Handwriting versus keyboarding in first grade:

Which modality best supports written composition performance and learning?

by

Eivor Finset Spilling

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)

Faculty of Arts and Education
Norwegian Reading Centre
2023

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

©2023 Eivor Finset Spilling

ISBN: 978-82-8439-157-1
ISSN: 1890-1387
PhD: Thesis UiS No. 692

Acknowledgements

The research reported in this thesis was funded by the Norwegian research council (grant number 273422) through the *DigiHand* project, a cooperation between Volda University College and the Reading Centre at the University of Stavanger.

First and foremost, I wish to express my gratitude to my supervisors and co-authors. It has been a privilege to work with you and learn from you. My main supervisor, Vibeke Rønneberg, you have been the ultimate supervisor. I am extremely grateful for your advice and support throughout. Thank you for everything you have taught me about research, including how to combine it with family life, and for always being available when I needed to discuss an issue or ask a question. You have been a constant inspiration! Co-supervisor Wenke Mork Rogne, thank you for inviting me into the *DigiHand* project. Without you, this thesis would not have come into being. Your positivity, encouragement, care and guidance from beginning to end have been invaluable! Co-supervisor Mark Torrance, thank you for sharing your inexhaustible knowledge and expertise. I am enormously grateful for your thorough and astute feedback, which has pushed me forward again and again. Jens Roeser, thank you for engaging in my project and generously teaching me how to perform a Bayesian analysis.

I have been fortunate to be part of the *DigiHand* research group, and I would like to thank all of you for your professional cooperation and wonderful companionship. Special thanks go to the project management, Wenke Mork Rogne and Siv M. Gamlem, for believing in me and for facilitating the execution of the PhD by providing equipment and arranging inspiring seminars.

I would also express my sincerest thanks to all the school leaders, teachers and students who participated in the project. This research would not have been possible without you.

I owe thanks to colleagues at Volda University College, particularly all fellow workers at Lettbygget, for chats and discussions, both professional and personal. I am also grateful to colleagues at Stavanger University for welcoming me when I visited. Moreover, I am indebted to Oddny Judith Solheim and Gustaf B. Skar for their valuable comments on my midway evaluation.

Finally, I would like to thank my friends and family for their encouragement and interest in my project. Mam and Dad, I deeply appreciate your support and enthusiasm. Ingvild, thank you for being there for me – I always feel better after talking to you. Anne Marie and Rolf, thank you for helping us and, particularly, for your contributions when I had to go away for courses and seminars. My warmest thanks go to Tor Arne, for your love, patience and endless support and for always being willing to discuss issues and read text; and to Agnar and Einar, for your unconditional love and for reminding me of what is the most important in life.

Volda, 10 January 2023

Eivor Finset Spilling

Summary

An important background for the present thesis is the increasing digitalisation in school, and more specifically, the Norwegian first-grade reality, where a growing number of schools provide students with personal digital devices to be used in initial writing instruction. The research that compares effects of handwriting and keyboarding on children's early writing is, however, scarce, findings are inconsistent, and many of the studies suffer from methodological problems, for example, inadequate control of children's prewriting experience (Wollscheid et al., 2016).

The aim of the present thesis was therefore to investigate whether modality – handwriting on paper or keyboarding on digital tablet with text-to-speech functionality – affects first grader's written composition performance and written composition learning, and whether these effects depend on children's literacy skills (grapheme-phoneme mapping, first sound segmentation, blending, word reading, spelling and vocabulary) measured at school start. This was examined in a sample of Norwegian first graders from 18 schools, where five schools taught children to write by hand, five schools taught children to write by digital tablet postponing handwriting, and eight schools taught children to write both by hand and using a digital tablet. Children's compositions were analysed for length and quality by formally assessing a set of text features related to both transcription (spacing, spelling and punctuation) and narrative sophistication (vocabulary, syntax and narrative structures). The text quality measures were specifically developed for assessing narratives by beginning writers which typically are short and simple. The statistical modelling was done using Bayesian methods, which allow for demonstrating evidence in both the presence and absence of effects.

This thesis includes four articles. Article 1 is a philosophical discussion of how texts by beginning writers can be analysed from a quantitative viewpoint. The three remaining articles contribute to the

thesis by empirically investigating the effects of modality on first graders' written composition performance and written composition learning. Article 2 shows that first graders who are taught writing in both modalities from the start of school are likely to produce compositions of similar length and quality in both modalities. This article also shows that the lack of a modality effect on written composition performance does not depend on children's literacy skills. For example, students with weaker literacy skills did not produce stories of higher quality in one or another modality.

Article 3 demonstrates that first-grade students receiving instruction based on handwriting or digital tablets with otherwise minimal change to instruction, overall learn to compose text at the same rate throughout the first year of formal writing instruction. The students showed similar development in text length, syntactic complexity and accuracy, and narrative structures, regardless of learning to write by hand or with a digital tablet. Students writing with a digital tablet showed better performance in transcription accuracy (spelling, spacing and terminator accuracy), but showed little or no development of these text features through the first grade. Students writing by hand started at a lower performance level for transcription accuracy but showed improvement throughout the year. This difference in performance can probably be attributed to the text-to-speech functionality offered by the digital tablets. Article 4 shows that there were no interaction effects between modality and students' literacy skills on learning to compose text. This means that there were, for example, no advantages related to learning to compose text with a digital tablet, or by hand, for students with weaker literacy skills.

The conclusion of the thesis is that, in a context similar to the one studied here, modality does not substantially affect first-grade students' written composition performance or written composition learning. Thus, it seems that instruction based on handwriting and instruction based on digital tablets can provide children with similar opportunities to develop their written composition skills in their first year of school. Before clear

recommendations about the choice of modality for initial writing instruction can be made, future research should investigate the potential transition effects of going from learning to write in one modality to the other.

Sammendrag

En viktig bakgrunn for denne avhandlingen er den økende digitaliseringen i skolen, og mer spesifikt den norske første-klasse-virkeligheten, der et økende antall skoler utstyrer elevene med personlige digitale enheter til bruk i skriveopplæringen. Forskningen som sammenligner effektene av håndskrift og tastaturskriving på barns tidlige skriving er imidlertid knapp, funn er inkonsistente og mange av studiene lider av metodologiske svakheter, for eksempel utilstrekkelig kontroll av deltakernes tidligere skriveerfaring (Wollscheid et al., 2016).

