

# Computer-Based Versus Pen-Based Testing

An Experimental Study of Reading Comprehension and  
Essay Writing in Swedish Lower Secondary School

Jakob Skoglund Olsson  
Ämneslärarprogrammet



Degree essay: 15hp  
Course: L9EN2A

Level: Advanced level  
Term/year: VT/2023  
Supervisor: Monika Mondor

Examiner: Anna-Lena Fredriksson

---

Keywords: Computer-based testing, Pen-based testing, Reading  
Comprehension, Writing skills, EFL

## Abstract

The Swedish government has made significant efforts to introduce digitalization in schools, including national tests. However, research has highlighted negative effects of digital testing on reading comprehension and understanding. Adding to this, recent research presented by Skolverket (2023), shows that, during this period, a concerning decline in reading comprehension test scores among Swedish students has occurred, with no mention of digitalization as a possible explanation. Based on these findings and the existing knowledge gap regarding the impact of digitalization, this study aims to investigate the differences between computer-based and pen-based tests in terms of student performance and preferences. The study was conducted at a highly digitalized lower secondary school in Sweden where two different classes participated in two tests: one on the computer and one using pen and paper. The tests were analyzed and compared by using average scores, the Flesch Kincaid readability test, LexTutor, as well as a manual analysis. Generally, the students performed better on the computer-based test, except for girls who produced slightly more complex texts with pen and paper. Additionally, a questionnaire revealed that most students preferred using a computer in test situations. However, the results may have been impacted by small group sizes and the high use of computers at this school. Furthermore, long-term research and a larger sample is necessary to examine the impact of computers on learning and to identify the best test format.

# Table of Contents

<b>1</b>	<b>Introduction</b> .....	<b>1</b>
<b>2</b>	<b>Literature Review</b> .....	<b>2</b>
<b>3</b>	<b>Method</b> .....	<b>4</b>
<b>3.1</b>	<b>Data Collection</b> .....	<b>5</b>
3.1.1	Quasi-Experiment .....	5
3.1.2	Participants: Characteristics and Inclusion Criteria .....	5
3.1.3	Material and Software .....	6
3.1.4	Reliability and Validity .....	7
<b>3.2</b>	<b>Data Analysis</b> .....	<b>8</b>
3.2.1	Flesch Kincaid Readability Test .....	8
3.2.2	LexTutor – Vocabulary Profile.....	9
<b>3.3</b>	<b>Questionnaire</b> .....	<b>10</b>
<b>3.4</b>	<b>Limitations</b> .....	<b>10</b>
<b>4</b>	<b>Results</b> .....	<b>10</b>
4.1	Analysis – Reading Comprehension .....	11
4.2	Analysis – Written Essay – Flesch Kincaid .....	13
4.3	Analysis – Written Essay – Gender Comparison .....	15
4.4	Analysis – Written Essay – LexTutor.....	17
4.5	Analysis – Written Essays - Gender Comparison .....	18
4.6	Questionnaire .....	20
4.7	Manual Analysis.....	22
<b>5</b>	<b>Discussion</b> .....	<b>23</b>
<b>6</b>	<b>Conclusion</b> .....	<b>25</b>
	<b>References</b> .....	<b>27</b>
	<b>Appendices</b> .....	<b>30</b>
	Appendix A - Negative Consent Letter .....	30
	Appendix B - Questionnaire.....	31

# 1 Introduction

Computers have revolutionized the world in many ways, not least, education. When computers became available to the public, schools in Sweden were quick to adapt and utilize the benefits which computers offer. Promises of a more cost-efficient school as well as an increase in the quality of education leading to the possibility of new jobs led the way to the development of the digitalized school and has not slowed down since (Skolverket, 2023a). Adding to this, Skolverket (2023a), *The Swedish National Agency for Education*, are now in the process of digitizing their national tests instead of running them the traditional way with pen and paper. The reason for doing this is that the Swedish government wants to create a centralized assessment, which is thought to ensure reliability and a more effective assessment (Regeringsbeslut, U2017/03739/GV).

Although digitization has many advantages, the emergence of ChatGPT, a new artificial intelligence tool capable of generating natural and fluid essays, has raised concerns about whether the traditional pen-and-paper method needs to be utilized more than is the case today. The development has made it increasingly challenging for teachers to distinguish between texts written by students and those generated by tools such as ChatGPT, potentially enabling students to cheat on assignments. As a result, it has become important to conduct a study to determine whether the use of a pen-and-paper approach in school can match or surpass the efficacy of the computer, given the risk of academic dishonesty.

With regard to the aspects mentioned above, the decision to digitize the Swedish school is also noteworthy, especially since previous studies (Mangen et al., 2013; Van Der Velden, 2021) have indicated that a pen-and-paper-based education has positive effects regarding a deeper understanding of what is being taught, as well as students scoring better on reading comprehension tests. Given these findings, one might question the empirical basis that Skolverket relies on to justify the decision of digitizing the Swedish school.

Adding to this, more recent research presented by Skolverket (2023b), shows that Sweden has not been able to reach its highest score (561) in reading literacy since 2001, with the lowest score of 542 in 2011 and the most recent score of 544 in 2021. It should be noted, however, that no country improved its score in the most recent survey. Several different factors, such as demographic changes as well as socioeconomic differences explain this decline. However, Skolverket (2023b) have not considered digitalization, which has changed

the Swedish school since the early 2000s, to be a part of the decline in results. Furthermore, Skolverket (2023b) also mentions that girls have been outperforming boys since 2001. These two factors make it important to consider and investigate if computers are the best way for students to show their proficiency.

Given the questionable advantages of digitization on language proficiency, this study aims to investigate if digitized tests or non-digitized tests are the best way for a student to show their proficiency and how they perceive the two test formats. To do so, I will examine differences and similarities in student performance in a digital vs. non-digital setting. In addition, this study will look at whether there are any differences between boys and girls regarding their proficiency with a computer or with a pen and paper. If the results point toward students being more proficient using a pen and paper, digitization might be unnecessary and should be rolled back, and the worries regarding academic dishonesty will disappear. However, if the results indicate that students are more proficient with a computer, questions must still be asked regarding how exams should be constructed to benefit all and how to make digital exams fair and reliable.

## **2 Literature Review**

There are many advantages when it comes to using a computer in test situations compared to pen and paper. Mangen et al. (2015) mention one of them as students being able to write faster and more efficiently compared to using a pen and paper. Because of this, Mangen et al. (2015) argue that using a computer during timed, written tests might be more suitable than a pen-and-paper approach since it enables students to finish their essays faster and create an opportunity to edit what they have written. However, there are also studies which show that a computer might bring some disadvantages regarding concentration. If the computer is not in a “locked mode” where students cannot avoid distractions, the result of their test might be impacted negatively. This is suggested by Muller and Oppenheimer (2014) who found that students who used a laptop when completing a comprehension test received a lower score compared to students who took it by hand. This could be an indication that there might be something on the computer which distracts students from focusing on the task, which has an impact on their results.

Chua and Don (2013) found comparable results to what Mangen (2015) found when conducting a study consisting of undergraduate students who either completed a pen-based or a computer-based test. When the students had finished their tests, they also completed a

questionnaire assessing their motivation in taking the test. The results showed that students who took the computer-based test were able to finish the test faster while also being more motivated compared to students who used the pen-based format. These results suggest that the use of computers in test situations might offer benefits in terms of efficiency and student engagement.

Furthermore, Mueller and Oppenheimer (2014) suggest that employing a pen-and-paper approach allows students to engage with the text in a more hands-on and direct manner. In their study, they found that students who took notes by hand remembered and could recall the information to a higher degree compared to the ones who typed the information on a computer. They mention that the reason for this is that using a pen requires the student to be more selective regarding incoming information, as writing by hand is slower than typing. Mueller and Oppenheimer (2014) also argue that even if typing is faster, a student might end up with something called, “verbatim notetaking” (p.1160) which means that students are only copying what is being taught in class and not processing or understanding the information. This is also supported by Van Der Velden (2021) who also found that students who took notes by hand enabled a “deeper cognitive understanding” (p.15) of what they wrote. Van Der Velden explains the loss of information when typing on a computer due to, “overloading of the cognitive process when information is written with a keyboard, compared with the slow method of handwriting” (p.2). According to this, our brains cannot handle the speed at which our hands are typing which causes a loss of information when transcribing. However, there are not only advantages when using a pen and paper either. Manwaring (2019) mentions some counterarguments such as, “the slowness; the time-consuming transcribing process... and for some writers, the physical strain or impossibility of such an approach” (p. 141) could make a case for the use of computers.

