by chance or random factors, rather than being an effect of using the spell and grammar checker. As a result, we cannot reject the null hypothesis that there is no statistically significant difference in verb-related error rates when the pupils access spell and grammar checker.

4.2.1.2 Other grammatical errors

Table 5 gives an overview of the mean, standard deviation, p-value in error rates for other grammatical errors in task one and task two.

	Mean	**Standard deviation**	**P-value**
Task 1	1.0	1.1	P=.004
Task 2	0.5	0.6	

Table 55: Descriptive statistics for other grammatical errors counted in task one and task two

The mean number of errors in this category counted per hundred words is 1.0 in task one, and 0.5 in task two. This is equivalent to a 50% decrease in error rates from task one to task two. The standard deviation for task one is 1.1 and 0.6 for task two, demonstrating a greater deviation from the mean in task one compared to task two.

Similarly to the findings in verb-related error rates, the statistical significance of the figures above also had to be tested. Based on these figures, task one appears to have a higher error rate and greater variation from the mean compared to task two. This could indicate that the spell and grammar checker reduced the number of errors the pupils made. The one-sided t-test calculated that p=.004 in this category, suggesting the difference in error rates to be statistically significant. This means that the difference in error rates when looking at other grammatical errors are not to be considered an effect of chance, as the p-value is lower than 0.05. Therefore, we can reject the null hypothesis that there is no statistically significant difference between error rates in other grammatical errors in task one compared to task two. However, we find it necessary to note that the mean numbers of errors per hundred words are low in both tasks, with one error per hundred words in task one and 0.5 in task two, meaning

that testing the significance could be viewed as less meaningful. For instance, a 50% decrease in error rates could be viewed as significant if the mean number of errors per hundred words was 30, thereby reducing the mean to 15 in task two. In contrast, a decrease of 0.5 errors per hundred words could be argued to be negligible, although the p-value can be used to argue otherwise.

4.2.2 Spelling errors

Calculated on the basis of the raw figures, spelling errors account for 67.3% of all errors. In task one, spelling errors comprise 70.7% of all errors. In task two, this category comprises 61.6% of the total number of errors. Figure 4 illustrates how the different error types in the spelling category are distributed, with the left chart representing task one and right chart representing task two.

Figure 4: Distribution of the types of spelling errors made in task one and task two

In the pie chart for task one, approximately one-third of the errors represent spelling errors related to speech sound. These errors account for 70.13% of all spelling errors in task one. Capitalisation errors comprise 25% of the errors, whereas apostrophe errors only make up 4.87% of spelling errors counted. The right chart visualises the distribution of errors in the same three categories in task two. Here, spelling errors account for 66.62%, capital letter errors for 29.97%, and apostrophe errors for 3.41%. When comparing the two pie charts, it can be observed that they share similarities in the distribution of the errors. There are, however, some changes in the categories. In the category consisting of spelling errors related to speech sound, there is a decrease of 3.68% from task one to task two. In the category addressing apostrophe errors, there is a 1.46% decrease. By contrast, capitalisation errors increased with 4.97%. Each of these subcategories will be addressed in the following chapters.

4.2.2.1 Capitalisation errors

Table 6 provides descriptive statistics based on the capitalisation errors made per hundred words in task one and task two.

	Mean	Standard deviation	P-value
Task 1	1.6	1.4	
Task 2	0.9	1.0	P=.003

Table 6: Descriptive statistics for capitalisation errors counted in task one and task two

In task one, the mean number of errors per hundred words is 1.6. In task two, the mean number of errors has decreased to 0.9. This indicates that the pupils made a higher number of capitalisation errors in task one than in task two, with a 43.75% decrease. The standard deviation for task one is 1.4, and 1.0 for task two, suggesting a greater variance around the mean in task one compared to task two.

While the mean and standard deviation indicate a clear difference between capitalisation errors counted in task one and task two, there is a need to test the significance of these findings as done for the grammatical categories. Here, we found that p=.003, suggesting the change to be statistically significant. This means that we can reject the null hypothesis that there is no significant difference in capitalisation errors counted in task one versus task two. However, similar to our findings in chapter 4.2.1.2, the number of capitalisation errors are low enough for one to question how meaningful statistical testing is. Another consideration to take into account, specifically in this category, is whether the data material is suitable for providing an indication of the pupils' use of capitalisation. In many of the texts, the candidates wrote multiple sentences without separating them with a full stop. Therefore, there are multiple cases where a capital letter would be needed if the punctuation had been done correctly. As we cannot make assumptions regarding whether the pupils would have made an error in these cases or not, it can be viewed as missing data.

4.2.2.2 Apostrophe errors

Table 7 gives the descriptive statistics of apostrophe errors counted in task one and task two.

	Mean	Standard deviation	P-value
Task 1	0.3	0.5	
Task 2	0.1	0.2	P=.025

Table 7: Descriptive statistics for apostrophe errors counted in task one and task two

The mean number of errors counted in task one is 0.3 and the standard deviation is 0.5. In task two, the mean number of errors counted is 0.1 and the standard deviation is 0.2. The figures indicate that the pupils made a higher number of apostrophe errors in task one compared to task two, and that the variation around the mean is greater in task one than in task two. When conducting a t-test in this category, we found that p=.025, indicating a statistically significant difference in error rates from task one to task two. One can, however, argue that the numbers are too low to find value in analysing this category separately from the spelling category as a whole, with there being close to zero errors per hundred words.

4.3 Correlation between error rates and reported usage of grammar control programs

To test the correlation between the differences in error rates from task one to task two and the answers to the questionnaire, we did a point-biserial correlation test, as explained in chapter 3.5.4. The null hypothesis is that there is no correlation between the answers to the questionnaire and the difference in error rates from task one to task two. First, we tested the correlation between the difference in error rates between task one and task two and the responses to question one. This question asked whether the pupils normally use the grammar control that is accessible in Microsoft Word. The point-biserial correlation test found that rpb= .05, indicating a positive correlation. This means that when the pupils answered that they normally use the grammar control program in Microsoft Word, the difference in error rates increased from task one to task two, suggesting that the spell and grammar checker corrected more of their mistakes. However, we also found that p=.80, resulting in the correlation not being statistically significant. This means that we cannot reject the null hypothesis, and we must conclude that there is no correlation between the answers to question one and the difference in error rates.

We also did a point-biserial test on the correlation between the difference in error rates and the pupils' answers to question three, which asked whether the pupils understand how to use the grammar control programs. Here, we found that rpb= -.051, suggesting that there is a negative correlation between the two variables. This means that when the pupils claimed that they knew how to use grammar control programs, the difference in error rates between task one and task two decreased, indicating that the grammar checker had less of an influence on their error rates. The test also found that p=.797, meaning that similarly to the results when testing the correlation to question one, the correlation is not statistically significant. Therefore, we must accept the null hypothesis that there is no correlation between the difference in error rates from task one to task two and the responses to question two.

While the results of the point-biserial correlation tests indicated no statistically significant correlation between the two variables, a careful scrutiny of individual submissions could suggest the contrary. The test described in this chapter is done to identify patterns in the sample collectively. While these tests can be argued to be meaningful in quantitative studies, we also find it valuable to look at individual submissions without generalising their results. The goal of doing so is to address variations amongst some the pupils in the sample, as we identified possible patterns which the quantitative figures do not give insight into. Therefore, the following chapter will address the results of four chosen candidates' written submissions, in conjunction with their responses to the questionnaire.

4.4 Individual effects of using, and knowing how to use, grammar control programs

In addition to the quantitative results presented in the previous chapters, we found four of the candidates' submissions to be particularly interesting to compare. While two of these candidates had significantly reduced their error rates from task one to task two, the other two candidates' written submissions had an increased number of errors per hundred words when the spell and grammar checker was activated compared to when it was deactivated. The two candidates whose error rates had increased from task one to task two, both gave indications that they did not know how to use the grammar control programs when asked in the questionnaire. In contrast, the two candidates who had decreased their error rates from task one to task two both stated that they know how to use the grammar control program. Based on this, we separated the four candidates into two pairs, viz pair one and pair two. Pair one, being candidate number 1 and candidate number 16, represent the group with a decreased number of

errors from task one to task two. Pair two, consisting of candidate number 12 and candidate number 25, represent the group with an increased number of errors from task one to task two.

To illustrate the changes in the written submissions, we have included the four pupils' texts in this part, both those written with spell and grammar checker activated and those written with the feature deactivated. As already mentioned, we colour coded the written submissions based on the categories described in chapter 3.5.1. The following texts include the coloured markings to indicate different types of spelling and grammatical errors found within them. Some words may have multiple colours, which indicate multiple errors. The colour codes are as follows:

> <mark style="background:green">Green</mark> = verb-related grammatical errors
> <mark style="background:cyan">Blue</mark> = other grammatical errors
> <mark style="background:magenta">Pink</mark> = spelling errors related to speech sound
> <mark style="background:yellow">Yellow</mark> = capitalisation errors
> <mark style="background:grey">Grey</mark> = apostrophe errors

4.4.1 A decrease in error rates with spell and grammar checker activated

As mentioned above, we chose to include candidate number 1 and candidate number 16 in pair one, as they had produced a much higher number of errors in task number one than in task number two.