Målet med denne avhandlingen var derfor å undersøke om modalitet – håndskrift på papir eller tastaturskriving på nettbrett med tekst-til-tale funksjonalitet – påvirker førsteklassingers prestasjon i og læring av tekstkomposisjon, og om disse modalitetseffektene avhenger av barnas literacyferdigheter (grafem-fonem-kunnskap, framlydsanalyse, fonologisk syntese, ordlesing, staving og vokabular) målt ved skolestart. Dette ble undersøkt i et utvalg av norske førsteklassinger fra 18 skoler, hvorav fem skoler lærte barna å skrive for hånd, fem skoler utsatte håndskriftsopplæringen og lærte elevene å skrive på digitalt nettbrett, og åtte skoler lærte barna å skrive både for hånd og på digitalt nettbrett. Elevenes tekster ble analysert for lengde og kvalitet gjennom formell vurdering av et sett av teksttrekk knyttet både til transkripsjon (staving, mellomromsbruk og tegnsetting) og narrativ kompleksitet (vokabular, syntaks og narrative strukturer). Tekstkvalitetsmålene ble utviklet spesielt for å vurdere begynnerskriveres fortellinger, som typisk er korte og enkle. Den

statistiske analysen ble gjort gjennom Bayesianske metoder, som kan bevise både tilstedeværelse og fravær av effekter.

Avhandlingen inkluderer fire artikler. Artikkel 1 er en vitenskapsteoretisk diskusjon av hvordan tekster av begynnerskrivere kan analyseres fra et kvantitativt perspektiv. De tre resterende artiklene bidrar til avhandlingen gjennom å empirisk undersøke modalitetseffekter på førsteklassingers prestasjon i og læring av tekstkomposisjon. Artikkel 2 gir evidens for at førsteklassinger, som fra starten av første klasse lærer å skrive i begge modaliteter, etter all sannsynlighet produserer fortellinger av lik lengde og kvalitet i begge modaliteter. Denne artikkelen viser også at mangelen på en modalitetseffekt på prestasjon i tekstkomposisjon ikke avhenger av elevenes literacyferdigheter. For eksempel skrev ikke elever med svakere literacyferdigheter fortellinger av høyere kvalitet i en av modalitetene.

Artikkel 3 viser at førsteklasseelever som får undervisning basert på enten håndskrift eller digitalt nettbrett, med ellers minimal forandring i undervisningen, i hovedsak lærer å komponere tekster i samme takt gjennom det første året med skriveopplæring. Elevene viste lik utvikling av tekstlengde, syntaktisk kompleksitet og nøyaktighet og narrative strukturer, uavhengig av om de lærte å skrive for hånd eller på digitalt nettbrett. Elever som skrev på nettbrett, presterte bedre på transkripsjonsnøyaktighet (stave-, mellomroms- og tegnsettingsnøyaktighet), men viste liten eller ingen utvikling av disse teksttrekkene gjennom førsteklasse. Elever som skrev for hånd, startet på et lavere nivå i transkripsjonsnøyaktighet, men viste utvikling gjennom året. Denne forskjellen i prestasjon kan sannsynligvis tilskrives tekst-til-tale funksjonaliteten på de digitale nettbrettene. Artikkel 4 viser at det ikke var noen interaksjonseffekter mellom modalitet og elevenes literacyferdigheter på læring av tekstkomposisjon. Det vil si at det var, for eksempel, ingen fordeler knyttet til å lære å komponere tekst på digitalt nettbrett, eller for hånd, for elever med svakere literacyferdigheter.

Konklusjonen i avhandlingen er at, i en kontekst lik den som er studert her, påvirker ikke modalitet førsteklassingers prestasjon i tekstkomposisjon eller læring av tekstkomposisjon i vesentlig grad. Det ser altså ut som at skriveopplæring basert på håndskrift og skriveopplæring basert på nettbrett kan gi elever like muligheter for å utvikle ferdigheter i tekstkomposisjon det første året på skolen. Før klare anbefalinger om bruk av modalitet i begynneropplæring kan gis, bør framtidig forskning undersøke mulige overgangseffekter i å gå fra å lære å skrive i en modalitet til den andre modaliteten.

Table of Contents

Acknowledgements iii

Summary v

Sammendrag vii

1 Introduction 13

1.1 Background and knowledge needs 13

1.2 The present thesis 20

1.3 Articles of the thesis 22

1.3.1 The individual articles in this thesis 22

1.3.2 Relationship between the research questions and the articles 23

1.4 Key concepts 26

1.4.1 Written composition performance and learning 26

1.4.2 Modality 27

1.4.3 Transcription and inscription 28

1.4.4 First graders 28

1.5 Outline of the thesis 28

2 Theoretical foundation 30

2.1 Transcription in different modalities 30

2.2 The link between transcription and text quality 31

2.3 Effects of text-to-speech support 34

2.4 Learning to compose in different modalities 35

2.5 Do modality effects depend on children's literacy skills? 36

3 Methodological perspectives and reflections 38

3.1 The Norwegian context 38

3.2 The DigiHand project 39

3.3 Design and participants 40

3.3.1 Conditions 41

3.3.2 Treatment fidelity 43

3.4 Writing tasks and procedures for assessment 44

3.4.1 Developing writing tasks 44

3.4.2 Written composition assessment (procedures) 45

3.5 Text analysis – foundational issues and measures applied 46

3.5.1 Philosophical stance 47
3.5.2 Measurement of meaningful texts 48
3.5.3 Text quality assessment – rationale 49
3.5.4 Text quality measures – selection and operationalisation 50
3.6 Literacy related measures 65
3.7 Statistical methods 65
3.7.1 Bayesian statistics 65
3.7.2 Reasons for choosing the Bayesian approach 66
3.7.3 The Bayesian data analysis 67
3.7.4 Bayesian methods and philosophy of science 69
3.7.5 Mixed-effects modelling 71
3.8 Ethical considerations 72
4 Findings 75
4.1 Overall findings 75
4.2 Research question 1: Modality effects on performance 76
4.3 Research question 2: Interaction effects on performance 77
4.4 Research question 3: Modality effects on learning to compose 77
4.5 Research question 4: Interaction effects on learning to compose 78
5 Discussion 79
5.1 Limitations and recommendations for future research 86
5.2 Implications 90
5.2.1 Implications for theory and methodology 90
5.2.2 Implications for practice 92
5.3 Conclusion 95
References 98
Article 1 114
Article 2 135
Article 3 161
Article 4 162
Appendices 163
Appendix 1 – Author contributions of the four articles 163
Appendix 2 – Writing assessment tasks 165

Appendix 3 – Transcription and coding manual 170
Appendix 4 – Approval from the Norwegian Centre for Research Data..... 201

Table of Figures

Figure 1. Illustration of how the research questions are related to the articles 25

List of Tables

Table 1. Transcription of Texts..... 52
Table 2. Coding of Texts 53
Table 3. Examples of narrative coding 61
Table 4. Example texts rated for holistic quality 63

Appendices

Appendix 1 – Author contributions of the four articles 163
Appendix 2 – Writing assessment tasks..... 165
Appendix 3 – Transcription and coding manual..... 170
Appendix 4 – Approval from the Norwegian Centre for Research Data..... 201

1 Introduction

1.1 Background and knowledge needs

The use of digital technology is widespread in educational systems in developed countries. Recently, there has been an increase in 1:1 computing implementations, where students are equipped with personal digital devices, such as digital tablets or computers, to be used in school (Islam & Grönlund, 2016). In Scandinavia, there has been a turn towards introducing personal digital devices in schools even from the first grade and using such devices in initial writing instruction (Andresen, 2017; Gamlem et al., 2020; Genlott & Grönlund, 2013). This change means that traditional writing instruction with pencil and paper is no longer the only existing practice in first-grade classrooms. Other practices involve postponing handwriting and teaching writing by keyboard first or teaching children handwriting and keyboarding in parallel.