In both Sweden and Norway, schools have transitioned from using physical books to computers when reading in school to a high degree. Therefore, Mangen et al. (2013) set out to investigate if there are any differences in using printed text vs digital versions in a reading comprehension test. The study found that students who used printed text to complete the test scored much better compared to the ones using a computer. The researchers offer various explanations for these findings, such as added cognitive load when switching between windows on a computer, as opposed to flipping pages in a book. Additionally, they mention that screens can cause visual fatigue and interfere with the cognitive process while at the same time creating navigational issues since the spatial cues of the text are not there on a screen.

Due to these findings, the researchers highlight the importance of paying attention to potential consequences when using digital assessment methods.

The research by Seok and DaCosta (2017) on the variations in digital literacy between females and males has paved the way for investigating the differences between boys and girls in this study. In their study, they found that males' "digital propensity was significantly higher than that of females" (Seok & Dacosta, 2017, p.175). While this was the case, they also found that females, on average, held more favorable perceptions towards digital reading compared to males. These findings are explained through Hilton and Canciello's (2015) research which mentions that males tend to use digital technology for fun while females use it for more functional use in their daily life. Due to these differences, it is vital to see how boys and girls perform in test situations using both pen-and-paper- and computer-based tests, since it might have an impact on their results.

Lastly, Backes and Cowan (2019) conducted a large study which was based on the rollout of digital exams in Massachusetts. During the rollout, schools had the opportunity to choose between the digital format as well as the traditional one with pen and paper. The results of their study indicate that students who took the computer-based test performed worse compared to students who used a pen and paper. To see why this was the case, Backles and Cowan (2019) considered students' computer habits. However, their investigation revealed no major differences in performance based on the participants' computer proficiency levels. Furthermore, it should be noted that the second year the digital tests were used, they saw that the negative effects of the online format decreased suggesting the development of computer proficiency among participants. However, despite these improvements over time, a notable negative impact on scores due to online testing persisted even in the second year.

### **3 Method**

In this section, the method of the study conducted will be described. The section will be divided into four different topics, data collection, data analysis, questionnaire, and limitations. The different topics will include a description of the quasi-experimental design, the characteristics and inclusion criteria of participants, the material and software used, validity and reliability, analysis of the result, the Flesch Kincaid readability test, LexTutor (n.d.), the questionnaire as well as limitations. The analysis is mostly quantitative.

## 3.1 Data Collection

In this section, several components are included: an overview of the quasi-experimental design; a comprehensive discussion of the participant characteristics along with the inclusion criteria; a description of the specific materials and software utilized for data collection; and, finally, an account of the measures implemented to enhance the validity and reliability, thereby securing the robustness of the data gathered.

### 3.1.1 Quasi-Experiment

In this study, two classes of students took part in two different tests, thus making it possible for them to serve as their own control group. This correlates with a quasi-experimental design (Bryman et al., 2021). The reason for this approach is that the design offers validity and reduces selection bias, as each student experiences both test formats. This not only enhances the robustness of the findings but also helps to neutralize the potential impact of individual differences (Bryman et al., 2021). However, Bryman et al. (2021) also mention that there could be potential drawbacks to this type of design. One must carefully consider whether observed differences are due to the manipulation of the independent variable - in this case, the distinct tests - or whether uncontrolled confounding variables could influence the results.

### 3.1.2 Participants: Characteristics and Inclusion Criteria

The study was conducted at a secondary school in southwest Sweden, involving two ninth-grade classes of 15 to 16-year-old students. One of the classes (hereafter referred to as class A) took two reading comprehension tests, one conducted on the computer and the other with pen and paper. The other class (hereafter referred to as class B) also took two tests using different formats, but instead of a reading comprehension test, they took two written tests where they wrote a short essay for 50 minutes. On one of the tests, students were asked about their gender to enable the possibility to see if there are any differences between these groups. The reason for only asking this once is that it seems logical that the students identify as the same gender on both occasions. Furthermore, initially, 22 students from Class A and 23 from Class B participated in the experiment. However, exclusions were made for students who only participated in one test (making their results non-comparable), those caught cheating, or those who opted out of the study. Seven students from Class A and nine from Class B were excluded, with the primary reason being participation in only one test. Consequently, Class A was left with 15 students: 10 boys and 5 girls. Similarly, Class B comprised 14 students, with



an even gender distribution of 7 boys and 7 girls. Lastly, while all of the students were familiar with writing with a pen and typing with a keyboard, it should be noted that the school provides the students with a computer that they bring to all classes, which might give an advantage to the computer in this study.

For students to be able to participate in the study, principals, teachers, parents, and students were informed about the study and what it aimed to investigate. Hence, the study follows the guidelines mentioned by Vetenskapsrådet (2002): “The researcher shall inform the data provider and the research participant about their role in the project and the conditions that apply to their participation.” (p.7, my translation). Since the study consists of participants who are not adults, a negative consent letter (see Appendix A) was created and sent out to all participants and parents to give them the opportunity to exclude themselves or their child from the study. If the parents or the students themselves had not said anything before the day the study started it was interpreted as them giving their passive consent. In addition to the negative consent letter, the participants were also informed that they could retract their submission at any time, even if it had already been handed in. Finally, to prevent any misunderstandings, the negative consent letter was written in Swedish.

### 3.1.3 Material and Software for Collecting Data

The material and software used for this study were pens, paper, old national tests, computers, Google Forms (Google LLC, n.d.), and Exam.net (Exam.net, n.d.). The two classes which participated in the study completed two tests each focused on either reading or writing. The first reading comprehension test was created in Google Forms (Google LLC, n.d) which made it possible for the students to submit their answers using a computer. The second reading comprehension test was printed out on paper to make it possible for students to read everything and submit their answers on paper. The construction of the written test was similar, except that it was administered through a website called Exam.net. (Exam.net, n.d.) This website was used since it is possible to lock the computer on that website when the students are doing the test. This stops students from using text-generating tools, such as ChatGPT, during the test which could corrupt the results. The website also makes it possible to decide if tools like autocorrect can be used or not. In this study, it has not been allowed since the results would not be comparable if the computer format allows something which is not possible to utilize when using a pen and paper.

Furthermore, the goal was to find two different reading comprehension tests as well as two different written essay assignments, each pair bearing similarities yet not being identical.

This was to ensure that the results collected could be validly compared. As a result, old national tests, developed by Skolverket in Sweden, were utilized since these tests contribute to both the reliability and validity of the study (for more details, refer to section 3.1.4). These are no longer used or confidential (Skolverket, 2022). The reading comprehension tests, in both formats, included a text and either multiple-choice questions or short-answer questions related to the text. The essay instructions on both test formats involved a future-related topic, allowing students to get creative with their writing. The tests didn't require any specific subject knowledge from the students; only their reading and writing skills.

### 3.1.4 Reliability and Validity

Several measures were taken to increase reliability and validity. First, to make sure that the results collected were reliable, the students were given a number by their mentor to anonymize their results. This ensured that my assessment and analysis of the results is not biased, while at the same time ensuring the possibility to analyze and compare the results one student produced.

Second, to ensure the validity and reliability of the collected material, old national tests, developed by Skolverket in 2007 and 2013, in Sweden, were used. These tests are no longer protected by the privacy act, which makes it possible to use them in this study (Skolverket, 2022). The utilization of the outdated national tests allows for the possibility of using the assessment guides which are included with the tests upon their distribution. This makes the assessment of the students' results more valid since the tests are developed by people who are proficient in creating and assessing these tests every year. Skolverket (2016) mentions that these tests should be used to calibrate a teacher's assessment of a student's performance, which increases the validity of my assessment of the student's results. Furthermore, Barmark and Djurfeldt (2020) mention that "the fundamental principle is that a measurement exhibits a high degree of reliability when it consistently produces the same result during repeated measurements." (p.47, my translation). Hopefully, this is what is going to be achieved by using old national tests.