4.4.1.1 Candidate number 1

We counted a total of 29 errors in candidate number 1's submission for task one, which is equivalent to 16.8 errors per hundred words. In task number two, we identified a total of 2 errors, making the number of errors per hundred words 1.7. This means that the pupil has had a 90.17% decrease in error rates from task one to task two. In task one, candidate number 1 wrote the following:

> Task one, candidate number 1 - Candidate 1
> The best trip I have ever had was when I was in 7th grade and my class <mark style="background:green">were</mark> at camp school. The trip was the best <mark style="background:grey">Ive</mark> ever had with my school because <mark style="background:yellow">There</mark> was very social and everyone eas together and had fun at <mark style="background:magenta">everytime</mark> at the day. Phones <mark style="background:green">was</mark> not allowed so everyone <mark style="background:green">were</mark> social and <mark style="background:green">were</mark> outside playing with <mark style="background:magenta">eatch</mark> other. The food was very good, we had chickenwings, pizza and <mark style="background:magenta">mutch</mark> more good food. We had fun activities, we <mark style="background:green">shoot</mark> with airguns and <mark style="background:cyan">bow</mark> and arrow, we <mark style="background:magenta">paddeld</mark> in <mark style="background:magenta">canos</mark> and we

climbd climbing wall. On wednesday we had jog as an activity, we got told a rute to run. It was an arrow in the forest that told us the way so me and my friend turnd the arrow so everyone ran wrong way and the next round we turnd it back so noone were gonna know. the camp school was one of the funniest trip Ive been to with my school.

The candidate wrote a 173-word submission for task one, we identified 6.36 spelling errors related to speech sound per hundred words (11 tokens). 36% of these (4 tokens) involve the lack of the letter "e" when attempting to use the -ed suffix, indicating that the pupil has spelled the words according to their pronunciation. The candidate also inserts a "t" in "eatch" and "mutch", which is presumably also due to the fact that there is a "t" in the pronunciation. In addition to spelling errors related to speech sound, we counted 1.7 capitalisation errors per hundred words (3 tokens). The lack of capitalisation when writing the proper noun "wednesday" could indicate that the pupil is not familiar with the rule that words denoting weekdays must be capitalised in English, as weekdays are not capitalised in Norwegian. However, as we only identified one of these errors and no other proper nouns were used in the text, no generalisation can be made about such errors. We also counted 1.2 apostrophe errors per hundred words (2 tokens), where all involve the lack of apostrophe when using contractions. Rather than contracting "I have" to "I've" they have written it without the apostrophe, viz "Ive.".

In addition to spelling errors, we counted 3.5 verb-related grammatical errors per hundred words (6 tokens). 71.43% of these errors are errors related to subject-verb agreement, i.e. the choice between "was" and "were". We also identified 4.0 other grammatical errors per hundred words (7 tokens). Some of these involve the non-use of determiners where they are needed, for instance in "ran wrong way" and "climbd climbing wall" rather than "ran the wrong way" and "climbed the climbing wall".

To illustrate how the spell and grammar checker affected candidate number one's error rates, we have included their submission for task two. The candidate wrote the following:

Task two, candidate number 1
The best things that has happened is that I met some of the best handball players. In the summer of 2023, I met some of the best players in handball and I even got their signatures. One of the players even played for the club I play for now named Bravo. I met them at a handball training camp named Camp Sagosen. The camp was to be

better at handball and learn new tactics and have fun. There were almost 200 other
players at the camp, and everyone <mark style="background-color:#00ff00">were</mark> trying to get better at handball. My whole
team was at the camp, and we slept at a camping place when we weren't training.

In this task, the candidate submitted 118 words. As mentioned above, we counted 1.7 errors
per hundred words (2 tokens) in candidate number one's submission for task two, which
means that the pupil decreased their errors with 15.6 errors per hundred words. While the
pupil still has one incident of incorrect subject-verb agreement with the verb "to be", this error
type has decreased with 80% (4 tokens). The incorrect subject-verb agreement could be a
result of the pupil confusing "everyone" for plural rather than singular, since the word has
plural meaning. In task two, the pupil has also included the letter "e" when using the -ed
suffix. The error coded in blue was interpreted as an incorrect use of "things". This
interpretation was based on the context, there being apparently only one thing that happened.
Additionally, the pupil has capitalised proper nouns correctly in task two, and there are no
further capitalisation errors, which could indicate that the spell and grammar checker has
effectively assisted the pupil with these corrections.

The decrease in error rates in candidate number 1's submission can be viewed in connection
with the pupil's answers to the questionnaire. When the candidate was asked whether they
normally use the spell and grammar checker in Microsoft Word, they wrote "yes". In the
second question, the candidate claimed that they do not use any other grammar control
program than the one accessible in Microsoft Word, which could suggest that the pupil has
exclusively familiarised themselves with the spell and grammar checker used in this study. In
their response to question three, candidate number 1 claimed to understand how to use the
grammar control programs, while in question number four they answered that they do not
read the descriptions that follow recommended changes in Microsoft Word. The decrease in
error rates and the responses to the questionnaire could indicate that the pupil benefits from
being familiar with using the grammar control program, given that many of the errors counted
frequently in task one are not identified in task two.

4.4.1.2 Candidate number 16

We identified a total of 31 errors in candidate number 16's submission for task one, which is
equivalent to 11.3 errors per hundred words. In task number two, we only identified 1 error,
making the number of errors per hundred words 0.9. This means the pupil has had a 92%

decrease in error rates from task one to task two. In task one, candidate number 16 wrote the following:

Task one, candidate number 16

The best school trip i have been on we're back in seventh grade when me and my class went on a field trips to haraldvoll. We went with the seventh grade of Tromsdalen school. When we arrived at the camp we had to meet outside the main house to get groups we were going to live with, i got in a good group and it was very fun. We where at the field trip from monday to friday and every day we did fun activity, we had to make our bed ready early in the morning to get ready for breakfast.

The best vacation i have ever been on was this years vacation i went to croatia with my friend and his family we went on a katamaran for 10 days and visited cool places we went to an island called vis its were the movie mamma mia was filmed i thought it was filmed in greece since its where it said it was but it was actually filmed in croatia. We also went to a yacht club where we saw Tommy hillfigers private yacht and we saw the third largest yacht in the world the katamaran looked small compared to the big yachts and the katamaran were about 60 foot. And we had a small boat on the katamaran.

In this task, the candidate submitted 274 words. We counted a total of 9.19 spelling errors per hundred words (25 tokens) in candidate number 16's submission for task one, meaning that these errors account for 83.18% of all errors in the submission. Amongst these, we counted 2.19 spelling errors related to the spelling of speech sound (6 tokens). While the pupil has used the words "where" and "were" correctly multiple times throughout the submission, the two words have been confused with each other twice, indicating uncertainty regarding the meaning of the words. In addition to spelling errors related to speech sound, the pupil has made 5.5 capitalisation errors per hundred words (15 tokens). The personal pronoun "I" lacks capitalisation in all incidences where it has been used. Additionally, with the exception of "Tromsdalen", the candidate has not capitalised proper nouns in their submission. While the lack of capitalisation when writing "monday" and "friday" could indicate that the pupil has followed Norwegian spelling rules, both languages require proper nouns such as "Croatia" and "Mamma Mia" to be capitalised. Moreover, the pupil has made 1.5 apostrophe errors per hundred words (4 tokens). These errors include lack of apostrophe when using contractions and when referring to the possessive form "year's". Instead, the pupil has written "years" and

"its", with "years" confusing the possessive and plural form, and "its" confusing the possessive form with the contracted form. There is also an overuse of contractions with "we're", indicating that the pupil does not know when contractions are appropriate.

We counted 2.2 grammatical errors per hundred words (6 tokens), where verb-related errors account for 28% (2 tokens) of the errors and 71% of the errors are other grammatical errors (4 tokens). All the four non- verb related grammatical errors involve problems in choosing between the singular and plural form of nouns. Regarding verb-related grammatical errors, the pupil has used the plural "were" with the singular subjects "the vacation" and "the trip".

To demonstrate the effects of activating the spell and grammar checker on the candidate's spelling and grammar, we have included their written submission for task two. Here, candidate 16 wrote the following:

> Task two, candidate number 16
> The best memory I have from this year is from Easter when I went to Switzerland and went skiing in the mountains, we went to a ski resort in Switzerland, and it was extremely fun. In the middle of the day, we went skiing and, in the evening, we went out to eat. I saw many expensive cars one car worth 45 million Norwegian kroners it was a Koenigsegg and about 5 Bugatti
> And when we went home, I didn't go in the ski resort in Tromsø at all that season because it was so small and boring compared to the ski resort in Switzerland. One time when I was in kroken

Regarding task number two, the candidate submitted a written piece of 117 words, we identified only one error in candidate number 16's submission, which is equivalent to 0.9 errors per hundred words. The single error identified is a capitalisation error, where the pupil has not capitalised the proper noun "Kroken". As Kroken is the name of an area in Tromsø, and can therefore be considered local, the spell and grammar checker might not have identified the term as a proper noun. There are no other capitalisation errors. All instances of "I" and all proper names are spelt with initial capitals. In contrast to the candidate's submission in task one, the text written with the spell and grammar checker activated uses correct subject-verb agreement, and we identified no grammatical errors or spelling errors related to speech sound. There are also no apostrophe errors, and in the one use of apostrophe, in the contracted form "didn't", it has been used correctly.