The aim of this thesis is to investigate the effects of these three practices on first graders' written composition performance and written composition learning. Broadly, I address two questions. First, I explore whether modality – writing by hand or writing by keyboard on a digital tablet – affects the length and quality of first-grade students' compositions. Second, I investigate whether one of the modalities better supports first graders' learning of written composition – whether learning to compose text in one modality leads to faster written composition learning compared to learning to compose text in the other modality.

In Norway, where this PhD study has been conducted, children are taught both handwriting and keyboarding in school. The curriculum for Norwegian (literacy) prescribes that, by the end of second grade, students shall already be able to compose simple texts by hand and using a keyboard (Ministry of Education and Research, 2019). The curriculum, however, does not prescribe the order of teaching children to write in the two modalities, with the result that some schools teach handwriting first,

some teach typing first and some teach both handwriting and keyboarding simultaneously.

In the last few years, there has been a large increase in schools buying digital devices for all students, including from the first grade. There is no complete overview of how many Norwegian schools have implemented 1:1 computing, as such decisions are made by local educational authorities without national control. A survey of the 100 largest municipalities in Norway shows that in the academic year 2021–2022, 81% of the students in primary and lower secondary schools in these municipalities had their own digital device provided by the school (UiO, 2021).

In general, the motivation for providing students with personal digital devices in school is to enhance learning and equip students with the skills necessary to participate in the rapidly changing and highly digital society of the 21st century (Islam & Grönlund, 2016; Ricoy & Sánchez-Martínez, 2020). A more specific argument for applying digital devices in initial writing instruction is that typing, which entails easier motor actions than handwriting, lets students concentrate on the cognitive processes in writing, i.e. spelling, while the training of the motor skills necessary for handwriting can be postponed (Genlott & Grönlund, 2013; Trageton, 2003, pp. 80–81). Also, several applications with text-to-speech functionality directed towards beginning writers have been developed, for example Skoleskrift (Ohlis, 2018), Lingit STL+ (Lingit, n.d.) and IntoWords (Vitec, n.d.). The text-to-speech functionality pronounces letter sounds, words and sentences while students write, and this immediate feedback is argued to support the spelling of students who are learning to write (Genlott & Grönlund, 2013).

Studying written composition in first graders is important. For most children, the ability to commit words on paper or screen develops when they enter school (or, in some educational systems, kindergarten). Initial writing instruction usually focuses on transcription skills, namely, the ability to spell words and output these words in written language.

Additionally, initial writing instruction will involve composition of simple texts – where children have to learn to generate ideas, translate these into linguistic entities and transcribe these sentences as coherent text. Mastering such basic writing skills is necessary to develop more advanced writing skills, and longitudinal studies have found that children's early writing skills are predictive of later writing performance (Juel, 1988; Kim et al., 2015). In general, much of what children are to learn in school depends on having good writing skills. Therefore, the early formal phase of children's writing development deserves to be studied, and in times of increasing digitalisation, it is particularly necessary to investigate the modality effects on children's writing.

Previous studies of modality effects on early writing outcomes have different foci. Some studies have compared writing performance in the two modalities measured at a single time point without any preceding intervention. Other studies have compared the effects of writing instruction based on different modalities on children's learning, either of written composition or of low-level skills, such as writing letters or words.

The studies that systematically have compared modality effects on writing performance at a single point in time sampling children from the early primary grades typically apply measures of fluency, speed or productivity. In this thesis, only measures of production (number of words written) and text quality are applied. However, studies of fluency are not irrelevant – if there are differential effects of modality on length or quality, fluency might be the mechanism by which such effects occur. In a study of children from reception class to Year 6 (N = 312), Connelly et al. (2007) found that, across all grades, children produced more correct letters in less time when handwriting compared to typing in a sentence-copying task. Berninger et al. (2009) found that second, fourth and sixth graders produced longer essays with faster word-production rates by hand than when typing. In both of these studies, however, the children had done most of their previous writing by hand. Further, Crook and Bennet (2007) found that second graders (N = 72) produced text more

quickly when copying text by hand, even though they all had experience using computers in class.

Very few studies have systematically compared modality effects on text quality in primary-grade children independently of any writing intervention, and to my knowledge, no such studies have examined modality effects in first graders. Connelly et al. (2007) found that Year 5 and 6 children (N = 48) produced texts of higher quality scored analytically for both surface and substantive features when writing by hand, compared to when using a keyboard. However, these children were more experienced in handwriting. Read (2007) found in a pilot study (N = 18) that 7- and 8-year-old students wrote texts that were longer and rated as better by the teacher when writing by hand than by typing. These children had, however, not written on a keyboard before. In another small-scale study (N = 16), fourth-grade children who had experience in writing in both modalities wrote texts with higher linguistic correctness, but not better structural content or more words, when writing on a digital tablet compared to writing by hand (Dahlström & Boström, 2017).

Studies that have investigated the effects of writing instruction built on different modalities can be divided into two main groups: studies that evaluate the effects of writing instruction based on either handwriting or keyboarding with minimal additional change in instruction and studies that evaluate specific computer-assisted learning interventions. Only the former group is considered in this thesis. A few studies have examined primary-grade classrooms where early writing skills have been established through handwriting instruction and word processors have been introduced in third, fourth or fifth grade, for periods spanning from six months to three years. Moore and Turner (1988) and Owston and Wideman (1997) found that texts improved more in terms of holistic quality when students were trained in typing compared to handwriting (N = 204 and 110, respectively). Dybdahl et al. (1997), on the other hand, found no differences in holistic quality between children (N = 47) receiving handwriting or typing instruction.