Another measure taken to ensure a valid result was the consideration of the potential fatigue effect when planning. Lim et al. (2010) explain that the part of the brain that helps us focus does not recover quickly after taking part in a mentally demanding task. This also applies to students who perform more than one test on the same day, arguing that the results from tests that are done in the afternoon might decline since the students are more tired at that time compared to the morning. The decline in a result could be the effect of mental or

physical exhaustion, i.e., the fatigue effect. Because of this, the tests were not conducted on the same day; instead, the students had one or two days off to rest before the next test was done. Additionally, it should be noted that class A and class B took the tests in different sequences. Specifically, class A completed the computer-based test first and then the pen-and-paper-based test, while class B followed the reverse order. This was done to see if the different format had an impact depending on when it was used.

## **3.2 Data analysis**

This section describes two approaches used for the analysis of the data. To assess the students' results on the reading comprehension test, the grading followed the guidelines given with the old national tests. For the evaluation of the written essays, a digital analysis was employed. Class A did two reading comprehension tests which resulted in two different scores based on the number of correct answers on their tests. These scores were later converted into an average score of the group to see if there were any differences between the two tests. Class B participated in two tests which gives the students a chance to show their proficiency in writing an essay. These two batches of essays have been analyzed by using different digital tools such as the Flesch Kincaid readability test which can be found in Microsoft Word software (Microsoft Corporation, 2023), and LexTutor (n.d.) which is available online. The first one mentioned is a readability score system that is used to determine the complexity and difficulty of a written test, while the second one is an online tool that aims to aid learners, educators, and researchers in working with vocabulary learning and analysis of texts. The different texts students wrote were entered into these digital tools to see if there are any differences in writing a text on the computer versus using a pen and paper. The texts which were produced by hand have manually been transferred by me into a Word document retaining all misspellings and other mistakes made by students, to enable the possibility of using digital tools to analyze the text. In addition to this, using face validity as a guiding principle, a manual analysis was undertaken to supplement the Flesch-Kincaid readability test by addressing certain aspects that may not be captured by the test's measures alone.

### **3.2.1 Flesch Kincaid Readability Test**

The analysis relies on three measures from the Flesch Kincaid readability test. These indicators include word count, Flesch Kincaid Grade Level (hereafter as FKGL), and Flesch Kincaid Reading Ease (hereafter as FKRE). The FKGL measures the approximate school

years of education in the U.S. needed to understand a particular piece of text. If a text consists of words with more syllables and longer sentences, which are generally more difficult to read and comprehend than shorter ones, the FKGL score will increase. This also applies to the FKRE score which rates the text on a scale from 0-100, with higher scores indicating an easier text to read. The scores are calculated by two different formulas which analyze factors such as the length of a sentence and the average syllables used per word. An FKGL score of 8.0 indicates that the text is readable by an average eighth-grade student, while an FKRE score of 90-100 indicates a text which is easy to read. However, even if the FKGL and the FKRE might point towards a difficult text to read, the scores may not indicate a well-structured text due to the limitations of the test. The reason for this is that the test does not evaluate semantic meaning or coherence, which means that it fails to recognize that some sentences can be nonsensical or consist of non-words. (Flesch, 1948). This is important to understand when looking at the results provided in this study since a lower FKRE and a high FKGL score might indicate a high-level text, while the content is gibberish. The results students got on these different measures were later turned into an average for the class, both on the computer-based and the pen-based test, to be able to be compared to each other. Adding to this, an individual perspective was also used to see if any students distinguish themselves compared to the result of the group.

### 3.2.2 LexTutor – Vocabulary Profile

LexTutor (n.d.) is an online tool that can be used on texts or speech to evaluate vocabulary usage. One of the tools included in LexTutor (n.d.) is the so-called Vocabulary Profile which can be used to classify words into different frequency bands on an established word list based on corpus data. In this study, the frequency bands of K1, K2, K3 and off-list words are going to be used due to the students not writing any words beyond K3 usage. The first frequency band represents the 1000 most frequent word families used in English (0-1000), the second one represents the next 1000 words in English (1001-2000) and the last one represents the next 1000 most frequent words used in English (2001-3000). The off-list words frequency band represents words that are not included in the K1, K2, and K3 which can be misspelled words, proper nouns, specialized vocabulary, and low-frequency words (Cobb, n.d.). These results were also converted into an average for the class and compared to each other.

### **3.3 Questionnaire**

After the students had finished their test, they were also given the opportunity to answer a questionnaire (see Appendix B) that contained five questions. Three questions were multiple choice which investigated what format they feel is easiest to use, what format they feel they are most proficient in using, and what format they want to use in the future. For the next two questions, the students got to express, on a scale from one to five, how motivated they feel depending on which format they are using. The result from the questionnaire was turned into a circle chart and bar charts displaying different percentages based on students' answers. In total, 23 students took part in the questionnaire. Finally, it should also be added that the questionnaire was written in Swedish to prevent any misunderstandings.

### **3.4 Limitations**

Despite the valuable insights presented in this study, it is also important to acknowledge its limitations. This is vital due to the emphasis placed by Vetenskapsrådet (2017) on the importance of adhering to good research practices, specifically stating that, "Possible sources of errors should be identified and discussed" (p.25). One limitation which is important to consider is the sample size of the girls in group A. Since it is small, it will be hard to generalize the result to a greater group. Another drawback is the fact that the answers provided in the questionnaire section are not directly correlated with the results regarding the reading comprehension test and the written test. The questionnaire was conducted at a different time which resulted in some students not being there to provide their answers. Lastly, it is also important to acknowledge that the results provided by class B have been impacted due to some students' handwriting being hard to read. This resulted in gaps in the text which might have given an advantage to the computer-based test since the letters are displayed flawlessly.

## **4 Results**

In this section, the result will be presented. First, the result regarding the reading comprehension tests will be presented on a group level, followed by differences between boys and girls. Secondly, the result regarding the written test will be presented in a similar way, where the group and the results broken down by gender will be accounted for, analyzed, and compared with the average word count, FKGL, FKRE, and LexTutor (n.d.). Furthermore, the

results from the questionnaire will be presented and analyzed to see what format students prefer. Lastly, the results from the manual analysis will also be presented with regard to face validity.

## 4.1 Analysis – Reading Comprehension

In this section, Class A’s reading comprehension test scores will be presented. First on a group level before transitioning into an individual perspective. Following this, the groups will be categorized based on gender to explore if there are any potential differences in performance between boys and girls.

Furthermore, while looking at table 1 some explanation is needed. Overall, 15 students were able to complete two tests which later could be analyzed. The column to the left represents the student’s number which was provided to all students to make them anonymous. The following two columns to the right indicate the points and percentage of students' scores out of the maximum, distinguished by whether they were obtained using a pen and paper (indicated by P) or a computer (indicated by C). The last two columns, from left to right respectively, represent the test on which the students managed to perform best on and if the student was a boy or a girl. To show which test the students demonstrated the strongest performance on, the percentage has been written in bold. Lastly, the average percentage score is given at the bottom of the table which will be used to demonstrate differences between different formats.

*Table 1. Results of the reading comprehension tests on a group level.*

Student #	24 ~ max points P	21 ~ max points C	Highest %	Gender
Student 2	19/24 ~ 79%	21/21 ~ <b>100%</b>	C	Boy
Student 3	7/24~ 29%	13/21 ~ <b>62%</b>	C	Boy
Student 5	16/24 ~ 67%	16/21 ~ <b>76%</b>	C	Girl
Student 8	22/24 ~ 92%	20/21 ~ <b>95%</b>	C	Girl
Student 9	13/24 ~54%	12/21 ~ <b>57%</b>	C	Girl
Student 12	14/24~ 58%	17/21~ <b>81%</b>	C	Boy
Student 15	2/24 ~ 8%	7/21 ~ <b>33%</b>	C	Boy
Student 16	20/24 ~ <b>83%</b>	15/21 ~ 71%	P	Boy
Student 17	10/24 ~ <b>42%</b>	8/21 ~ 38%	P	Boy
Student 23	21/24 ~ <b>88%</b>	11/21 ~ 52%	P	Boy
Student 24	13/24 ~ <b>54%</b>	7/21~33%	P	Girl
Student 26	23/24 ~ <b>96%</b>	16/21 ~ 76%	P	Boy
Student 27	11/24 ~ <b>46%</b>	8/21~38%	P	Boy

Student 28	9/24 ~ 38%	8/21 ~ 38%	Same	Boy
Student 30	14/24 ~ 58%	13/21 ~ <b>62%</b>	C	Girl
<b>Average</b>	<b>14,3/24 ~ 60 %</b>	<b>12,9/21~ 61 %</b>	<b>C – 53% P – 40% Same – 7%</b>	

As a group, the results show that 53% of the students performed better on a test using a computer, 40% of the students excelled when using a pen and paper while 7% managed to get the same result on both test formats. The reason for the difference in the result comes down to two students excelling on the test with a computer, which is a small difference. Furthermore, the average percentage score for the whole group did not vary much either. The average percentage score for the computer test was 60% while the average percentage score for the test using a pen was 61%, which also is a small difference. These numbers suggest that there is no considerable difference in test performance between students using a computer and those using pen and paper at the group level.