When answering the questionnaire, candidate number 16's responses were identical to candidate number 1's responses. When asked whether they normally use the grammar control program that is accessible in Microsoft Word, the candidate answered "yes". The pupil also claimed to understand how to use the grammar control programs and stated that they do not use any other grammar control program than the one accessible in Microsoft Word. These responses could indicate that the pupil is confident in using the grammar checker in Microsoft Word, which could suggest that the grammar checker has contributed to the decrease in error rates from task one to task two in candidate number 16's submission. When asked whether they read the descriptions following the suggested changes, the candidate answered "no". It is therefore not possible to predict the extent to which the pupil understands how to use the grammar checker solely based on the answers to the questionnaire. We nonetheless discovered that two pupils with similar error rates as candidate number one and candidate number 16 in task one, who did *not* claim to understand how to use grammar control programs, had increased error rates in task two. These candidates' submissions will be addressed next.

4.4.2 An increase in error rates with spell and grammar checker activated

As mentioned above, we have chosen to include candidate number 12 and candidate number 25 in pair two, as we counted a higher number of errors per hundred words in task number two than in task number one for these candidates' written submissions.

4.4.2.1 Candidate number 12

We counted a total of 49 errors in candidate number 12's submission for task one, which is equivalent to 24.4 errors per hundred words. In task number two, we counted 60 errors in total, making the number of errors per hundred words 26.0. This means that the pupil has had a 6.6% increase in number of errors from task one to task two. In task one, candidate number 12 wrote the following:

Task one, candidate number:12

One time i was on a school trip when i was inn 7 grade, vi went to a place named Sommarøya. We tok a buss to get there, and when we arrived vi got a tour on the place and the cabins we where going to sleep inn, each grup got a cabin, and i think we where 6 gups i think, and then when vi packt our tings out in the cabins and choose rooms, then we made our selfs some hotdogs at the beach. When vi had eaten vi runed

off and play and visited the other grups, after that vi all vent to our own cabins and vi whare with our grup and ate candy and drank soda. Then it was time to sleep, but we warent planing on sleeping, we sneek out, we troed all of our blankets and then we jumped down, then we wisit our friends and hang out with them and walk to a light tower and then vi hed back and to get up the window we found water hoase and we troed it up to the to friends that stayd there and vi climed up the water hoase

In this task, the candidate submitted a text consisting of 201 words. We counted a total of 16.92 spelling errors related to speech sounds per hundred words (34 tokens). Therefore, this is the category where the pupil made the highest number of errors. With some of these spelling errors it is apparent that the pupil has spelled the words as they are spelt in cognate words in Norwegian. For instance, "vi", "inn", "buss" and "tok" are in accordance with Norwegian spelling. Some of the incorrect spellings also indicate that the pupil spells words phonologically. For instance, "grup" and "stayd" are both spelled according to the pronunciation of the words.

Candidate number 12 also made 2.0 capitalisation errors per hundred words (4 tokens) in task one. These are all errors where the pupil has not capitalised the pronoun "I". While the pupil demonstrated accuracy with capitalisation when starting new sentences in this task, the issue of missing capitalisation data is relevant in this submission. By missing data, we mean that it is not possible for us to comment on the matter, as there are too few incidences that allowed the candidate to demonstrate their knowledge on this. However, as the pupil has capitalised correctly the three times they started a new sentence, the data suggests that the pupil manages to do this independently. In task one, candidate number 12 does not have any apostrophe errors. While this could indicate that the pupil knows how to use apostrophes, it could also be a result of the task not requiring the pupil to demonstrate their skills in these areas as they have not used any contractions or possessive forms that require it.

In addition to the spelling errors, we counted 5.0 verb-related grammatical errors per hundred words (10 tokens) in task one. This might be a result of the pupil having to use irregular verbs frequently throughout their text, while still trying to conjugate the verbs as if they were regular. "Runed" and "troed" are both examples of this. The pupil has also failed to use past tense forms in in general, using base forms such as "choose", "play" and "walk". There is only a single instance of a non- verb related grammatical error, which was identified as a lack of determiner.

To illustrate the spell and grammar checker's effect on candidate number 12's written submissions we must look at what they wrote in task two. With the spell and grammar checker activated, the pupil wrote the following:

> Task two, candidate number:12
>
> It was a time in the sommer breake me and my friend whent to his cabin in malangen his father drived and vi stoped at eidehandel to bought lotts of food and then we drived along. When vi got to his gabin vi packed out our tings, when we ware done we fild up the jetskis bacaus they have 4 of them so vi got our gir in and he tok hi's bout and i tok the jetski. Then vi drived and i jumped on some waves and tok a wheeilie and playd around on the jetski, when we had drived for a litle bit vi went to pick up Patrick that lives there in malangen and we drove and drove and drove then we almoust ran out off gass so vi had to drive back to hes cabin to fill up. When we came there i drive the jetski on to the jetski duck and we filled it up and chilled a bit in the cabin, and we grabed a snack. Afther that vi tok two of the jetskis out on the wather then me and my friend drove together and vi played a little game, that ve were going to tryed squirt on eachother with the jetskis. And there was so fun, then vi drove back ant vi goth the bout and whe went fishing.

In task number two the candidate wrote 231 words, we counted a total of 19.0 spelling errors related to speech sound (44 tokens), which is an increase of 12.3% from task one. Some of the errors the pupil made in task one have been corrected, for example "packt" and "inn" being changed to the correct "packed" and "in". However, it is apparent that the candidate's phonological spelling has not been notably improved by the spell checker. For instance, the fact that the candidate spells "played" and "filled" as "playd" and "fild" indicates that the pupil writes the words according to how they are pronounced, rather than using the rules for past tense -ed. Additionally, the pupil uses Norwegian spelling also in this task. While there are some examples from the text where the pupil has corrected "vi" to "we", the error has remained uncorrected 10 times. Other examples of the pupil spelling English words according to Norwegian orthography are when the words "gas" and "summer" are spelled "gass" and "sommer".

There are 2.6 capitalisation errors (6 tokens) and 0.4 apostrophe errors (1 token) per hundred words in candidate 12's submission for task two. In this submission, the candidate has included more full stops compared to task one, and the beginning of new sentences are still

capitalised. It is unknown whether this is because the candidate manages to follow this rule, or because the spell checker automatically turns small letters that follow a full stop into capital letters. The omitted capitalisation of "I" has not been corrected and the pupil has not capitalised proper nouns, indicating that the spell and grammar checker has had little to no effect on the candidate's capitalisation.

When looking at grammatical errors, we counted 3.0 verb-related errors (7 tokens) and 0.9 other grammatical errors (2 tokens) per hundred words. While the incorrect conjugation "runed" has been corrected to "ran", the form "drived" occurs 4 times, accounting for 57% of all verb-related errors. Interestingly, the text also contains the sequence "drove and drove and drove", as well as an instance where the base form "drive" is used, i.e. where the candidate has not attempted to conjugate the verb for past tense. The latter use is, however, the only use of the base form to express past tense, of which there were several examples in text number one.

Some of candidate number 12's responses to the questionnaire align with the results of the text study. Specifically, although the candidate claims to use the grammar checker in Word and to read the descriptions of the suggested changes, they also admit to not understanding how to use the grammar checker. When comparing the responses to the questionnaire to the results of the analysis of the texts, it can be argued that the pupil's lack of knowledge on how to use grammar-control programs can have contributed to their ineffective use of the spell and grammar checker. The findings when comparing candidate 12's responses to the questionnaire with the difference in error rates between two tasks, can be compared to the results found in candidate number 25's submissions. Therefore, candidate number 25's submissions will be addressed next.

4.4.2.2 Candidate number 25

Similarly to candidate number 12, candidate number 25 increased their number of errors from task one to task two. We counted a total of 51 errors in candidate number 25's submission for task one, which is equivalent to 14.0 errors per hundred words. In task number two, we counted 80 errors in total, making the number of errors per hundred words 17.7. This means that the pupil has had a 26.4% increase in error rates from task one to task two. In task one, candidate number 25 wrote the following:

Task one - Canditade number 25

In the seventh grade we went to leierskole we went the second week of school. So we packed home 1 to 2 days before with things like a sleeping bag and clothes. So on Monday week 2 of school we went on the bus we where the first then we went to the other schools to pick up the other kids. We drove tru bardufos to our leirskole. I was shooked why because I thout it was different but then we went in a big meeting in the senter of the whole leirskole and they of course had a list of names lets say eskil gabby and Lil maggy you vil go to the birds cabbin but there wasent a birds cabbin. Then ive got my name and a friends name and other lads. We go inside open up the bags make our beds and then we get called up for food this time it wasent a ring it was a line with our group and the others we wait our turn. The our turn is we had to take off our shoes then we go inside by our table. The groups go the us the food was lagsania we and I ate it was one of the top foods in the 5 day we were there. I don`t remermber yhe rest of the day but I know we ate and I met folk. On the 2 day we got woke up ate brekfast on the menue was bread and crisp bread . The intresting thin was our lunch was the food we took from the cafeteria and I dident so we went to the lake to make pankakes and everybody made mine was perfect after I ate I began to get hungry I was lake a dog begging for a another pancake but the som boy got a fish thet I got to eat it I took a bite forgetting there was bone I an fish then I picked out the fish bone ate the fresh fish it cured my hunger I was happy for the boy he was the only one to catch a fish by the river after the we went up again

In task one, the candidate submitted a 364-word text. We identified 7.4 spelling errors related to speech sound (27 tokens) in candidate number 25's submission. Similarly to candidate number 12, this pupil seems to spell some words phonologically. The spellings "brekfast", "som" and "intresting" all indicate that the pupil has spelled the words according to how they are pronounced. Moreover, the pupil has added an "e" in some contractions, for instance when writing "dident" and "wasent" rather than the correct "didn't" and "wasn't".