Two early studies looked at modality effects on learning to compose in first-grade children, though not from the very start of the school year, and found benefits for typing. Larter et al. (1987) compared 60 first graders who, for the second semester of first grade, received writing instruction based on either handwriting or typing. All students had been taught traditional handwriting in the first semester. The results showed that children in the typing condition produced texts that were longer and of better holistic quality than students in the handwriting condition. Jones and Pellegrini (1996) investigated the development of first graders' narratives in students' writing both by hand and by keyboard through a 10-week programme, starting in the winter of first grade. When typing, during instruction and at assessment, children had the opportunity to use vocabulary support and text-to-speech to have their compositions read aloud. The results showed that narratives written digitally were lexically denser and more grammatically and lexically cohesive than narratives composed by hand, while the texts did not differ on measures of endophora, temporal conjunctions or narrative structure. A more recent study, Genlott and Grönlund (2016), investigated students (N = 255) who, from the start of first grade, received writing instruction based either on handwriting methods or on personal digital devices in otherwise business-as-usual classrooms. Scores on national literacy tests in the third grade showed no differences between the conditions. However, this study did not report specific measures of written composition.

In sum, studies of the modality effects of writing instruction based on different modalities have tended to indicate that children have produced better texts when the instruction is based on typing. However, these studies sampled students who had already received some amount of formal literacy instruction by hand. It is also worth to note that many of the studies were conducted several years ago. Newer practices of initial writing instruction which include text-to-speech have scarcely been investigated. The study of Jones and Pellegrini (1996) referred to above compared handwriting instruction and typing instruction that

allowed for read-back of written text, but in combination with vocabulary support. A recent study where two first- and second-grade classrooms were observed for two days indicates that students seem to focus on orthography rather than content while writing on a digital tablet with text-to-speech support (Bjørkvold & Svanes, 2021). A few studies have compared effects of typing to typing with text-to-speech on written composition in students older than first graders. No clear benefits have been found for productivity or text quality (Borgh & Patrick Dickson, 1992; Dahlström & Boström, 2017), but there is some evidence that text-to-speech in combination with word prediction might support spelling accuracy in writers with spelling difficulties (MacArthur, 1998, 1999).

A body of studies has explored the effects of modality on children's early writing acquisition. These studies have typically sampled preschool children and assessed modality effects on learning to recognise or write letters or words. Longcamp et al. (2005) and Mayer et al. (2020) found that handwriting training led to better letter recognition performance in young children aged 3–6 compared to keyboard training. Other studies, however, have not found significant differences in the letter recognition performance of 4–6-year-old children after training in different modalities (Duiser et al., 2022; Kiefer et al., 2015). In addition, two studies comparing the effects of training in different modalities on word-writing ability in kindergarteners differed in their findings. Kiefer et al. (2015) found that, in a group of 23 kindergarteners, the handwriting group outperformed the keyboarding group in a word-writing-accuracy task tested in the trained modality. In a larger sample of 147 kindergarten children, Mayer et al. (2020) did not find that handwriting training resulted in better word-writing-accuracy performance than keyboarding training.

Similarly, the results of modality effects on spelling in primary-grade children show either benefits for handwriting or no difference between modalities. For example, Cunningham and Stanovich (1990) found that, in first-grade children, learning to spell real words by hand led to better spelling performance than learning by typing. On the other

hand, Ouelette and Tims (2014) found no difference in spelling performance among second-grade students who were trained in either printing or typing. This study controlled for children's prior knowledge of words by training their ability to spell non-words. If handwriting training can result in better letter learning or spelling of words, it might be that at the very start of formal literacy development, learning composition by hand is beneficial.

It is possible that the effects of modality on written composition performance and on learning to compose text are not the same for all children but that they, for example, depend on children's literacy skills. A body of studies has investigated the relationship between literacy skills and writing performance. For kindergarten and first-grade children, there is evidence that skills in spelling, reading and oral language are associated with children's productivity/fluency, for example, number of words produced (Abbott & Berninger, 1993; Kent et al., 2014; Kim et al., 2011, 2014), and writing quality (Abbott & Berninger, 1993; Juel et al., 1986; Kent et al., 2014; Kim et al., 2013, 2014). There is also some evidence that spelling ability, word-reading ability, and letter knowledge can predict the rate at which first graders' learn to compose text (Torrance et al., 2021). Therefore, it is possible that modality and literacy skills might have a combined effect on written composition performance and learning.

The above summary illustrates that previous studies of modality effects on children's writing show mixed results. The studies of modality effects of writing instruction using different modalities have typically found more gains in quality when children have been trained by typing, but these studies have not sampled students from the very start of school and they are quite old, which means that newer practices that include text-to-speech when children type are scarcely explored. On the other hand, research on modality effects on learning low-level skills in writing has found advantages for handwriting or no modality differences, but these studies have typically investigated children younger than first graders. Studies that have looked at modality effects on written

composition performance at a single point in time have tended to find advantages for handwriting, but most of these have sampled students who have already written a substantial amount by hand and less by typing. Wollscheid et al. (2016) pointed out this limitation in their review of studies of modality effects on early writing outcomes, together with other methodological problems in many existing studies: lack of control for nesting effects and small sample sizes below 50 participants. These three methodological issues were considered when designing the present PhD project.

1.2 The present thesis

The research reported in the present thesis explores modality effects on written composition in first graders exposed to three different approaches to initial writing instruction: learning to write by keyboard on a digital tablet, learning to write by handwriting, and learning to write by hand and by keyboard on a digital tablet in parallel. This was possible because the PhD project was part of the larger *DigiHand* project (Gamlem et al., 2020), which had recruited a relatively large sample of schools representing these writing instruction practices. Thus, the research reported here involved no researcher-made intervention concerning how typing can be taught from the start of school or how specific software can be used to enhance written composition skills. Rather, the present thesis investigates written composition in first graders' writing predominantly using a keyboard, predominantly by hand or in both modalities in otherwise business-as-usual classrooms. As this study is a natural experiment (quasi-experiment; Bordens & Abbott, 2005, p. 305–308), the digital writing reflects an established practice in Norwegian schools, namely with digital tablets with touchscreen keyboards and with applications that offer text-to-speech support.

The thesis investigates modality effects on written composition by studying the effects on both written composition performance and written composition learning. First, this study investigates whether one

of the modalities can better support the text production of writers who are at the very start of their formal literacy development. Second, as first graders are learning to write, their improvement in composition ability is likely to be substantial in the first school year. It is therefore of interest to investigate whether modality affects the rate at which their compositional writing grows. Unlike previous studies of modality effects on written composition in young writers, the present study samples students from the very start of school, thereby minimising possible problems with children being more experienced in handwriting.

An important contribution of the present thesis is the assessment of children's narratives using a text-analytic approach. The quality of the students' compositions was measured using a set of text features specifically aimed at describing short and inaccurate texts typically produced by beginning writers. A text-analytic approach promotes explicit and transparent evaluation and makes it possible to investigate whether some, and not all, features of texts are affected by modality.