However, there are considerable differences between individual students which do not correlate to the result of the whole group. For example, student 3 managed to increase his result with the use of a computer, from 29% to 62%. Similarly, student 23 managed to obtain a score of 83% with the use of a pen and paper compared to 52% on the computer-based test. This suggests that even if the group managed to get similar results regardless of the test format, some students benefit from variation on different levels.

Furthermore, Tables 2 and 3 present the same information as Table 1 (see the beginning of section 4.1 for details) with regards to the reading comprehension test results, but with a gender-based division. The primary focus of the analysis regarding Tables 2 and 3 will be centralized around the most major factors such as average score and the test format.

*Table 2. Results of the reading comprehension tests divided by gender.*

Student #	24 = max P	21 = max C	Highest %	Gender
Student 2	19/24 = 79% p	21/21 = <b>100%</b> c	C	Boy
Student 3	7/24= 29% p	13/21 = <b>62%</b>	C	Boy
Student 12	14/24= 58% p	17/21= <b>81%</b> c	C	Boy
Student 15	2/24 = 8% p	7/21 = <b>33%</b> c	C	Boy
Student 16	20/24 = <b>83%</b> p	15/21 = 71% c	P	Boy
Student 17	10/24 = <b>42%</b> p	8/21 = 38%	P	Boy
Student 23	21/24 = <b>88%</b> p	11/21 = 52% c	P	Boy
Student 26	23/24 = <b>96%</b> p	16/21 = 76% c	P	Boy
Student 27	11/24 = <b>46%</b> p	8/21=38% c	P	Boy
Student 28	9/24 = 38% p	8/21 = 38% c	Same	Boy
<b>Average</b>	<b>14.1/24 ~ 57%</b>	<b>12.4/21 ~ 59%</b>	<b>C-50% P-40% Same-10%</b>	

*Table 3. Results of the reading comprehension tests divided by gender.*

Student #	24 = max P	21 = max C	Highest %	Gender
Student 5	16/24 ~ 67% p	16/21 ~ <b>76%</b> c	C	Girl
Student 8	22/24 ~ 92% p	20/21 ~ <b>95%</b> c	C	Girl
Student 9	13/24 ~54% p	12/21 ~ <b>57%</b> c	C	Girl
Student 24	13/24 ~ <b>54%</b> p	7/21~33% c	P	Girl
Student 30	14/24 ~ 58% p	13/21 ~ <b>62%</b> c	C	Girl
<b>Average</b>	<b>15,6/24 ~ 65%</b>	<b>13,6/21 ~ 65%</b>	<b>C – 80% P – 20%</b>	

Based on the results provided in tables 2 and 3, we can conclude that 50% of the boys performed better on an exam using a pen and paper while 40% of the boys excelled when using a computer. The last 10 % represents one boy who managed to get the same percentage score on both exams. Additionally, there was no major difference in the average test score for boys, who scored an average of 59% on the computer-based test and 57% with the use of a pen and paper. These differences seem to correlate with the findings demonstrated before regarding the whole group. On the other hand, it is noticeable that 80% of the girls performed better on the computer-based test while only 20% of the girls were more successful using a pen and paper. This does not correlate with the average test score for girls, which is equal for both tests (65%). This seems to suggest that it does not matter if boys use a computer or a pen to show their proficiency regarding reading comprehension, while girls seem to increase their scores with a computer. However, one should note that the difference between some of the results is small, e.g., student 8, got 92% on the pen-based test compared to 95% on the computer-based. Unfortunately, the reliability of the results is undermined due to the slight difference in scores between the pen-based and computer-based tests as well as the small sample size of girls.

## 4.2 Analysis – Written Essay – Flesch Kincaid

In this section, the results of Class B's written essays will be presented. The analysis method involved a two-step process. Firstly, the total number of words were counted in each essay to be able to calculate the average number of words the class produced. Secondly, a digital tool called Flesch Kincaid (see section 3.2.1 for details) was used to gather specific metrics. This tool generates values regarding how readable a text is based on the length of sentences as well as the number of syllables in the text. These values are called the FKGL (Flesch Kincaid Grade Level) and the FKRE (Flesch Kincaid Reading Ease). These scores were later converted into average class scores, allowing a comparison between the two testing formats.



Table 4 presents all of these metrics, accompanied by pen format, digital format, student numbers and gender, that will be subjected to comparison. The analysis will begin by examining the results at a group level, then delve deeper to analyze individual performance, and finally, it will segregate the data by gender.

*Table 4. Findings related to written essays conducted in both digital and pen-and-paper formats on a group level.*

Student	Gender	Pen - FKGL	Digital – FKGL	Pen -Words	Digital - Words	Pen - FKRE	Digital - FKRE
Student 5	Girl	8.0	10.8	236	357	69.2	68.4
Student 7	Boy	5.9	7	210	359	80.7	71.87
Student 9	Girl	8.1	7.1	277	404	55.3	69.1
Student 11	Girl	10.6	13.1	335	554	60.7	58.6
Student 13	Boy	6.1	9.4	352	554	74.6	73
Student 14	Boy	9.5	16.2	248	330	66.1	51.7
Student 15	Boy	6.3	5.5	101	140	78.8	82
Student 16	Girl	5.3	5.2	462	726	84.0	81.6
Student 17	Boy	6.5	7	259	352	75.2	70.6
Student 20	Girl	5.6	5.4	307	394	81.7	86.1
Student 22	Boy	6.1	6	382	841	75.6	81.5
Student 23	Girl	9.2	9	356	525	58.9	65.1
Student 25	Girl	7.4	5.8	179	406	70	83.6
Student 27	Boy	7.9	13.3	159	283	66.7	61.6
<b>Average</b>		<b>7.32</b>	<b>8.6</b>	<b>276</b>	<b>445</b>	<b>69.6</b>	<b>77.5</b>

When looking at the number provided in Table 4, the conclusion can be drawn that 100% of the students who participated in the study produced longer texts when using a computer to complete the test. The number of words produced in the test which used a pen and paper was on average 276, while the test that was taken on the computer showed an average of 445 words produced. On average, boys managed to produce 40 %, and girls 36 % more text when using a computer compared to pen and paper.

Furthermore, the average FKGL score was also higher when the students used a computer, showing a result of 8.6 compared to 7.3. The result of class B shows that the texts produced with a computer, on average, rendered a higher FKGL than with a pen and paper indicating that it is a more complex text than the one produced with a pencil.

Moreover, it is also possible to see that the FKRE increased when the students used a computer. An average of 69.6 in the test using a pen versus 77.5 with a computer. These numbers could mean that students are using an easier type of English when writing on a

computer and more complex use of words with a pen and paper. However, one should not forget that these scores do not indicate a well-structured or coherent text, only longer sentences and words consisting of more syllables.

By breaking down these scores to the individual level, it is also possible to see that, even if the average of all students showed an increase in all numbers regarding the digital format, this does not apply to all students. For example, student 9 had a pen and paper FKGL of 8.1 and a digital FKGL of 7.1, which indicates that the digital test was not the ideal format for this student to demonstrate her competence. This also applies to student 25 who has a pen and paper FKGL of 7.4 and a digital FKGL of 5.8 pointing out that the student likely performs better on a test using pen and paper. However, it is worth noting that student 25's wordcount also increased when she did the digital test, from 179 to 406, which suggests that a higher number of words does not necessarily correlate with a more complex text, since, in student 25's case, the FKGL proved to be lower when the number of words increased.

### 4.3 Analysis – Written Essay – Gender Comparison

Furthermore, Tables 5 and 6 present the same information as Table 4 (see section 4.2 for details). However, in this section the results will be divided by gender. The analysis regarding Tables 5 and 6 will focus on the average score of FKGL and FKRE to see if there are any differences between the genders when they are using one test format or the other.