Other spelling errors made in the submission are capitalisation and apostrophe errors. We counted a total of 1.1 capitalisation errors (4 tokens) and 2.2 apostrophe errors (8 tokens) per hundred words. All capitalisation errors represent a lack of capitalisation in proper nouns. This has been done correctly only once in the text, which could suggest that the pupil is not familiar with the rule that all proper nouns should be capitalised, or that they do not know what qualifies as a proper noun. Notably, the proper nouns which the pupil has not capitalised

are nouns which are also capitalised in Norwegian, while the noun which is correctly capitalised, viz "Monday", would not have been capitalised in Norwegian. 62.5% of all apostrophe errors counted in the submission represent a lack of apostrophes in contracted words (lets, dident, wasent), which the pupil has done correctly once throughout the text. This could indicate that the pupil does not know when the rule applies. In two cases, the apostrophe is left our with possessive '*s*, as in "birds cabbin" and "friends name".

We counted 3.0 verb-related grammatical errors per hundred words (11 tokens) in task one, and 0.3 other grammatical errors (1 token). While the pupil has conjugated some verbs correctly in task one, such as "we drove" and "I picked out", there are several verb-related errors. All the verb-related errors represent the pupil not conjugating verbs into past tense where the context requires them to do so. For instance, "go", "make" and "ive" are all written in present tense or the base form in a past tense context.

With the spell and grammar checker activated, candidate 25 wrote the following:

> Task two canntidate 25
> it was summer we it was our first tornument of the year I play football so we started by getting teksts about the tornument in piteå and when we got close at 3 to 2 days we packt like sleeping bag and an air mattres and when the day was we woke up at 10 and started driving 11 and how to get to piteå it is long you have to drive the kvaløya bridge and tru the city take the tromsø bridge and and drive we stoped at the shoping senter we call paremiden I got deodorant from normal and a fizzy drink on how to get to piteå you have to roudes going right and left you have to take the right you drive past a dealer ship full of cars and othe stuf and you just drive a long time no rights of lefts until you get to volan most times almot mandetori you stop at the triangle kjosk and stop to get the delisios milkshake but then you go just forward til you see a roud to the left saying kilpisjårvi you drive up the mountain then you see after a 30 min the a fine river with a very light blue colour but you carry on thru the mouthin and then a right or left you have to take the right carry n the again drive until the round is just straight you see a finnish flag and you are soumi land then you see a a pub and a big place known as the k market where you got beer meat drinks fizzy drinks and 18 plus things you have a toy play and knives ushakbas and by a warm desk you can order some kinda ribs witch are amazing full of fat but more or less good again driving there are of course butifull scines like a very round mountin after a while you have a gas

satasion take the right you in Sweden but I am not telling moore. Then when we arrive in piteå at the school I will be staying in we pass time by doing stuff but day to we wke up eat our food we already have all of our clothes on for a match but we eat the go back wait the we go on a bus that is taking us to the pitch we where playing on a real grass pitch that was a first one the match begins after I was subbed on I start to get the fire of the moves I had dribbling everybody I saw but it was very hot the game end we loose but I was good and am glad and not glad.

For task number two, the candidate submitted a 453-word submission, in which we identified a total of 9.7 spelling errors related to speech sounds per hundred words (44 tokens). The phonological spelling strategy is used also in this text, with spellings such as "scines" and "packt". There are also examples of words being spelled according to Norwegian orthography, such as "teksts" and "senter", indicating that the spell and grammar checker has not been successful in assisting the pupil in correcting these errors. Additionally, many of the spelling errors represent a lack of double consonants. For instance, the pupil has written "stoped", "shoping" and "mattres" rather than the correct "stopped", "shopping" and "mattress". Failure to use double consonant was not found in the text submitted for task one, making these errors new in task two. The lack of capitalisation of proper nouns also remains uncorrected in this submission, with 3.1 errors per hundred words (14 tokens). This is an increase of 181.8% from task one. In contrast, the pupil has a decrease in apostrophe errors from task one to task two, as they did not make any errors in the given category in task two. However, as the pupil has not used any contractions in the second submission, this could be a result of the pupil not feeling confident about using apostrophes.

When counting grammatical errors, we found that candidate number 25 has made 4.0 verb-related grammatical errors (18 tokens) and 0.9 other grammatical errors per hundred words (4 tokens) in task number two. The numbers of errors have therefore increased in both categories from task one to task two. Also, this text contains present tense verb forms like "is talking" and forms which are either base forms or present tense forms, such as "arrive" and "pass", in a past tense context. Among the four instances of other grammatical errors, three involve the incorrect use of determiners ("like sleeping bag", "got deodorant, "after a 30 min"), while the fourth involves the lack of the adverbial suffix -ly in "amazing full".

Candidate number 25's responses to the questionnaire are somewhat comparable to candidate number 12's responses. When asked whether they normally use the grammar checker

accessible in Microsoft Word, candidate number 25 answered "yes". However, we do not know whether the pupil regards the grammar checker being activated as synonymous with using it. To question number two, where they were asked whether they used another grammar control program than the one accessible in Microsoft Word, the pupil answered "yes, not Grammarly but Opera GX". As Opara GX is a browsing program for gamers, which does not seem to have a built-in grammar control program, candidate number 25's response to this question could indicate that the pupil does not have a comprehensive understanding of what a grammar control program is. In response to question three, where they were asked if they understand how to use the grammar control programs, candidate number 25 answered "a little". This answer could suggest that the candidate is familiar with some of the features of a grammar control program, while not being fully confident that they can use it competently. When asked whether they read the descriptions following the red or blue lines that appear when the spell and grammar checker suggests changes, the candidate responded "no". This aligns with the results when comparing the two texts, as the pupil seems to have ignored the recommended changes frequently. Although we do not know for sure what changes the program recommended, one can assume that some of the errors made in task two would have been attempted corrected by the spell and grammar checker, as suggested changes appeared multiple times when copy pasting the text into our version of Microsoft Word.

5 Discussion

In this chapter, the results presented above are discussed in relation to our research questions and the theory and research presented in chapter 2. As explained in chapter 1.2, our research questions are the following:

To what extent, and how, does the use of the spell and grammar checker in Microsoft Word affect spelling and grammar in texts written by pupils in Norwegian schools aged 13-14 years?

RQ1) To what extent do we identify differences in error rates in texts written by pupils with and without the spell and grammar checker activated?

RQ2) To what extent do the participants in our study claim to use, and understand how to use, the spell and grammar checking software in Microsoft Word?

We focus on RQ1 in chapter 5.1 and on RQ2 in chapter 5.2. Then, in chapter 5.3, we discuss some of the implications of our data.

5.1 The effect of the spell and grammar checker on the pupils written submission

In this subchapter, we discuss to what extent the data presented in chapters 4.2 and 4.4 enables us to answer research question one. In our study, we found that activating the spell and grammar checker resulted in a statistically significant decrease in overall error rates. The error rates decreased by 43.18% in the submissions where the pupils had activated the spell and grammar checker, from a mean of 8.8 errors in task one to 5.0 errors in task two. Our findings are consistent with the findings of Rynning-Tønnesen (2023, p. 64), who found a significant decrease in error rates when learners were allowed to write in Microsoft Word rather than by hand. In this subchapter, we discuss the difference in error rates regarding both spelling and grammar.

5.1.1 A significant decrease in spelling errors

The category which had the most statistically significant decrease in error rates from task one to task two, is spelling errors. We found a 51.6% decline from task one to task two, indicating that the spell checker is effective in correcting spelling errors in pupils' texts. This is in agreement with the main findings of Rynning-Tønnesen (2023), who found that the pupils'

spelling errors decreased from a mean number of 4.29 to a mean number of 0.80 errors per hundred words. Similarly, Wedge (2021) found the percentage of misspelled words to be higher when the pupils wrote by hand in CORYL (with 3.0%) compared to the control corpus where the participants used Microsoft Word (with 0.05%). The findings of Rynning-Tønnesen (2023) and Wedge (2021) strengthen the external validity of our results, as the similarity of our conclusions demonstrates consistency across different contexts and populations. This could suggest that the decrease in spelling errors when using the spell checker in Microsoft Word is generalisable beyond our study.

We found that the spelling errors that were made most frequently, in both tasks, were spelling errors related to speech sound. In our study, the pupils made an average of 4.3 errors per hundred words in this category for task one. This is almost identical to the results of Rynning-Tønnesen (2023, p. 54), where the pupils made an average of 4.29 errors per hundred words in the category "misspelled words" (cf. chapter 2.4.4). Our results demonstrate a less significant decrease in errors when the pupils activated the spell checker than the difference demonstrated by Rynning-Tønnesen (2023, p. 54), with a mean of 2.0 errors per hundred words compared to 0.80. Our results could differ due to non-identical categories or a difference in research design. However, this is not possible to state with certainty.