Another strength of the present work is that Bayesian methods were used in the statistical analysis. These permit robust inferences, both in the presence and absence of a true effect. As is clear from the summary of previous research, there is a possibility that the true modality effect is zero, and it is therefore of value to use statistical methods that allow for finding evidence in favour of both the alternative and null hypotheses.

This thesis aims to answer the following overall research question: *Does modality – handwriting with pencil on paper or keyboarding on a digital tablet with text-to-speech functionality – affect first graders' written composition performance and written composition learning?*

This overarching question is unpacked through four subordinate research questions:

1. Does modality affect written composition performance in first graders?

2. If modality has an effect on first-grade children's composition performance, is this dependent on children's literacy skills? For example, do children with weaker literacy skills produce compositions of better quality when composing in one or another modality?
3. Does modality affect first graders' learning of written composition?
4. If modality has an effect on first-grade children's written composition learning, is this dependent on children's literacy skills? For example, do children with better literacy skills learn written composition faster when learning to write in one or another modality?

These four research questions are answered in three empirical articles. In addition, the thesis includes an article that discusses fundamental questions concerning text analysis in quantitative studies. The next section presents the articles and shows how they together answer the research questions of the thesis.

1.3 Articles of the thesis

1.3.1 The individual articles in this thesis[1]

Article 1: Spilling, E. F. (2021). The measurement of text quality. In T. A. Haugen, S.-A. Myklebost, S. J. Helset, & E. Brunstad (Eds.), *Språk, tekst og medvit* [*Language, text, and consciousness*] (pp. 47–66). Cappelen Damm Akademisk. https://doi.org/10.23865/noasp.146

[1] Author contributions of each article can be found in Appendix 1.

Article 2: Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., and Torrance, M. (2022). Handwriting versus keyboarding: Does writing modality affect quality of narratives written by beginning writers? *Reading and Writing*, *35*(1), 129–153. https://link.springer.com/article/10.1007/s11145-021-10169-y

Article 3: Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., and Torrance, M. (2023). Writing by hand or by keyboard: Does modality affect rate of learning to compose text in first grade? Under review, 2nd revision, *Computers & Education*.

Article 4: Spilling, E. F. (2023). Effects of literacy skills on learning to compose narratives: A comparison of children writing by hand and by keyboard. Manuscript prepared for submission.

1.3.2 Relationship between the research questions and the articles

Figure 1 shows the relationship between the research questions and the articles included in this thesis. It also displays the sample and design of each empirical article.

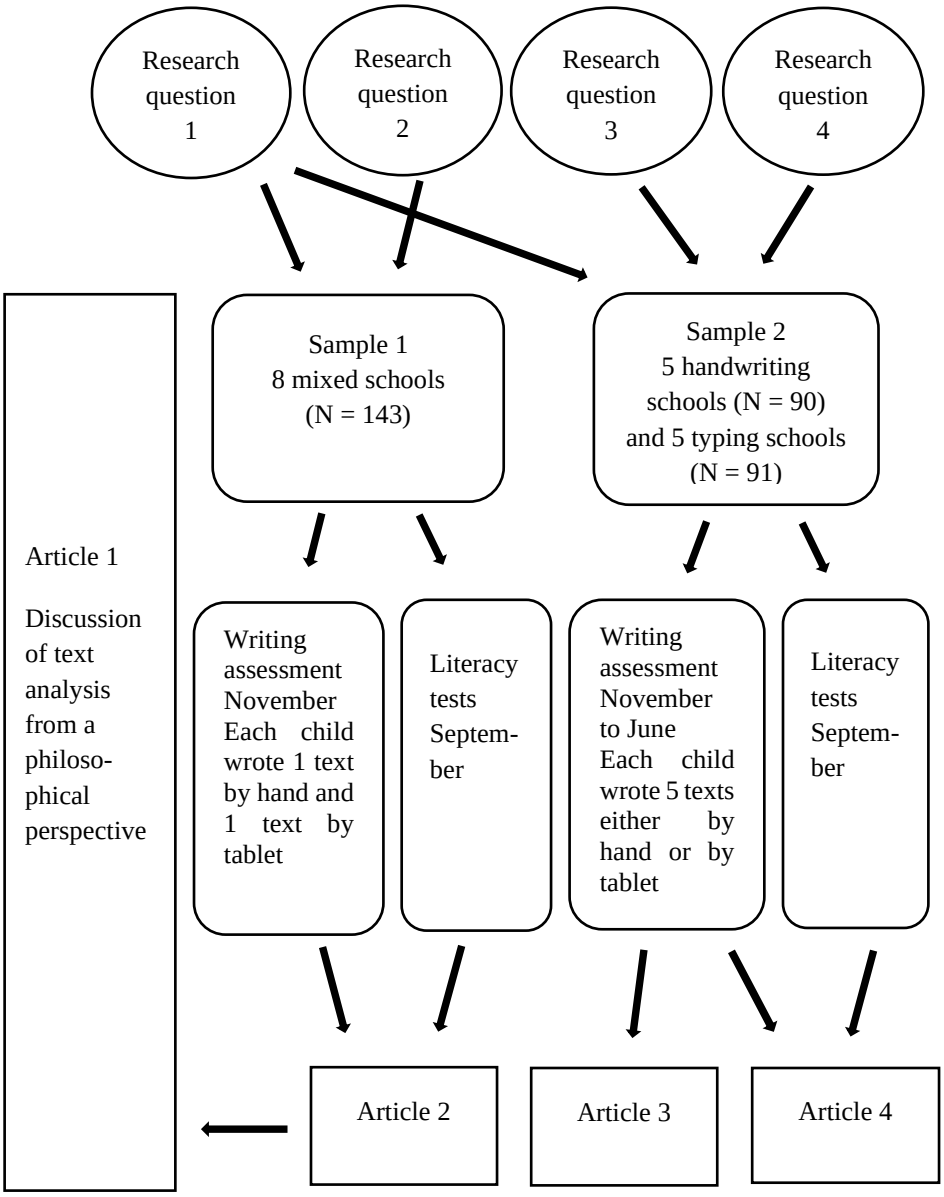
Article 1, placed in the margin of Figure 1, is a discussion of text analysis from a philosophical viewpoint. The thesis is grounded in a naturalistic, postpositivist worldview (cf. Section 3.5.1), and consistent with this, the three empirical articles have a quantitative design. The philosophical worldview also has consequences for the text analysis; for example, the texts should be analysed systematically and objectively. However, as texts are products of meaning expressed linguistically and beginning writers typically produce texts that are rudimentary, it is worth discussing how an analysis of beginning writers' texts can be conducted objectively. Article 1 discusses, with a philosophical backdrop, how I approached the texts in the empirical articles and the challenges involved

in this process. Article 1 directly discusses Article 2, and the text examples used were written by students in Sample 1 (hence, the arrow from Article 2 to Article 1 in the figure).