*Table 5. Findings related to the written essays conducted in both digital and pen-and-paper formats have been categorized based on gender.*

Student	Gender	Pen - FKGL	Digital – FKGL	Pen – Words	Digital - Words	Pen - FKRE	Digital - FKRE
Student 5	Girl	8	10.8	236	357	69.2	68.4
Student 9	Girl	8.1	7.1	277	404	55.3	69.1
Student 11	Girl	10.6	13.1	335	554	60.7	58.6
Student 16	Girl	5.3	5.2	462	726	84.0	81.6
Student 20	Girl	5.6	5.4	307	394	81.7	86.1
Student 23	Girl	9.2	9	356	525	58.9	65.1
Student 25	Girl	7.4	5.8	179	406	70	83.6
<b>Average</b>		<b>7.7</b>	<b>8.1</b>	<b>307</b>	<b>481</b>	<b>66</b>	<b>73.2</b>

*Table 6. Findings related to the written essays conducted in both digital and pen-and-paper formats have been categorized based on gender.*

Student	Gender	Pen - FKGL	Digital – FKGL	Pen – Words	Digital - Words	Pen - FKRE	Digital - FKRE
---------	--------	------------	----------------	-------------	-----------------	------------	----------------

Student 7	Boy	5.9	7	210	359	80.7	71.8
Student 13	Boy	6.1	9.4	352	554	74.6	73
Student 14	Boy	9.5	16.2	248	330	66.1	51.7
Student 15	Boy	6.3	5.5	101	140	78.8	82
Student 17	Boy	6.5	7	259	352	75.2	70.6
Student 22	Boy	6.1	6	382	841	75.6	81.5
Student 27	Boy	7.9	13.3	159	283	66.7	61.6
<b>Average</b>		<b>6.9</b>	<b>9.2</b>	<b>244</b>	<b>408</b>	<b>74</b>	<b>70</b>

When dividing the results based on gender the results show several differences. When it comes to FKGL with regards to using the pen and paper format, girls got an average of 7.7 versus an average of 8.1 when using a computer. This indicates a slight advantage for the computer. Boys, likewise, got an average of 6.9 with a pen and paper and an average of 9.2 with a computer, which suggests that boys do better on computer-based exams for essays than on pen-based exams. In addition, these findings seem to suggest that both boys and girls are more proficient when writing an essay using a computer versus pen and paper.

Adding to the FKGL score for boys, the analysis regarding the FKRE revealed contrasting patterns between pen-based and computer-based text. In the pen-based tests, boys had an average FKRE of 74 while in the computer-based test, their score was an average of 70. This is a slight indication that the boys write more complex texts when using a computer compared to using a pen. On the other hand, girls produced more complex texts when writing with a pen compared to a computer, scoring an average of 66 with a pen and 73.2 with a computer. These findings might suggest that boys and girls are more proficient in using one or the other format when writing an essay. Compared to each other, boys perform better when using a computer and girls when using a pen and paper. However, one should note that a lower score for the FKRE could represent a more complex text or a text with an underuse of punctuation, rendering long but faulty sentences.

Another key factor to look at is the correlation between word count and the FKRE results. Girls managed to get an FKRE score of 66 and a word count of 307 on the pen-based test. On the computer-based test, they got an FKRE score of 73.2 with a word count of 481. This might be an indication of girls producing more complex text in a pen-and-paper setting since the FKRE score becomes lower when the text contains a lower number of words. Looking at the same correlation for the boys, the analysis shows the opposite pattern. With an FKRE score of 74 and a word count of 244 on the pen-based test versus an FKRE score of 70

and a word count of 408 on the computer-based test. It thus seems like boys generate more complex texts with a computer even if the word count goes up, contrary to the girls' results.

#### 4.4 Analysis – Written Essay – LexTutor

Table 7 presents the results of the written essays, analyzed through LexTutor (n.d.), on a group level in class B. The test results are divided into four categories: K1, K2, K3, and Off-List Words, which are indicative of various levels of vocabulary proficiency and words that are not included in the language corpus. Each category is reported for two different tools used by the students, namely Pen and Computer. Table 7 also shows that the group consists of both boys and girls, who were assigned a unique number for identification purposes. For each student, the percentage of vocabulary falling into the K1, K2, K3, and Off-List Words categories are provided, both for their handwritten work (Pen) and their digital work (Computer). Lastly, the average score of the group is provided at the bottom of the table. These numbers will be compared to see if there are any differences regarding test format.

*Table 7. The results of the written essays that were analyzed in LexTutor.*

	Pen	Computer	Pen	Computer	Pen	Computer	Pen	Computer	
Student #	K1 %	K1 %	K2 %	K2 %	K3 %	K3 %	Off-List Words %	Off-List Words %	Gender
Student 5	86.32	88.77	5.13	6.03	2.56	1.1	5.98	4.11	Girl
Student 7	83.81	84.39	3.33	1.95	1.43	0	11.43	13.66	Boy
Student 9	81.7	83.61	1.08	4.34	3.58	2.17	13.62	9.88	Girl
Student 11	87.71	87.69	1.43	5.03	3.14	2.25	7.71	5.03	Girl
Student 13	84.15	83.46	3.66	7.52	0	0.75	12.2	8.27	Boy
Student 14	88.4	89.29	5.6	5.95	3.6	2.68	2.4	2.08	Boy
Student 15	87.38	97.2	3.88	1.4	0.97	0	7.77	1.4	Boy
Student 16	85.62	86.08	5.36	3.75	0.21	0.94	8.8	9.24	Girl
Student 17	83.1	86.94	5.41	5	4.25	3.06	7.34	5	Boy
Student 20	90.28	93.67	3.45	3.65	2.51	0.24	3.76	2.43	Girl
Student 22	86.22	88.71	4.26	3.76	4.76	1.88	4.76	5.65	Boy
Student 23	82.96	86.67	3.35	3.62	3.07	2.48	10.61	7.24	Girl
Student 25	87.15	91.41	3.35	4.06	5.03	0.72	4.47	3.82	Girl
Student 27	79.5	88.77	4.35	2.81	7.45	1.75	8.7	6.67	Boy
<b>Average</b>	<b>~ 85.3</b>	<b>~ 88.3</b>	<b>~ 3.8</b>	<b>~ 4.2</b>	<b>3.04</b>	<b>~ 1.43</b>	<b>~ 7.8</b>	<b>~ 6</b>	

When the results were analyzed on a group level it shows that the K1 and K2 word use, on average, increased when the students used a computer to write (85.3% – 88.3%, 3.8% - 4.2%). On the other hand, K3 words decreased when a computer was used (3.04% - 1.43%). These findings might suggest that when students write an essay with a computer, they tend to use more basic words, while also lowering their use of K3 words and Off-List words. This might be because the students are more used to typing with a keyboard compared to a pen and paper, resulting in a faster pace of writing which explains the more frequent use of basic words and the decrease of K3 words. The decrease of off-list words might be due to the unfamiliarity of using a pen since the text produced with a pen often proved to be more difficult to read due to messy handwriting, which might have resulted in errors when transcribing the text from paper to computer.

#### 4.5 Analysis – Written Essays - Gender Comparison

Furthermore, Table 8 presents the same data as Table 7 (see section 4.4 for details), but with the distinction of presenting results specific to the girls in class B. The comments made regarding Table 8 will center around the average score achieved by the girls, which is displayed at the bottom of the table.

*Table 8. The results of the written essays that were analyzed in LexTutor have been divided by gender.*

	Pen	Computer	Pen	Computer	Pen	Computer	Pen	Computer	
Student #	K1 %	K1 %	K2 %	K2 %	K3 %	K3 %	Off-List Words %	Off-List Words %	Gender
Student 5	86.32	88.77	5.13	6.03	2.56	1.1	5.98	4.11	Girl
Student 9	81.7	83.61	1.08	4.34	3.58	2.17	13.62	9.88	Girl
Student 11	87.71	87.69	1.43	5.03	3.14	2.25	7.71	5.03	Girl
Student 16	85.62	86.08	5.36	3.75	0.21	0.94	8.8	9.24	Girl
Student 20	90.28	93.67	3.45	3.65	2.51	0.24	3.76	2.43	Girl
Student 23	82.96	86.67	3.35	3.62	3.07	2.48	10.61	7.24	Girl
Student 25	87.15	91.41	3.35	4.06	5.03	0.72	4.47	3.82	Girl
<b>Average</b>	<b>~ 86</b>	<b>~ 88.2</b>	<b>~ 3.3</b>	<b>~ 4.35</b>	<b>2,9</b>	<b>~ 1.4</b>	<b>~ 7.9</b>	<b>~ 6</b>	

When looking at the girls' results in Table 8, the conclusion can be made that it follows the pattern for the whole group. With the use of a computer, the girls managed to increase their K1 word use from 86 % to 88.2 % while at the same time showing an increase in K2 word use as well, from 3.3 % word usage to 4.35%. The girls also showed a decrease in

K3 words, from 2.9% to 1.4%, as well as off-list words from 7.9% to 6 % when a computer was used. This might also be because the girls are used to typing on a computer at a faster pace which results in more basic word usage. The same also applies to the girls, compared to the whole group, when it comes to messy handwriting which might have impacted the results.