In chapter 4.4, we delve deeper into errors by discussing four specific cases, each addressing error types in more detail compared to chapter 4.3. In these cases, we found that many of the errors involve phonological spellings and spellings which are in accordance with the Norwegian rather than the English spelling system. For instance, the (non-)use of double consonants, the spelling of double vowels, distinguishing between homonyms, and non-use of silent letters were spelling errors made frequently. Moreover, issues concerning the use of capital letters in proper nouns extend beyond nouns treated as proper nouns in English but not in Norwegian, such as "Monday" and "Christmas Eve". For some learners it even seems hard to learn that "I" is spelt with a capital. As we found a statistically significant difference in capitalisation errors from task one to task two (p=.003), one can argue that the spell checker inhibits these spelling errors from surfacing, consequently restricting the potential for pupils to acquire the English rules of capitalisation as the errors are obscured from the teacher.

While the decrease in spelling errors from task one to task two is significant when considering the whole dataset, our results demonstrate a clear variation amongst the participants in our study. Some pupils' error rates regarding spelling went up, while others had a significant

decrease. Similarly, Rynning-Tønnesen (2023, p. 79) found that in six of the participants' texts, the total number of spelling errors reduced by 80.74%, from 135 errors in the handwritten submissions to 26 errors in the typewritten submissions. These participants' errors accounted for 50% of all errors. Although these pupils had a significant error reduction, our study shows that some of the pupils do not benefit from activating the spell checker. This variability could be attributed to several factors, including differences in spelling proficiency or familiarity with the spell checker tool.

5.1.2 A significant decrease in grammatical errors

The results of our study demonstrate a statistically significant difference in grammatical errors overall from task one to task two ($p = .037$). We found a 23% decline from task one to task two, indicating that the grammar checker is effective in correcting grammatical errors in pupils' texts. However, the p-value is notably higher compared to the spelling category (where $p = <.001$). Moreover, we found no significant difference when testing the two subcategories of grammatical errors rather than the grammatical category as a whole. For instance, the majority of grammatical errors made in both tasks were verb-related grammatical errors (as was the case in John & Woll, 2020, cf. chapter 2.4.3), yet the difference in error rates was not significant ($p = .32$). Our findings could be a result of Microsoft Word's inadequacies in effectively identifying such errors, as highlighted by John and Woll (2020)'s comparison with other grammar checking programs. As explained in chapter 2.4.3, John and Woll (2020, p. 180) found that Microsoft Word has clear shortcomings in error detection compared to Grammarly and Virtual Tutor.

The results of our case study demonstrate that several of our participants had a high number of errors related to conjugating irregular verbs and demonstrated struggles with the past tense -ed suffix. This is consistent with both McDonald (2008, p. 264) and Li (2012, p. 123)'s findings that regular past tense causes trouble for learners. Although our results indicate that the change in error rates when activating the spell and grammar checker is not statistically significant, the number of errors the pupils made in this category could indicate that the participants struggle with verb tenses. Additionally, some of the errors related to the -ed suffix have been coded as spelling errors, as the form is correct while the spelling is incorrect. As John and Woll (2020, pp. 179- 180) found tense shifts to be one of the seven categories Microsoft Word was unable to spot and give feedback on, and we found errors related to tense

shifts to be frequent, the non-significant difference in verb-related grammatical errors is not unexpected.

Our case study, combined with the figures demonstrated in table 2, indicates that the difference in grammatical errors from task one to task two varies amongst the participants. This could be a result of factors influencing their grammatical performance, for instance their individual grammatical proficiency or familiarity with the grammar control program. The latter is discussed in the next subchapter.

5.2 The pupils' reported knowledge of and experience with using spell and grammar checking software

Our second research question is the following:

RQ2) To what extent do the participants in our study claim to understand how to use the spell and grammar checking software in Microsoft Word?

This research question will be answered through using the results from the questionnaire, and through comparing the pupils' responses with the results from the pupils' written submissions (cf. chapters 4.1 and 4.3).

Through our study, we have gained an insight into our participants' knowledge of and experience with using grammar control programs, and into whether the pupils read the explanations that follow the suggested corrections. Our study found that 90% of the participants claim to normally use the grammar control program in Microsoft Word. Moreover, our results indicate that although 23 of the pupils, equivalent to 76%, claim to understand how to use the grammar control program in Microsoft Word, some of the participants do not read the accompanying descriptions explaining the recommended changes. This is consistent with Cavaleri and Dianati (2016, p. 233)'s findings when testing the effectiveness of Grammarly, which indicated that while the tool is effective for 70.8% of the pupils in the study, some of the pupils who participated found the feedback provided by Grammarly challenging to comprehend (cf. chapter 2.4.1). Additionally, Lin et al. (2017, p. 1501) state that pupils fail to fully utilize the assistance or guidance Microsoft Word offers, as the spell and grammar checker may occasionally misinterpret your intentions and provide the wrong suggestion (cf. chapter 2.5). As 43% of our participants claimed that they do not read the descriptions following suggested changes, and seven of the participants had a higher error rate with the grammar checker activated than when it was deactivated, our results could be

argued to support Lin's suggestion that pupils fail to use the spell and grammar checker effectively.

The high percentage of pupils in our study who claim that they do not read the descriptions, could be a result of insufficient training in using the program. As explained in chapter 2.5, Dypedahl and Hasselgård (2018, p. 170) state that spell and grammar checkers can be useful for pupils if they are used correctly and with sufficient knowledge. They describe grammar checkers as powerful tools, but state that there is a need for enhanced training and guidance to aid pupils in effectively utilizing the feedback provided by spell and grammar checking software. Although the point-biserial correlation test found no correlation between the responses to the questionnaire and the difference in error rates from task one to task two, the results could indicate that being familiar with using the grammar control program benefits the pupils' accuracy. While our study found that the error rates in average decreased by 43.18% when activating the spell and grammar checker, the pupils who did not claim to understand how to use the program only had a 23.6% decrease in error rates.

As addressed in chapter 3.2.1, the validity of our findings regarding the pupils' claimed understanding of grammar control programs is compromised due to reliance on self-reported responses, impacting the construct validity of our research. Many of the pupils who reported that they did not understand how to use the programs benefited from activating them, and some of the pupils who reported that they understood how to use the programs had an increase in error rates from task one to task two. The weakened validity that comes from not being able to fully trust that the respondents answered truthfully, or that they understood the questions, can be one of the reasons that we did not find statistically significant correlations between the results of the questionnaire and the difference in error rates.

5.3 Implications

Our results suggest that the pupils in our study benefited from using spell and grammar checkers, with a high decrease in error rates when the spell and grammar checker was activated compared to when it was deactivated. Our sample of data also illustrate individual differences in the extent to which the pupils benefited from using the spell and grammar checker, and in some cases it is apparent that being unfamiliar with using the tool is unhelpful. In the following subchapters, we will discuss possible implications of our data, both relating to assessment and relating to language learning.

5.3.1 Assessing digital competence rather than language skills

While our study illustrates a clear statistically significant decrease in error rates amongst the pupils when the spell and grammar checker is activated from when it is deactivated, not all pupils had a decrease in error rates. The results of this study demonstrate a clear variation in both pupils' knowledge and experience with using spell and grammar checkers, and in the effects of activating the program. The results suggest that if pupils normally have the spell and grammar checker activated when writing submissions for summative assessment, this may negatively impact the teacher's ability to give a fair assessment of each pupil's language skills and their ability to identify which competence the pupils are being assessed on - their linguistic or digital competence?

As the competence aims involving digital competence explicitly mention the use of digital tools, and as digital skills and writing skills are referred to as two separate types of skills (Utdanningsdirektoratet, 2019), one can argue that the aims directed at writing skills are meant to address competence the pupils should have without relying on digital aids. Our findings indicate that independent language production will not be assessed or practiced when pupils use spell and grammar checkers regularly.

As we do not know the extent to which the teachers rely on texts written in Microsoft Word to assess their pupils' spelling and grammar skills, we cannot state with confidence whether submissions such as those in our sample are routinely used to assess language. Therefore, this discussion is most relevant in cases where texts written with the spell and grammar checker activated are being used to assess spelling and grammar. Using one of the competence aims for after year 10 as an example, that pupils are expected to be able to follow rules for spelling, word inflection, syntax and text structure (Utdanningsdirektoratet, 2019), the texts submitted in task two can be argued to not represent each pupil's ability to follow these rules independently. This is due to the pupils' error rates having significantly decreased from task one, and there being incidences where errors made repeatedly in task one have been corrected in task two. Therefore, assessing pupils' ability to follow these rules through their submissions written with spell and grammar checker activated could be misleading.

To give an example of the implication discussed above, we will use candidates number 1's and number 25's submissions, which were both addressed in chapter 4.4. We found that candidate number 1 had a 90% decrease in error rates from task one to task two, decreasing from 16.8 errors per hundred words to 1.7. In contrast, candidate number 25 increased their

error rates with 25.9%, from 13.5 to 17.0 errors per hundred words. If these pupils had been assessed on their ability to follow rules for spelling and grammar in their submissions for task two, their knowledge on using the program would give candidate number 25 a disadvantage compared to candidate number 1. With the spell and grammar checker deactivated, candidate number 25 would be considered more accurate in their spelling and grammar than candidate number 1. With the spell and grammar checker activated, candidate number 1 would be viewed as significantly more accurate in their spelling and grammar compared to candidate number 25.