Research question 1 of this thesis concerns the modality effects on written composition performance, which was first investigated in a sample of students who were taught to write by hand and use a digital tablet in parallel ('mixed schools'). These children, whom I call Sample 1, were tested for literacy skills at school entry and their written composition performance was measured in both modalities in November. As these children learned to write in both modalities, Article 2 investigated whether modality per se affects written composition performance in children who are at the very beginning of their formal writing development.

Modality effects on written composition performance were further investigated in Sample 2, which consisted of five schools that taught children to write by hand ('handwriting schools') and five schools that taught children to write by typing on a digital tablet ('typing schools'). The written composition performance of these students was tested at five time points throughout the first grade in the modality in which they learned to write, the first time point concurrently with the students in Sample 1. These data provide knowledge about the effects of modality on written composition performance in students who learn to write in one modality. If a difference in written composition performance between students in handwriting and typing schools is found, this might be a result of both modality effects at assessment and modality effects on learning to write. For example, if students in typing classrooms perform better than students in handwriting classrooms at a specific point in time, this might be due to a modality effect on how students write on the specific assessment occasion, but also because of a modality effect on learning to write up to the time of testing. Modality effects on written composition performance in handwriting and typing classrooms are reported in Article 3.

Figure 1. Illustration of how the research questions are related to the articles



Research question 2 explores whether the modality effects on written composition performance depend on (interact with) children's literacy skills. This was investigated in Sample 1, where the data from the literacy tests at school start were combined with the writing performance data. This analysis is presented in Article 2.

Research question 3 examines whether modality affects first-grade children's learning of written composition. It was important to sample children from handwriting schools and typing schools when answering this research question, because a between-subject design makes it possible to isolate the effect of each modality on children's learning to compose. If students from mixed schools were sampled, it would not have been possible to isolate the effect of each modality on children's learning of written composition, because both modalities were allowed to affect the learning simultaneously. Research question 3 is answered by Article 3, which used the written composition performance data from the five time points and investigated the rate at which performance improved for children learning to write in the two modalities.

Research question 4 concerns whether the modality effects on learning to compose text depend on children's literacy skills. This was investigated in Sample 2 and Article 4. Article 4 used the same data as Article 3 – writing samples from five time points – but added the scores on the literacy tests at school entry. Therefore, this article could answer whether the modality effects on learning to compose text explored in Article 3 are the same for all children independently of their literacy skills.

1.4 Key concepts

1.4.1 Written composition performance and learning

In this thesis, I study children's writing and use the term written composition to refer to meaningful text that consists of more than single

graphemes or a few single words, or, more precisely, a written coherent multi-sentence text. For first graders, written composition typically means a narrative text, which is also why the writing tasks used in this work invited children to write stories. Being able to compose a narrative requires a child to master several writing skills in parallel: to generate and structure ideas in line with genre conventions, to formulate accurate and complex clauses, to apply appropriate words and to spell, space and punctuate the text.

In this thesis, children's written composition performance is measured by the length and quality of the composition children produce on a specific writing assessment occasion. The quality was assessed analytically by scoring a set of text features related to narrative composition (vocabulary, syntax and narrative structures) and transcription (spacing, spelling and punctuation).

The children's learning to compose text was measured by changes in written composition performance over time. Children's written composition performance was tested on several occasions – the first chosen to be late November, when children had received approximately three months of instruction and one could expect that children would manage to produce simple narratives. The last test occasion was in June, just before the children were to finish the first school year. Therefore, children's written composition learning – the improvement in length and composition quality from the first to last test occasion – could be measured. Again, quality was assessed by a set of text features, reflecting the ability to generate and structure content, form accurate and complex clauses, apply appropriate words, and spell, space and punctuate text.

1.4.2 Modality

Modality refers to handwriting and keyboarding/typing, which are two output media or modes that can be used to transcribe ideas and oral language into written text. In this thesis, keyboarding was done using a

digital tablet with an electronic touchscreen keyboard. Writing both for instruction and assessment allowed for text-to-speech, which means that while writing students would hear both letter sounds, words and sentences read aloud (see Section 3.3.1 for more details). I use the terms *keyboarding* and *typing* interchangeably.

1.4.3 Transcription and inscription

Transcription refers to the process of both spelling and handwriting or keyboarding. Inscription refers to this latter component without specifying the modality – the process of motor planning and execution of the retrieved word-spellings (Rønneberg et al., 2022).

1.4.4 First graders

In Norway, the first grade is the first year of formal writing instruction. Children start school the year they turn six (see Section 3.1 for more details). I also refer to the first-grade students as beginning writers and young writers to make the text more readable. In contexts where it is important to separate first graders from younger or older students, I specify information about age or grade.

1.5 Outline of the thesis

This thesis consists of an extended abstract and four articles. The extended abstract presents the coherence of the thesis by compiling, elaborating and discussing the research questions, methodological aspects and findings from the individual articles. The extended abstract is organised into five chapters. This first chapter has outlined a brief background to the thesis and introduced the research questions of the thesis. Chapter 2 presents some theoretical perspectives that might explain why modality affects children's written composition performance and learning. In Chapter 3, I discuss the methodological choices, present the philosophical foundation on which this thesis rests

and introduce the Bayesian framework used for the statistical analysis. Chapter 4 provides a summary of the findings of the three empirical articles that answer the research questions. In Chapter 5, the findings are discussed before the limitations, recommendations for future research and implications are outlined. Finally, the four articles are presented in their full form.

2 Theoretical foundation

In the following, I first point out how handwriting and keyboarding differ in their processing demands. Then, I discuss some theoretical perspectives that contribute to understanding why modality might affect first graders' written composition performance and written composition learning.

2.1 Transcription in different modalities

Many of the underlying processes in composing text are the same, regardless of which modality is used to output the text. The inscription process, however, is different when composing by hand compared to composing by keyboard.

Van Galen (1991) described the process of writing by hand as a set of processing modules that each address a specific feature of the message. The modules are structured hierarchically, with the output from the preceding module functioning as input for the next lower module. The three top modules concern the activation of intentions (ideas), retrieval of relevant semantic concepts and syntactical construction. Next comes the spelling stage, which involves finding a graphemic representation of the relevant words. Further, the writer needs to select allographs – variants of the graphemes – that represent the actual shape the grapheme will take. The selection of allographs is then followed by a motor programme that involves finding the appropriate letter size (the size control module) and outputting on paper the letters through specific letter strokes (the muscular adjustment stage). In this model, all modules work concurrently. Higher levels, which have passed on input to the lower stages, start processing related to the forthcoming parts of the message (van Galen, 1991).