Moreover, Table 9 also portrays the same information as Table 7 (see section 4.4 for in-depth description), however, the comments regarding Table 9 will solely focus on the average performance of the boys in class B, displayed at the bottom of the table.

**Table 9.** *The results of the written essays that were analyzed in LexTutor have been divided by gender.*

	Pen	Computer	Pen	Computer	Pen	Computer	Pen	Computer	
Student #	K1 %	K1 %	K2 %	K2 %	K3 %	K3 %	Off-List Words %	Off-List Words %	Gender
Student 7	83.81	84.39	3.33	1.95	1.43	0	11.43	13.66	Boy
Student 13	84.15	83.46	3.66	7.52	0	0.75	12.2	8.27	Boy
Student 14	88.4	89.29	5.6	5.95	3.6	2.68	2.4	2.08	Boy
Student 15	87.38	97.2	3.88	1.4	0.97	0	7.77	1.4	Boy
Student 17	83.1	86.94	5.41	5	4.25	3.06	7.34	5	Boy
Student 22	86.22	88.71	4.26	3.76	4.76	1.88	4.76	5.65	Boy
Student 27	79.5	88.77	4.35	2.81	7.45	1.75	8.7	6.67	Boy
<b>Average</b>	<b>~ 84.7</b>	<b>~ 88.4</b>	<b>~ 4.4</b>	<b>~ 4.1</b>	<b>3.2</b>	<b>~ 1.5</b>	<b>~ 7</b>	<b>~ 6.1</b>	

Similar to the girls and the whole group, by using a computer, the boys increased their K1 word use from 84.7% to 88.4%. However, the boys decreased their word use in K2 (4.4% - 4.1%), K3 (3.2% - 1.5%) and off-list words (7% - 6.1%), resulting in even more basic words used when using a computer. These findings might also suggest that boys have a faster pace of writing due to being more used to the format of a computer which might explain the result. This also correlates with the findings in the Flesch Kincaid readability test, which suggested that boys excel in FKGL, FKRE and word count when using a computer, indicating a familiarity with computers.

## 4.6 Questionnaire

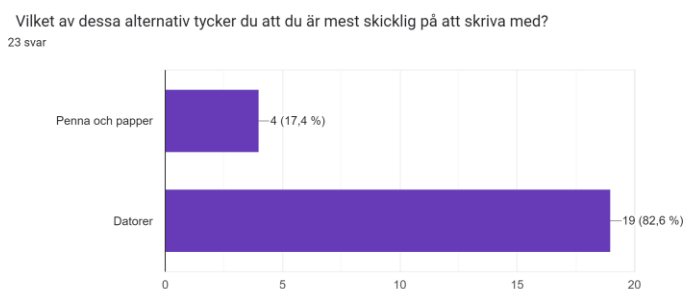
Following the completion of their tests, the students were also provided with the chance to respond to a questionnaire. Figure 1 represents options, either pen and paper (top graph) or computers (bottom graph), to the question, “Which of these options do you think is the easiest to take an exam on?”. Comments pertaining to Figure 1 and subsequent figures will primarily center on the format students prefer.



**Figure 1.** Results regarding which format students perceive as the easiest to perform a test on. Pen and paper bar on top and computer bar on the bottom.

Looking at Figure 1, it is evident that the majority of students find computer-based testing easier. Specifically, 82.6% (19 students) expressed that it is easier to take a test on a computer, compared to a mere 17.4% (4 students) who preferred the traditional pen-and-paper method.

Furthermore, with the same options and structure as for Figure 1, Figure 2 shows the students’ responses regarding the question, “Which of these options do you feel most proficient at writing with?”.

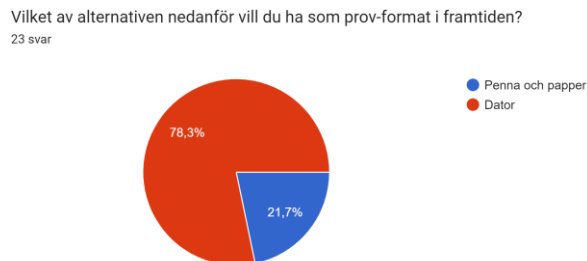


**Figure 2.** Results regarding which format students perceive themselves to be most proficient on. Pen and paper bar on top and computer bar on the bottom.

In line with the findings regarding Figure 1, Figure 2 illustrates that the majority of students, 82.6% (19 students), believe they have a higher level of proficiency in writing when

using a computer, whereas a smaller percentage of students, 17.4% (4 students), reported feeling more skilled when using pen and paper.

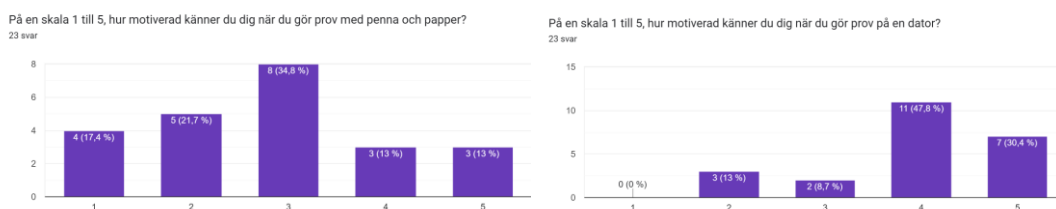
Moreover, the circle diagram in Figure 3 provides an illustration of what type of format students want to use in the future when taking a test. The red color of the circle diagram indicates the answer, computer, while the blue color represents, pen and paper.



**Figure 3.** Circle diagram which indicates what format students prefer to use in the future.

Figure 3 illustrates that most students prefer taking their exams using a computer rather than using pen and paper. 78.3% (18 students) reported that they want a computer as their future test format, while 21.7% (5 students) expressed that they would rather use pen and paper in the future.

Furthermore, Figure 4 illustrates the students' level of motivation, rated on a scale from one to five, during pen-and-paper testing (shown in the left bar chart), as compared to computer-based testing (represented in the right bar chart).



**Figure 4.** Bar charts demonstrating results regarding how motivated students feel from one to five.

The results, presented in Figure 4, indicated a major preference for computer-based testing. On the left-hand graph, only 26% (depicted in bars 4 and 5) of students reported feeling motivated or highly motivated during a pen-and-paper format. Conversely, 39.1% (depicted in bars 1 and 2) conveyed a lack of motivation or a marginal sense of motivation with this traditional method. An additional 34.8% (bar 3) fell into the neutral category. Shifting the focus to the right-hand graph, which represents the responses to the computer-based exam question, a considerable difference becomes evident. An overwhelming 78.2%



(bars 4 and 5) of the students reported feeling motivated or highly motivated when taking a test on a computer, while 13% (bar 2) expressed a low level of motivation, and a mere 8.7% stated they felt neutral about this test-format. This suggests that the majority of the surveyed students demonstrated a clear preference for computer-based testing over traditional pen-and-paper methods, suggesting that they feel more motivated when using a computer during a test.

## 4.7 Manual Analysis

The manual analysis is based on face validity, consisting of the impressions gathered when reading and transcribing the written pen-and-paper-based test into the computer. This manual analysis aims to fill some of the gaps that the Flesch Kincaid readability test does not account for.

Overall, transcribing certain texts proved challenging due to the poor handwriting of some students. This might have impacted the results in the Flesch Kincaid readability test and LexTutor (n.d.) since some words could not be deciphered, resulting in gaps in some of the students' text when digitized. With regards to handwriting, even if there are individual differences, the girls seemed to produce more readable texts compared to the boys, although, some girls produced less readable texts as well.