With the example given above, the question arises whether teachers can give a fair assessment of the pupils when they submit written texts with the spell and grammar checker activated. As we did not interview the candidates' teachers, we do not know the extent to which they have taught the pupils how to use the grammar control program. As stated by Dypedahl and Hasselgård (2018, pp. 168-170), spell and grammar checkers are helpful when utilized properly, but practical awareness regarding the program's possibilities and limitations must be created in the classroom. Thus, it is important that teachers give their pupils proper training in how to use spell and grammar checkers if they want learners to use them, to ensure that the pupils have equal starting points. Nonetheless, the challenge lies not in evaluating language skills, but in accurately assessing pupils' digital proficiency, regardless of their familiarity with spell and grammar checkers. In the example above, the pupils would be assessed on the competence aim that they are expected to be able to "use different digital resources and other aids in language learning, text creation and interaction" (Utdanningsdirektoratet, 2019), rather than on their ability to follow rules for spelling and grammar.

5.3.2 Implications for formative assessment

While the subchapter above is related to summative assessment, our results could also have implications for formative assessment. The aim of formative assessment is to allow the teacher and the pupil to monitor progress towards achieving the learning objectives (Kunnskapsdepartementet, 2019). While one can argue that the spell and grammar checker can help pupils correct surface-level errors, our study revealed that the tool could also mask gaps in the pupils' understanding of rules for spelling and grammar. For instance, our study found that some of the pupils made spelling errors where they followed Norwegian rules rather than English. As an example, English uses capitals in weekdays and apostrophes in contracted words. The former convention does not exist in Norwegian, and contracted forms

with apostrophes are rare and mostly dialectal (e.g. "Vå'lenga" and "'n Per"). Some of the participants in our study made such errors frequently in task one but not in task two. This could result in the teacher missing valuable insight into the pupils' competence and areas for growth regarding spelling and grammar. Therefore, one could argue that the significant decrease in error rates when activating the spell and grammar checker could limit the opportunities to provide ongoing feedback about areas of improvement, as, according to Munden et al. (2017, p. 122), English teachers must identify what pupils already know in order to help them learn more. Additionally, the question arises of whether teachers can know if a learner has moved from the zone of proximal development to self- regulation if the learner is never assessed on independent work, without digital tools to guide them.

As the goal of formative assessment is to monitor progress, the assessment should be beneficial for the pupils' learning. According to Fenner and Skulstad (2020, p. 225), the pupils should eventually gain what they refer to as procedural knowledge, where grammatical rules become automatised so that the pupils can use their knowledge about these rules in their own language production, which is also one of the competence aims in LK20 (cf. chapter 1.1). This knowledge is fostered through internalising, confirming, and consolidating the grammatical rules. As explained in chapter 2.1, Ellis (2006, p. 97) refers to this process as explicit to implicit knowledge conversion. When pupils depend on the spell and grammar checker in their language production, the conversion from explicit to implicit knowledge can be affected by them not having to use their grammatical knowledge in their own language production. As we have seen, the results from the questionnaire show that although 90% of the participants claim to use the grammar checker, many of the pupils do not read the descriptions following recommended changes. This could suggest that they do not access their knowledge about grammatical rules when they use the spell and grammar checker. Additionally, the grammar checker masking the pupils' knowledge gaps can affect the teacher's possibilities to design tasks that may contribute to developing procedural knowledge. If the goal is for formative assessment to be used to track progress, the teachers need an accurate representation of the pupils' competence to design tasks that will be beneficial towards achieving learning objectives. With the significant difference in error rates found in this study, one can argue that tracking the pupils' progress using texts written with the spell and grammar checker activated could be problematic.

A recognised approach related to formative feedback, is giving corrective feedback. One of the implications our results could have for formative assessment, is that they could limit the teacher's opportunities to give corrective feedback. This will be addressed next.

5.3.3 Limited opportunities to give corrective feedback

The significant decrease in error rates found in this study suggests that using the spell and grammar checker could limit the teachers' ability to give pupils corrective feedback where it is needed. As pointed out in chapter 2.3.1, Bitchener (2008, p. 116) found that corrective feedback of errors significantly improved the accuracy of the participants in his study. One could argue that the spell and grammar checker gives corrective feedback, but the feedback must be considered effective for there to be a purpose in using it. Hernández Puertas (2018, p. 82) found corrective feedback given by a teacher to reduce the number of errors short-term more than feedback given by a computer. As the pupils in our study made a significantly higher number of errors when the spell and grammar checker was deactivated, and 90% of them claimed to normally use it, our results indicate that the feedback they have previously received has had little effect on their accuracy. However, we do not know the extent to which their level of accuracy has improved throughout the years, and the results could also be explained by the pupils getting confused by the correction programs and their metalanguage and explanations.

Although we cannot state that the spell and grammar checker's feedback has no long term effects, Ellis (2009, p. 16) argues that the effect of the feedback is dependent on the learner's needs. As the spell and grammar checker does not consider whether the pupils need direct or indirect corrective feedback, one can argue that the feedback should be given by a teacher who is able to consider the pupils' needs in this respect. The issue then becomes whether it is useful for teachers to give corrective feedback on submissions written with spell and grammar checker activated. In our study, we found that when the spell and grammar checker was activated, many of the errors that were made frequently were corrected, which could result in the pupils' needs not being identified. Therefore, if we were to give corrective feedback on their submissions, we would not know whether the pupils would benefit more from direct or indirect corrective feedback. For instance, in candidate number 1's submissions, task one could indicate that the pupil needs direct corrective feedback as they make a high number of errors, suggesting that the learner is less advanced. In contrast, the reduced number of errors in task number two could mean that the pupil needs indirect corrective feedback, as the

number of errors could indicate that the learner is more advanced. This could suggest that the spell and grammar checker may mislead the teacher to believe that pupils are more advanced learners than they are, which could result in giving ineffective corrective feedback.

Ellis (2009, p. 16)' findings are consistent with Sheen and Ellis (2011, p. 597)' statement that, with a sociocultural view of corrective feedback, the corrections must be based on the individual learners' needs. Our study found that many of these needs cannot be identified with the spell and grammar checker activated. The type of corrective feedback the pupil needs is not only dependent on whether the pupil is a beginner or an advanced learner, but also the types of errors that the pupil needs corrections on. For instance, in their study, Ghandi and Maghsoudi (2014, p. 59) found that indirect corrective feedback is more effective than direct corrective feedback in correcting spelling errors. In our study we found that the spell and grammar checker was most effective in correcting spelling errors, with a 52.70% decrease in error rates from task one to task two. Regarding grammatical errors, there was a 23.45% decrease in error rates. This means that when the pupils use the spell and grammar checker, the corrective feedback they receive comes mainly from the program, as many of the spelling errors have been corrected when the texts are submitted to a teacher. Therefore, the corrective feedback the pupils receive is direct, making it less effective compared to indirect feedback given by a teacher. While our study strongly suggests that the use of spell and grammar checkers may negatively impact teachers' chances to provide fair and helpful summative and formative feedback, this question is conspicuously absent from discussions of the use of digital tools in language teaching.

6 Conclusion

In this thesis, we have investigated the following research questions:

> *To what extent, and how, does the use of the spell and grammar checker in*
> *Microsoft Word affect spelling and grammar in texts written by pupils in*
> *Norwegian schools aged 13-14 years?*

> *RQ1) To what extent do we identify differences in error rates in texts written by*
> *a sample of pupils with and without the spell and grammar checker activated?*

> *RQ2) To what extent do the participants in our study claim to use, and*
> *understand how to use, the spell and grammar checking software in Microsoft*
> *Word?*

The main research question was answered through RQ1 and RQ2. To answer RQ1, we used quantitative data gathered in a within-subject design through pupils' written texts. Each participant wrote two texts, where the first was written with the spell and grammar checker deactivated and the second was written with the feature activated. Our study found that while the grammar checker in Microsoft Word does **not** help pupils reduce the number of verb-related grammatical errors significantly, the software does help the pupils reduce the number of other grammatical errors significantly. Moreover, the research found that the spell checker in Microsoft Word does help pupils reduce the number of spelling errors related to speech sound, the number of capitalisation errors and the number of apostrophe errors significantly. However, the findings concerning non- verb related grammatical errors, apostrophe errors and capitalisation errors must be interpreted with reservation as some of the error rates are relatively low. Notably, the results related to RQ1 demonstrated a clear variation in the error rates and the effect of using the spell and grammar checker amongst the participants in the study.

To answer RQ2, we used data from a questionnaire that was given to the same participants who produced the written texts. The questionnaire included four questions that asked about the pupils' knowledge of, and experience with, using grammar control programs. The study found that 90% of the pupils claimed to use the grammar control program in Microsoft Word, and 77% of the pupils reported that they understand how to use the feature. This means that there is a high reported knowledge and experience with using the grammar control program.