Typing differs from handwriting in both the cognitive and the motor processes related to inscription – the processing that in the model of van Galen (1991) is at the stage of selection of allographs or lower.

When typing, the writer has access to external representations of the letters on the keyboard, which might cue retrieval. For beginning writers, it might be less demanding to recognise letters than to retrieve them from memory. On the other hand, when selecting the correct letter, a child has to be able to ignore the other, potentially distracting letters also found on the keyboard.

Handwriting presupposes having precise muscle control of the fingers, hand and arm to produce letters that consist of specific combinations of strokes (Dinehart, 2015; van Galen, 1991). For beginning writers, who typically type using one or two (index) fingers, writing by keyboard involves less fine motor control and the same motor action for producing all letters (Beschorner & Hutchison, 2013; Mangen & Velay, 2010). Acquiring keyboarding fluency, for example, writing through the touch-keyboarding method, involves learning more complex processes than typing with one finger (Freeman et al., 2005; Perminger et al., 2004).

Writing digitally can offer spelling support through spellchecking or text-to-speech technology. In the research reported in the present thesis, students writing by keyboard could use text-to-speech, which provided read-back while writing. It is not clear whether text-to-speech supports transcription in beginning writers or whether it might distract the writers (which I return to below in Section 2.3).

2.2 The link between transcription and text quality

As described by Juel and colleagues (1986) in the simple view of writing, the production of written text must, at a minimum, include the abilities to generate ideas and to output these as written language. If children cannot spell or form/find letters, they will not be able to express their ideas in written form. When struggling with inscription or spelling, this may lead to production of shorter texts. A very short text will probably not have well-developed ideas or include relevant details that may be important features of good texts. There is evidence that text length

correlates with holistic rating – that shorter texts typically receive lower quality scores than longer texts (Charney, 1984; Froese, 1989). The evaluation of text quality will in many cases directly or indirectly involve the assessment of handwriting and spelling, and struggling with these skills may therefore influence the quality rating negatively. If typing, compared to handwriting, makes transcription easier, this might help students write more words and produce neater letters, which again might lead to better quality in the compositions they produce.

There are also theories that link transcription skills to text quality based on the capacity-sharing hypothesis. Capacity theory understands writing as the coordination of various processes that draw on the same cognitive resources, and increased costs in one process will reduce resources available for other processes (McCutchen, 1996; Torrance & Galbraith, 2006). A problem with capacity explanations of writing processes is that they to a limited extent fulfil the criteria of falsifiability; that is, they can be used to explain almost any pattern observed in the data (Torrance & Galbraith, 2006). Still, the capacity-sharing idea in some form is incorporated in several theories of developmental writing, for example, the not-so-simple view of writing (Berninger & Winn, 2006) and the direct and indirect effects model of writing (DIEW; Kim & Park, 2019; Kim & Schatschneider, 2017).

The not-so-simple view of writing expands the simple view of writing by including executive functions (e.g. attention, planning, revising and self-monitoring) in addition to text generation (ideation) and transcription (spelling and handwriting/keyboarding). This model also posits that these three components are coordinated by working memory. Long-term memory is activated during planning, composing, reviewing and revising, and short-term memory is activated during reviewing and revising output (Berninger & Winn, 2006).

The DIEW builds on the simple and not-so-simple view of writing and extends them by including higher-order cognitive skills, background knowledge, and affect and motivation as components

contributing to writing development and by specifying the structural relationships between the component skills.

DIEW specifies two large categories of skills: discourse oral language (corresponding to ideation/text generation in the simple and not-so-simple view of writing) and transcription skills. Discourse oral language is hypothesised to draw on both foundational oral language skills (e.g. vocabulary and grammatical knowledge) and higher-order cognitive skills (e.g. reasoning, monitoring). Transcription skills depend on knowledge of phonology, orthography and semantics. The DIEW postulates that component skills are related in a hierarchy where higher-order skills depend on lower-order skills, and the effect of lower-order skills is partly or completely mediated by higher-order skills. For example, executive functions are seen as foundational for both discourse oral language and transcription skills. Further, it is assumed that even though the components of DIEW all contribute to overall writing development, their contribution will vary depending on both the developmental phase and the writing dimension. It is, for example, assumed that young writers use their mental resources primarily for transcription, but with the increasing automatisation of transcription skills, more resources can support processes related to discourse oral language skills (Kim & Park, 2019).

From a capacity-sharing perspective, the low-level processes of transcription put demands on children to such an extent that less attention is left for higher-level processes. If one of the modalities, for example, typing, places less costly inscription demands on the writers, it is possible that more resources can be freed to generate ideas, form accurate clauses, select appropriate words and spell more accurately.

It may also be that the relationship between children's transcription skills and the quality of their compositions is not causal. For example, Rønneberg et al. (2022) investigated the process-disruption hypothesis – the assumption that lack of fluency in spelling and/or typing leads to disfluency in producing text, which again may damage the quality of the final product. They found that sixth-grade students with

weaker typing abilities hesitated more when producing text, but they did not find evidence that this affected the quality of the resulting compositions. The authors discussed possible explanations, one of which is that motor planning and spelling retrieval do not share processing resources with other upstream (higher-level) processes. Further, they argued that there might exist child-level factors that can explain both transcription skills and the ability to produce high-quality compositions.

2.3 Effects of text-to-speech support

As noted in the Introduction there is limited research on the effects of text-to-speech on written composition. It has been argued that this functionality is beneficial for beginning writers by providing immediate feedback on spelling while children write (Genlott & Grönlund, 2013). However, it has been observed that text-to-speech might disturb children's writing and that children do not necessarily manage to correct the spelling errors they detect through the text-to-speech, which might cause frustration (Bjørkvold & Svanes, 2021). If text-to-speech functionality relieves the burden of spelling, more resources can be deployed to writing more words, or – as indicated by capacity theory (McCutchen, 1996) – to develop substantive features when composing. On the other hand, it is not clear that text-to-speech functionality supports the very production of words. This functionality does not necessarily make the spelling process less costly or more fluent. Rather, using text-to-speech might lead to an increased focus on spelling words accurately. This could mean that resources that otherwise can be deployed for developing high-level text features are occupied for monitoring spelling. Similarly, increased attention towards correct spelling may not entail increased production fluency. As Rønneberg et al. (2022) argued, misspellings are not necessarily produced disfluently, and correctly spelled words are not necessarily produced fluently. Hearing (misspelled) words read aloud while composing might disrupt

other ongoing processes, which again potentially might have negative effects on the final product.