With the use of a computer, it seemed that the students managed to produce more texts that lacked correct usage of punctuation which resulted in many run-on sentences. Since all students produced more text when writing on a computer, it seems logical that the number of run-on sentences also increased. The keyboard might make students type faster which could result in less thinking about when to use punctuation. This was the case for both boys and girls.

Furthermore, the correct usage of capitalized letters increased when the students used a computer. This might be due to students not being used to the format of a pen and paper, resulting in incorrect usage of lowercase letters as well as capitalized ones. By employing a computer, students no longer rely on muscle memory for letter formation; instead, they only need to understand the capitalization rules since computers consistently correctly display letters. Adding to this, it is important to note that students did not have access to autocorrecting tools when they wrote their essays. Since the students were able to demonstrate a more accurate usage of capitalization when using a computer, it seems that they possess knowledge of when to use capitalized letters, but their poor handwriting may mislead teachers when assessing their written texts.

## 5 Discussion

To summarize, the objective of this study was to investigate the differences between pen-based tests and computer-based tests. The research sought to determine if digitized tests are the most effective method for students to show their proficiency and if there are any differences in performance between boys and girls when comparing traditional and computer-based testing. This section will discuss the results presented in this study with regard to previous research mentioned in the literature review. The results, regarding groups A and B, will first be summarized and discussed according to the average score, Flesch Kincaid readability test and LexTutor (n.d.). Lastly, the results of the questionnaire as well as the manual analysis will also be put into perspective based on previous research.

The reading comprehension test relied solely on the average score based on students' performance, which revealed no major difference in the group's performance when using a computer compared to pen and paper. However, 80% of the girls in class A performed better on the computer-based reading comprehension test, which is supported by earlier research conducted by Hilton and Canciello (2016), who showed that females' attitudes towards digital reading are more positive than males. Although 80% of the girls managed to get a higher score when using a computer, the difference in the results were too slight to consider the outcome reliable. As an illustration, one girl obtained a 92% score on the pen-based test and 95% on the computer-based one. While this suggests an advantage for the computer, it is likely not a noteworthy one. Adding to this, it is also important to note that the sample size of girls in this study was small, consisting of five students, which makes it hard to generalize the results to a larger group. These findings contradict what Mangen et al. (2013) and Muller and Oppenheimer (2014) found in their studies regarding reading comprehension. According to their findings, students demonstrated better performance on the pen-and-paper-based test. However, in this study, it was observed that students either performed better when using a computer or achieved similar results compared to the pen-and-paper format. The reason for this might be due to students being more used to the format of a computer since the school is considered to be highly digitalized.

Furthermore, the analysis of the texts using the Flesch Kincaid readability test demonstrated that all students managed to produce more text when using a computer as opposed to pen and paper. These findings are supported by Mangen et al. (2015), who claims that the use of a computer enables a student to finish their tests faster, due to typing being more efficient. Likewise, the students as a group managed to increase their average FKGL

with the use of a computer, which indicates that they tend to write more complex texts while using this device. This contradicts what Mueller and Oppenheimer (2014), and Van Der Velden (2021) mention in their studies, regarding pen and paper enabling students to interact with the text in a more direct manner resulting in a deeper comprehension at a cognitive level. However, this was not the case for all students since some of them performed better with the use of a pen and paper.

When it comes to the FKRE score, boys seem to write more complex texts when using a computer while girls tend to perform slightly better with the use of pen and paper. Adding to this, it also seems as if boys produce more complex text when the word count is high (computer) while girls achieved more complex text when the word count was lower (pen and paper). Based on these findings, it seems as if the girls' results correlate to what Muller and Oppenheimer (2014) and Van Der Velden (2021) mention regarding a deeper understanding with the use of a pen, while boys do not. However, it is important to keep in mind that even if a lower score on FKRE can be an indication of a more complex text; the text might be nonsensical due to restrictions of the digital tool, such as not accounting for coherence or non-words. In addition, the FKGL does not take correct punctuation into account so a higher score can be obtained through a lack or underuse of full stops.

The results provided by LexTutor (n.d.) show that students tend to write texts with more common words with the use of a computer, while at the same time decreasing the use off-list-words. This result also might be due to the students being more used to the format of the computer, which seems to be the case when looking at the answers from the questionnaire and the finding of bad handwriting presented in the manual analysis. This correlates with what Backles and Cowan (2019) found regarding a negative effect on test performance when implementing a new test format. Since the students usually write with the use of a computer, the pen-and-paper-based format prohibits the students from showing their full potential and the result becomes impacted. Due to the bad handwriting of some students, it was hard to read some of the texts which impacted the off-word-list in the analysis, if this would not have been the case more words might have been added in the K1, K2, and K3 frequency bands.

While analyzing the results of the questionnaire it became clear that a majority of students prefer to use a computer when doing a test. The reason for this appears to be that students feel more comfortable using a computer than pen and paper, as 81.8% of the students reported feeling more proficient with computer usage. This is important to keep in mind when interpreting the results of this study since the school at which the study was conducted is

considered to be a highly digitalized school. Thus, the results might be misleading due to students being more used to typing on a computer compared to pen and paper. In addition, 77.3% of the students who completed the questionnaire expressed that they feel motivated to a higher degree when doing a test on the computer. These findings suggest that a majority of the students will perform better on a computer since their attitudes towards using a pen and paper might impact their results negatively. This correlates to Chua and Don's (2013) discoveries regarding students being more motivated with the use of a computer versus pen and paper. However, even if the results indicate that students prefer the use of a computer in test situations, one should remember that some prefer the traditional method (26.1 %). Teachers should view these results as an indication of different preferences when students take part in tests. Some students prefer to use a computer while others want to use a pen and paper.

Lastly, based on the face validity provided in the manual analysis, it is evident that the format of a test has an impact on a student's performance. The finding that students may write more run-on sentences with the use of a computer correlates to what Mueller and Oppenheimer (2014) found with regard to the faster-paced, non-processing "verbatim notetaking" (p.1160). In their study, students wrote at a quicker pace with the use of a computer compared to pen and paper, which led to students not processing the information they were writing down. This phenomenon could also have taken place when the students in this study were writing with a computer, resulting in more run-on sentences and unprocessed text. This also correlates with Backles and Cowan's (2019) findings, which suggest that a new test format might have a negative effect on students' performance. On a positive note, the manual analysis also shows that capitalized letters were, it seemed, used more correctly when the students were using a computer. In short, the format matters.

## **6 Conclusion**

In conclusion, this study reveals that the format of testing does indeed matter, impacting the results both quantitatively and qualitatively. While the results, on average, show that the computer gives an advantage to the students at this school, it also emphasizes that preferences and performance can vary among students. Moreover, there also seems to be some differences between boys and girls regarding what format is being used, since some of the average scores of girls pointed to more complex text with the use of a pen and paper while that was not the case for boys.

Due to these results, teachers who construct tests should consider the fact that not all students perform better in a computer-based format. Since the result does not point to a unified direction, teachers should not solely create tests in one format and expect that all students will excel using that method. Instead, teachers should view the results presented as students preferring different formats and let them show their proficiency in different settings, or let them practice in the other format to improve. This will be crucial for teachers to consider when constructing a test since the format might impact the student's performance.

Even though the results, generally, point towards an advantage for the computer, another challenge for teachers will be to construct reliable tests which cannot be compromised by academic dishonesty. This will be difficult to do if the employer does not pay for digital tools, such as Exam.net (n.d.), which makes it possible for the teacher to lock the computer as well as customize what tools the students can use.

Moreover, further research is needed regarding which format students should use in order to optimize their learning process in the most effective way. Even if the results of this study show that the majority of these students perform better on a computer in a test environment, it does not account for which format is best suited for learning. A longitudinal study regarding students who are proficient at both using a computer and pen and paper should be conducted, to see if one format outperforms the other with regards to learning. This could be interesting since some research mentioned in this study implies that a pen-and-paper format has proved to be better for note-taking regarding memory retention (Mueller & Oppenheimer, 2021).

Another area for improvement is to carry out more large-scale studies including a manual analysis of quality in terms of accuracy e.g. punctuation and coherence. One approach could involve having teachers evaluate texts in both modes (digital analysis and manual analysis) to triangulate the results and thus provide a comprehensive assessment. This additional step would provide a more holistic understanding of the texts' quality and enhance the accuracy of the overall evaluation process.