An important part of answering the main research question in this study involves addressing the generalisability of our findings. Determining the generalisability of our results presents a challenge as we used convenience sampling, and we had a relatively small number of participants. However, we found no reason to presume that the pupils included in our study deviate significantly from those in comparable contexts. Moreover, the alignment of our results with previous research, particularly related to spelling, demonstrates consistency beyond the research done in this thesis. Additionally, to answer the main research question we checked for correlations between the results of RQ1 and RQ2. The analysis revealed no correlations when considering the entirety of the sample. However, given that LK20 applies to each pupil within Norwegian educational settings, it is noteworthy that the impact of the spell and grammar checker on error rates exhibited variability, with instances where a correlation was observed between questionnaire responses and error rates.

The results of this study poses challenges for both summative and formative assessment practices, potentially hindering teachers' ability to provide accurate feedback and assess pupils' spelling and grammatical skills effectively. Furthermore, the findings highlight the importance of distinguishing between digital and language competence in educational assessments and the need for enhanced training and guidance in using spell and grammar checkers effectively. Overall, while the spell and grammar checker offers valuable support, careful consideration is required to ensure they complement, rather than replace, effective language learning and assessment practices.

7 Limitations of this study and suggestions for further research

In our study, we have researched the effects using the spell and grammar checker in Microsoft Word have on pupils' spelling and grammar. Although our research revealed some important facts, the study also has limitations. There are multiple aspects in this field that are yet to be explored, but that we have been unable to investigate due to the limited timeframe and resources available to us, the scope of our study and variables we did not consider. These limitations will be discussed in this subchapter, and they are also suggestions for further research that we believe can provide valuable insight in this field.

The first limitation of our study is that we used convenience sampling, which limited the generalisability of our findings. Future research employing random sampling methods could enhance the external validity of the results, allowing for broader generalisations to the wider population. Comparing the outcomes of studies conducted with different sampling techniques could provide a more robust understanding of the effects of spell and grammar checking software on pupils' language skills across diverse contexts.

The second limitation of our study is the design of the questionnaire. The limited scope of the questionnaire hindered our ability to capture the nuances of pupils' experiences with grammar checking software. A more comprehensive questionnaire could have explored factors such as the frequency of the usage of grammar control programs, specific features utilized, and challenges pupils encounter. Additionally, incorporating rating scales in the questionnaire could have allowed for a more nuanced analysis of pupils' perceptions and experiences.

Another limitation is that our study overlooked the crucial aspect of teachers' instructional practices regarding grammar control programs. Understanding how teachers integrate these tools into their curriculum and provide guidance to pupils could provide valuable insights into their impact on learning outcomes. Similarly, investigating whether teachers consider the use of grammar checkers when assessing pupils' written work could shed light on the influence of spell and grammar checkers on grading standards and feedback.

As a suggestion for further research, conducting longitudinal studies could offer valuable insights into the long-term effects of using grammar checkers on pupils' language proficiency. Tracking pupils' progress over time and examining changes in error rates, writing quality, and performance with spelling and grammar could provide a more comprehensive understanding

of the impact of these tools on language learning outcomes. However, we recognise that such a study is challenging to accomplish as in most countries people are exposed to a massive input of English, making it hard to isolate the spell and grammar checkers' effects. Additionally, integrating quantitative data from questionnaires with qualitative data from interviews or classroom observations would provide a richer understanding of pupils' experiences with grammar checking software and the extent to which they are used in the classroom. Qualitative insights could give insight into contextual factors influencing pupils' usage patterns, perceptions, and learning outcomes, which could complement the quantitative findings.

The spell and grammar checker in Microsoft Word is one of many digital tools available for pupils in Norwegian schools. With the constant digital development, research must be done on both the effects of using new tools and on how they are being used. Our research is limited to only one digital aid, thereby not covering the range of tools accessible to learners. With the increasing use of artificial intelligence, more research is needed on how digital aids are currently being used, how they can be used in the future as a supplement that contributes to learning, and on implications of ineffective use of these tools.

Works cited

Bakhurst, D., & Shanker, S. (2001). *Jerome Bruner : language, culture, self*. SAGE.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of second language writing*, *17*(2), 102-118. https://doi.org/10.1016/j.jslw.2007.11.004

Brown, R. (1973). *A first language : the early stages*. Harvard univ. press.

Cavaleri, M., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar check as perceived by students. *Journal of academic language and learning*, *10*(1).

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (Eighth edition. ed.). Routledge.

Creswell, J. W., Creswell, J. W., & Guetterman, T. C. (2021). *Educational research : planning, conducting, and evaluating quantitative and qualitative research* (Sixth edition. ed.). Pearson Education Limited.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.

DeKeyser, R. (1998). Focus on form in classroom second languag. In.

Dypedahl, M., & Hasselgård, H. (2018). *Introducing English grammar* (3rd ed.). Fagbokforl.

Dysleksi Norge. (2021). *Dysleksivennlig skole*. Retrieved 11. oktober. 2023 from https://dysleksivennlig.no/for-ledelsen/systemer-for-kartlegging/

Ellis, R. (2006). Current Issues in the Teaching of Grammar: An SLA Perspective.

Ellis, R. (2009). Corrective Feedback and Teacher Development. *L2 Journal*, *1*(1), 3-18. https://doi.org/10.5070/L2.V1I1.9054

Fenner, A.-B., & Skulstad, A. S. (2020). *Teaching English in the 21st century : central issues in English didactics* (2nd edition. ed.). Fagbokforlaget.

Flognfeldt, M. E., & Lund, R. E. (2016). *English for teachers and learners : vocabulary, grammar, pronunciation, varieties*. Cappelen Damm akademisk.

Ghandi, M., & Maghsoudi, M. (2014). The Effect of Direct and Indirect Corrective Feedback on Iranian EFL Learners' Spelling Errors. *English language teaching (Toronto)*, *7*(8), 53. https://doi.org/10.5539/elt.v7n8p53

Gleiss, M. S., & Sæther, E. (2021). *Forskningsmetode for lærerstudenter : å utvikle ny kunnskap i forskning og praksis* (1. utgave. ed.). Cappelen Damm akademisk.

Hammersley, M. (2013). *What is qualitative research?* Bloomsbury Academic.

Harmer, J. (2007). *The Practice of English language teaching : with DVD* (4th edition. ed.). Longman.

Hernández Puertas, T. (2018). Teacher's feedback vs. computer-generated feedback: A focus on articles. *Language value*, *10*, 67-88.

Hewa Fouad, A., Nakshbandi, L. J., Saadi, F., & Sami Hussein Hakeem, B. (2022). The Effect of Spell-Checker Features on Spelling Competence among EFL Learners: An Empirical Study. *International journal of social sciences & educational studies*, *9*(3), 101-111. https://doi.org/10.23918/ijsses.v9i3p101

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Lang. Teach*, *39*(2), 83-101. https://doi.org/10.1017/S0261444806003399

Imsen, G. (2014). *Elevens verden : innføring i pedagogisk psykologi* (5. utg. ed.). Universitetsforl.

John, P., & Woll, N. (2020). Using Grammar Checkers in an ESL Context: An Investigation of Automatic Corrective Feedback. *CALICO journal*, *37*(2), 169-196. https://doi.org/10.1558/cj.36523

Kunnskapsdepartementet. (2019). Curriculum in English. https://www.udir.no/lk20/eng01-04/kompetansemaal-og-vurdering/kv3?lang=eng

Li, M.-C. (2012). *The acquisition of tense and agreement by early child second language learners* ProQuest Dissertations Publishing].

Lin, P.-H., Liu, T.-C., & Paas, F. (2017). Effects of spell checkers on English as a second language students' incidental spelling learning: a cognitive load perspective. *Reading & writing*, *30*(7), 1501-1525. https://doi.org/10.1007/s11145-017-9734-4

Lira-Gonzales, M.-L., & Nassaji, H. (2020). The Amount and Usefulness of Written Corrective Feedback Across Different Educational Contexts and Levels. *TESL Canada journal*, *37*(2), 1-22. https://doi.org/10.18806/tesl.v37i2.1333

Lærd statistics. (2018). *Point-Biserial Correlation Using SPSS Statistics*. Retrieved 15.04.2024 from https://statistics.laerd.com/spss-tutorials/point-biserial-correlation-using-spss-statistics.php

Mackey, A., & Gass, S. M. (2015). *Second Language Research* (2 ed.). Routledge. https://doi.org/10.4324/9781315750606

McDonald, J. L. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *J. Child Lang*, *35*(2), 247-268. https://doi.org/10.1017/S0305000907008367

Munden, J., Sandhaug, C., & Keeping, D. (2017). *Engelsk for secondary school*. Gyldendal akademisk.

NVivo. (1997). *Computer software*. https://help-nv.qsrinternational.com/20/win/Content/welcome.htm

Oxford, U. o. (2014). *Style guide*. https://www.ox.ac.uk/sites/files/oxford/media_wysiwyg/University%20of%20Oxford%20Style%20Guide.pdf

Rynning-Tønnesen, A. (2023). *A comparative case study of 8th graders' spelling and grammar using handwritten versus computer-written texts with autocorrect/ Grammar-checking-software* Høgskulen på Vestlandet].

Sheen, Y., & Ellis, R. (2011). Corrective Feedback in Language Teaching. In.

Spada, N. (2011). Beyond form-focused instruction: Reflections on past, present and future research. *Language teaching*, *44*(2), 225-236. https://doi.org/10.1017/S0261444810000224

Truscott, J. (1996). The Case Against Grammar Correction in L2 Writing Classes. *Language learning*, *46*(2), 327-369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x

Utdanningsdirektoratet. (2019). *Competence aims and assessment*. Retrieved 2, Oktober, 2023 from https://www.udir.no/lk20/eng01-04/kompetansemaal-og-vurdering/kv3?lang=eng

Vygotskiĭ, L. S., & Cole, M. (1978). *Mind in society : the development of higher psychological processes*. Harvard University Press.

Wedge, A. Ø. (2021). *Wrighting in Englesh is harde: An error analysis of Norwegian pupils' L2 English spelling errors* The University of Bergen].

Appendix 1 - Correspondence with SIKT

Meldeskjema / A comperative study of students´ gram… / Vurdering

Vurdering av behandling av personopplysninger

🖶 Skriv ut   ⅀ 13.10.2023 ▾

Referansenummer	Vurderingstype	Dato
564386	Standard	13.10.2023

Tittel
A comperative study of students´ grammatical abilities in English, with and without using grammar-checking software

Behandlingsansvarlig institusjon
UiT Norges Arktiske Universitet / Fakultet for humaniora, samfunnsvitenskap og lærerutdanning / Institutt for lærerutdanning og pedagogikk

Prosjektansvarlig
Kristin Killie

Student
Kathrine Kristiansen

Prosjektperiode
04.10.2023 - 15.05.2024

Kategorier personopplysninger
Alminnelige

Lovlig grunnlag
Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 15.05.2024.

Meldeskjema 🗗

Kommentar
OM VURDERINGEN
Sikt har en avtale med institusjonen du forsker eller studerer ved. Denne avtalen innebærer at vi skal gi deg råd slik at behandlingen av personopplysninger i prosjektet ditt er lovlig etter personvernregelverket.

FORELDRE SAMTYKKER FOR BARN
Prosjektet vil innhente samtykke fra foresatte til behandlingen av personopplysninger om barn under 16 år.

FØLG DIN INSTITUSJONS RETNINGSLINJER
Vi har vurdert at du har lovlig grunnlag til å behandle personopplysningene, men husk at det er institusjonen du er ansatt/student ved som avgjør hvilke databehandlere du kan bruke og hvordan du må lagre og sikre data i ditt prosjekt. Husk å bruke leverandører som din institusjon har avtale med (f.eks. ved skylagring, nettspørreskjema, videosamtale el.).

Behandling av personopplysninger på private enheter må skje i samsvar med behandlingsansvarlig institusjon sine retningslinjer/lagringsguider.

Personverntjenester legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til oss ved å oppdatere meldeskjemaet. Se våre nettsider om hvilke endringer du må melde: https://sikt.no/melde-endringar-i-meldeskjema

OPPFØLGING AV PROSJEKTET
Vi vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Appendix 2 – Consent form

Vil du delta i forskningsprosjektet

"A comperative study of students´ grammatical abilities in English, with and without using grammar-checking software"?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å identifisere grammatiske forskjeller i engelsk, i tekster elever skriver med og uten tilgang til autokorrektur. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål
Formålet med denne forskningen er en masteroppgave for ilp – institutt for lærerutdanning og pedagogikk ved UiT. Vi ønsker å få et innblikk av elevers tilgang til digitale verktøy hvor de har tilgjengelig autokorrektur, og hvordan dette påvirker deres grammatiske ferdigheter.
Spesielt stiller vi spørsmål ved hvor nøyaktig lærere kan måle elevenes grammatiske ferdigheter ved bruk av digitale skriveverktøy med autokorrekturfunksjoner. Dette er avgjørende fordi vurdering skal identifisere elevenes utfordringer i akademisk arbeid og synliggjøre områder som trenger ytterligere forbedring. Derfor ønsker vi å undersøke hvilket innblikk lærere får i elevs grammatiske ferdigheter når elevene leverer tekster i programvarer som har tilgang til autokorrektur. Da masteroppgaven skrives på engelsk, er dette de foreløpige forskningsspørsmålene vi ønsker å finne svar på:

To what extent does students' use of grammar checkers in written assignments affect the ability of teachers of English to provide feedback on (students') grammar?
- Do we identify grammatical errors in texts written by hand that we do not identify in texts written digitally with access to autocorrect?
- Are the students´ digital submissions representative of their grammatical skills?

Hvem er ansvarlig for forskningsprosjektet?
Det er vi, Nadia Sigvaldsen og Kathrine Kristiansen som er ansvarlig for å gjennomføre dette prosjektet. Dette gjør vi med god hjelp fra veileder og prosjektansvarlig Kristin Kille som er professor i Engelsk ved UiT Norges arktiske universitet.

Hvorfor får du spørsmål om å delta?
Utvalget til denne undersøkelsen har vi valgt gjennom et bekvemmelighetsutvalg. Dette bekvemmelighetsutvalget baseres på bekjentskap til lærere slik at vi kan komme i deres klasser og samle inn data fra deres elever. Lærerne vi tar kontakt med er tidligere praksislærere og bekjente. Utvalget har vi allerede kontaktet, og vi ønsker å benytte oss av tre kontakter som jobber på forskjellige skoler slik at vi får ulik data til dette forskningsprosjektet.

Hva innebærer det for deg å delta?
Dersom du velger å delta i dette prosjektet vil du få et spørreskjema med 3-5 spørsmål om bruken av grammatiske hjelpemidler på pc, og du gjennomfører to oppgaver på engelsk som du ikke vil bli vurdert på i skolen. Begge oppgavene blir skrevet i et skriveprogram på pc, men kun en av oppgavene skal ha autokorrektur påskrudd (Engelsk Storbritannia). Disse to oppgavene er basert på at du skal

skrive om egne erfaringer/opplevelser som vil ta ca. 1 time totalt, så rundt 30 minutter på hver oppgave. Hovedfokus i dette forskningsprosjektet er grammatikk, og vi vil dermed ikke fokusere på faglig innhold.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Ved å delta på dette prosjektet vil det ikke påvirke din vurdering i engelskfaget eller på skolen. Det er kun masterstudentene som skal se ditt arbeid, ikke din faglærer.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Som nevnt tidligere er det kun forskningsansvarlig som skal lese og gå gjennom ditt arbeid. Din lærer skal anonymisere deg, slik at vi ikke har mulighet å vite hvem som har levert hva, dette blir gjort gjennom et individuelt kandidatnummer som kontakt lærer deler ut. Arbeidet som blir gjort markeres med et kandidatnummer slik at det er anonymisert, slik at ditt personvern blir ivaretatt.

Hva skjer med personopplysningene dine når forskningsprosjektet avsluttes?
Prosjektet vil etter planen avsluttes i uke 20, rundt 15. mai. Etter prosjektslutt vil all anonymisert data bli slettet og ingen andre vil ha tilgang til ditt arbeid. Denne masteroppgaven vil, dersom bestått, bli publisert. Dette innebærer ikke at personopplysninger blir utgitt, da bidrag vil være anonymisert gjennom kandidatnummer.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.
På oppdrag fra Ilp - institutt for lærerutdanning og pedagogikk har Sikt – Kunnskapssektorens tjenesteleverandør vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

Masterstudenter:
Nadia Sigvaldsen   473 49 880   nsi091@uit.no
Kathrine Kristiansen   450 99 806   kkr091@uit.no

Ansvarlig for forskingsprosjektet:
Kristin Killie   776 60 467   kristin.killie@uit.no

Vårt personvernombud:
Hvis du har spørsmål knyttet til vurderingen som er gjort av personverntjenestene fra Sikt, kan du ta kontakt via:

- Epost: personverntjenester@sikt.no eller telefon: 73 98 40 40.

Med vennlig hilsen

Prosjektansvarlig   *Eventuelt student*
(Forsker/veileder)

Kristin Killie

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «A comperative study of students´ grammatical abilities in English, with and without using grammar-checking software», og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i spørreskjema
- ☐ å delta i beskrevet forskningsprosjekt
- ☐ at mitt bidrag kan publiseres uten at jeg kan gjenkjennes (gjennom anonymisering)

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

(Signert av prosjektdeltaker, dato)

Appendix 3 - Questionnaire

Spørreskjema til forskningsstudiet

«A comperative study of students´ grammatical abilities in English with and without using grammar-checking software".

Svar ja/nei på spørsmålene under, eller spesifiser/begrunn kort.

1) Bruker du normalt grammatikk-kontrollprogrammet (autokorrekturen) som ligger inne i tekstbehandlingsprogrammet (Word)?

Svar:_____

2) Bruker du et annen grammatikk-kontrollprogram (autokorrektur) enn det som er i Word, f.eks. Grammarly?

Svar: _____

3) Skjønner du hvordan man bruker grammatikk-kontrollprogrammene?

Svar: _____

4) Leser du beskrivelsene for hvorfor programmet anbefaler endringer (beskrivelsene som kommer før forslag til løsning ved rød eller blå linje under ord eller setninger)?

Svar: _____

Appendix 4 – Task one and task two

Task 1

Write a text in English about the best school trip you have had. Include as many details as possible.

(Skriv en tekst på engelsk om den beste turen du har hatt på din skole. Inkluder så mange detaljer som mulig).

Task 2

Write a text in English about your best memory from this year. Include as many details as possible.

(Skriv en tekst på engelsk om ditt beste minne fra dette året. Inkluder så mange detaljer som mulig).