Lastly, considering that the results indicated that students tend to perform better with computers in a test setting, it would be intriguing to explore the reliability of computer-based assessments. Given the rapid advancement of artificial intelligence, teachers must have dependable guidelines to ensure the reliability of the assessments they administer.

## References

- Backes, B., Cowan, J. (2019). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review*, 68, 89–103. <https://doi.org/10.1016/j.econedurev.2018.12.007>
- Barmark, M., & Djurfeldt, G. (2020). *Statistisk Verktyglåda 0: Att Förstå och förändra världen Med Siffror*. Studentlitteratur AB.
- Bryman, A., Clark, T., Foster, L., & Sloan, L. (2021). *Bryman's Social Research Methods* (6th ed.). Oxford University Press.
- Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior*, 29(5), 1889–1895. <https://doi.org/10.1016/j.chb.2013.03.008>
- Cobb, T. (n.d.). LexTutor. Retrieved May 23, 2023, from <https://www.lextutor.ca/>
- Exam.net. (n.d.). Retrieved June 17, 2023, from <https://exam.net/>
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- Google LLC. (n.d). *Overview*. Retrieved May 5, 2023, from <https://www.google.se/intl/sv/forms/about/>
- Hilton, J. T., & Canciello, J. (2015). Male and female perceptions of an iPad-based digital literacy curriculum. *Journal of Educational Multimedia and Hypermedia*, 24(1), 23–41. <https://www-learntechlib-org.ezproxy.ub.gu.se/primary/p/130556/>.
- LexTutor. (n.d.). Retrieved June 17, 2023, from <https://www.lextutor.ca/>
- Lim, J., Wu, W., Wang, J., Detre, J. A., Dinges, D. F., & Rao, H. (2010). Imaging Brain Fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. *NeuroImage*, 49(4), 3426–3435. <https://doi.org/10.1016/j.neuroimage.2009.11.020>
- Mangen, A., Walgermo, B.R. and Brønneick, K. (2013). “Reading linear texts on paper versus computer screen: Effects on reading comprehension,” *International Journal of Educational Research*, 58, 61–68. <https://doi.org/10.1016/j.ijer.2012.12.002>.

- Mangen, A., Anda, L. G., Oxborough, G. H., & Brännick, K. (2015). Handwriting versus keyboard writing: Effect on word recall. *Journal of Writing Research*, 7(2), 227–247. <https://doi.org/10.17239/jowr-2015.07.02.1>
- Manwaring, K. (2019). “Lines of Desire: The Phenomenology of long-handwriting in creative praxis,” *New Writing*, 17(2), 133–145. <https://doi.org/10.1080/14790726.2019.1586958>.
- Microsoft Corporation. (2023). Microsoft Word. (Version 2303) [Computer software]
- Muller, P.A. & Oppenheimer, D.M. (2014). The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking. *Psychological Science*, 25(6), 1159-1168. <https://doi.org/10.1177/0956797614524581>
- Regeringsbeslut U2017/03739/GV. *Uppdrag att digitalisera de nationella proven m.m.* <https://www.regeringen.se/contentassets/03dee5c5cdf244afa053e26cf654a8d8/uppdra-g-att-digitalisera-de-nationella-proven-m.m..pdf>
- Seok, S., DaCosta, B. (2017). Gender differences in teens’ digital propensity and perceptions and preferences with regard to digital and printed text. *TechTrends*, 61(2), 171–178. <https://doi.org/10.1007/s11528-016-0134-4>
- Skolverket. (2016). *Nationella proven i grundskolans årskurs 6 och 9.* <https://www.skolverket.se/download/18.6bfaca41169863e6a65c5cb/1553967061504/pdf3695.pdf>
- Skolverket. (2022, 13 October). *Sekretess på prov i grundskolan.* <https://www.skolverket.se/undervisning/grundskolan/nationella-prov-i-grundskolan/forbereda-och-bestalla-prov-i-grundskolan/sekretess-pa-prov-i-grundskolan>
- Skolverket. (2023a, 27 Mars). *Leda digitaliseringen på skolor och förskolor.* <https://www.skolverket.se/skolutveckling/leda-och-organisera-skolan/leda-digitaliseringen-i-skola-och-forskola>
- Skolverket. (2023b). *PIRLS 2021 Läsförmågan hos svenska elever i årskurs 4 i ett internationellt perspektiv.* <https://www.skolverket.se/getFile?file=11490>

- Van Der Velden, M. (2021). 'I felt a new connection between my fingers and brain': A thematic analysis of student reflections on the use of Pen and paper during lectures. *Teaching in Higher Education*, 1–18. <https://doi.org/10.1080/13562517.2020.1863347>
- Vetenskapsrådet (2002). Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning. Stockholm: Vetenskapsrådet.  
[https://www.vr.se/download/18.68c009f71769c7698a41df/1610103120390/Forskning\\_setiska\\_principer\\_VR\\_2002.pdf](https://www.vr.se/download/18.68c009f71769c7698a41df/1610103120390/Forskning_setiska_principer_VR_2002.pdf)
- Vetenskapsrådet. (2017). God forskningssed. Stockholm: Vetenskapsrådet.  
<https://publikationer.vr.se/produkt/god-forskningssed/>



# Appendices

## Appendix A - Negative Consent Letter

### Studie om datorer kontra penna och papper i åk 9

Studieansvarig: namn, lärarkandidat i engelska genom Göteborgs universitet

Digitaliseringen av den svenska skolan är i full gång och verkar inte sakta ned. Datorer används flitigt och penna och papper får ofta sitta på avbytarbänken. Med detta som bakgrund kommer jag under min praktik på X-skolan genomföra en undersökning där syftet är att ta reda på om elevernas prestation skiljer sig när de använder penna och papper kontra datorer vid provsituationer. Tanken är att eleverna ska få göra ett gammalt nationellt prov på datorn och ytterligare ett gammalt nationellt prov med penna och papper. Resultatet från dessa prov kommer sedan analyseras och jämföras för att se vilka för- och nackdelar de olika tillvägagångssätten har. Detta blir även ett ypperligt tillfälle för eleverna att förbereda sig inför de nationella proven som kommer snart.

Detta informationsbrev skickas till er eftersom det krävs samtycke från föräldrar för att undersökningen ska kunna genomföras. Om du godkänner att ditt barn deltar i undersökningen, behöver du inte göra någonting. Om du däremot **inte** vill att ditt barn deltar i undersökningen, vänligen skicka ett mejl till mig namn (mailadress) eller Namn (mailadress), klassföreståndare i årskurs 9. Eleverna har förstås också rätt att säga nej till att medverka och avbryta sin medverkan närhelst de önskar.

Resultaten kommer att behandlas konfidentiellt och ingen kommer kunna utläsa vem som svarat vad i min färdiga text. Elevernas resultat i denna studie kommer heller inte påverka deras betyg i engelska.

Tack för att ni tog er tiden att läsa denna information!

Med vänliga hälsningar, namn

# Appendix B - Questionnaire

2023-05-24 09:17

Undersökning

## Undersökning

Besvara frågorna nedanför på ett så ärligt sätt du bara kan.

\* Anger obligatorisk fråga

---

1. E-post \*

---

2. Kön \*

*Markera endast en oval.*

- Kille
- Tjej
- Annan
- Vill inte säga

3. Skriv din siffra \*

---

4. Vilket av dessa alternativ tycker du det är lättast att göra ett prov på? \*

*Markera endast en oval.*

- Penna och papper
- Datorer

5.

\*

Vilket av dessa alternativ tycker du att du är mest skicklig på att skriva med?

*Markera endast en oval.*

Penna och papper

Datorer

6. På en skala 1 till 5, hur motiverad känner du dig när du gör prov på en dator? \*

*Markera endast en oval.*

Inte alls motiverad

1

2

3

4

5

Mycket motiverad

7. På en skala 1 till 5, hur motiverad känner du dig när du gör prov med penna och papper? \*

Markera endast en oval.

Inte alls motiverad

1

2

3

4

5

Mycket motiverad

8. Vilket av alternativen nedanför vill du ha som prov-format i framtiden? \*

Markera endast en oval.

Penna och papper

Dator

---

Det här innehållet har varken skapats eller godkänts av Google.

Google Formulär