2.4 Learning to compose in different modalities

Learning to write means learning several complex component skills, cf. the simple view of writing, the not-so-simple view of writing and the DIEW briefly presented above. For most children, learning to compose text requires explicit instruction (Kellogg, 2008). As described by Rijlaarsdam and Couzijn (2000), beginning writers are faced with a double task when composing – both producing a composition and simultaneously developing their writing skills. If writers focus their attention on forming or finding letters, there might be few resources left for learning spelling and higher-level processes. If one modality can relieve the inscription demands and more attention can be directed towards developing writing skills, the rate at which children learn might be affected positively. In addition, if the inscription is supported in one modality, children might produce more text. Learning to apply higher-level features in composition, such as organising text on the macro level, presupposes the ability to write texts of some length. In addition, being able to write more words might enhance students' motivation for writing, again positively affecting their learning. Additionally, longer compositions might yield more teacher feedback from which children can learn.

In a natural classroom setting, modality has potential to influence how writing is taught. For example, when children learn to write by hand, they must practice letter formation. If initial writing instruction is based on keyboard, time that otherwise would have been spent on letter formation might be used on other writing activities. However, changing instruction to exploit the potential offered by digital affordances might require much from the teacher, both in terms of motivation and competence. From previous research there is an indication that introducing digital tablets in primary-grade classrooms to a limited

extent change the classroom practices compared to traditional teaching practices (Ricoy & Sánchez-Martínez, 2020).

2.5 Do modality effects depend on children's literacy skills?

It is possible that modality affects written composition performance and learning differentially, for example, depending on children's literacy skills. As suggested by the models of writing cited above there is a range of skills that directly or indirectly may contribute to writing. For example, phonological awareness may be a prerequisite for being able to spell (Juel et al., 1986), and vocabulary skills may be central for generating content (Kim & Park, 2019). It might be that the effects of modality interact with the effects of children's literacy skills on their written composition performance. If inscription is easier in one of the modalities, children with weak letter knowledge may be supported in their text production in this modality. If a child struggles to remember letter shapes, visual cues on the keyboard might make it easier to produce words. Also, if typing can free resources for higher-level processing through easier inscription demands (cf. McCutchen, 1996), children with well-developed skills might better exploit their literacy skills to produce good compositions when typing.

Hypothetically, it might be that when learning written composition, the combination of children's literacy skills and the modality in which they learn to write influence the rate at which they learn to compose. If one of the modalities is more resource demanding, children who start school with weak inscription skills may be disproportionately affected when learning to write in this modality. They will struggle more with inscription and will have fewer resources left for learning higher-level composition skills (cf. Rijlaarsdam & Couzijn, 2000). Another example could be children with well-developed literacy skills who might take advantage of these skills if inscription demands are

relieved. Potentially, for these students, writing in this modality might result in faster learning of composition.

3 Methodological perspectives and reflections

Each of the empirical articles included in this thesis has detailed methodological accounts. In the following, I comment on the methodological aspects that concern the thesis as a whole and elaborate on aspects that require discussion. This section also presents the philosophy of science on which this thesis rests. I relate the philosophical stance to the account of the text analysis executed in this work, as the philosophical stance was decisive in how I analysed the texts. Further, some space is devoted to a presentation of Bayesian statistics, as this approach to statistical analysis is not as well known within writing research as the more mainstream approach of frequentist statistics. Finally, ethical aspects are considered.

3.1 The Norwegian context

Most Norwegian children attend kindergarten a few years before entering school (The Norwegian Directorate for Education and Training, 2020). The pedagogical content of kindergartens is nationally prescribed, and kindergartens are obliged to stimulate children's language development (The Norwegian Directorate for Education and Training, 2017). However, there is no formal writing instruction in kindergarten.

Children start school in August of the year they turn six. Usually, children go to their local state school. All education offered by state schools is free, and schools are financially resourced by the state. The content of primary-grade instruction is nationally prescribed, and so is the number of hours of teaching. As mentioned in Section 1.1, the Norwegian curriculum prescribes that children have to learn to compose texts by hand and keyboard by the end of the second grade, but the order of teaching writing in the two modalities is not specified.

The use of computers to support the writing of beginners has been argued for for some time through the reading-through-writing approach (Trageton, 2003). Recently, these thoughts have received renewed

interest along with the development of applications with text-to-speech technology that might support students' learning to write and with a large number of schools purchasing personal digital devices for their students.

The Norwegian language has relatively regular phoneme–grapheme mapping. The orthography is considered rather shallow, with Finnish being shallower but Swedish, French and English being deeper orthographies (Seymour et al., 2003). Norwegian has two written standards, Bokmål and Nynorsk, which are linguistically very similar. The core area of Nynorsk is in the western part of Norway, where the data were collected.

3.2 The DigiHand project

The present work has been part of the larger project, *DigiHand – The Emergence of Handwriting Skills in Digital Classrooms* (Gamlem et al., 2020). This project is a cooperation between Volda University College and the Reading Centre (Stavanger University) and financed by the Norwegian Research Council (grant number 273422). In total, 33 schools (and only one class from each school) from Western Norway participated in the project. To be included in the project, schools had to use Nynorsk as the main written standard in school, and there had to be a minimum of 10 students in the first-grade class. Further, schools applying three different approaches to initial writing instruction were sampled: (1) writing instruction based on a personal digital device postponing handwriting instruction (typing schools), (2) handwriting instruction based on traditional methods using pencil and paper (handwriting schools) and (3) a combination of handwriting instruction and instruction with a personal digital device (mixed schools). The decision to adopt a typing-first, handwriting-first or mixed-approach to writing instruction was made by the educational authority at the district (municipality) level. Thus, these three instructional practices represent naturally occurring practices. More details about the *DigiHand* project,

for example, about the sample of 33 schools, can be found in Gamlem et al. (2020).

3.3 Design and participants

The three empirical studies encompassing this thesis were developed within the frame of the *DigiHand* project. Including schools representing the three different writing instruction practices and applying both within-subjects design and between-subjects design made it possible to investigate modality effects on written composition from different perspectives (cf. Section 1.3.2). Of the 33 schools that took part in the *DigiHand* project, 18 comprised the sample in this thesis – eight mixed schools, five handwriting schools and five typing schools. See Figure 1 presented in the Introduction for a visual reminder of the sample and design of the articles. All students were tested for literacy skills in September, which was in their second to fifth week of school. Written composition assessment was done at five time points from November to June, with approximately seven-week intervals. Students in the mixed schools wrote two texts at the first time point, one by hand and one on a digital tablet. Students from the handwriting and typing schools wrote one text at each time point, all in the modality in which they learned to write.

The 18 schools were located in rural or semi-rural parts of Western Norway. The typing schools and the mixed schools came from regions where the educational authorities had decided that digital devices should be used from the start of school. These regions have implemented the use of digital devices from the first grade for pedagogical reasons, for example to enhance learning (cf. Section 1.1).

Students in the typing and handwriting schools made up one sample that was investigated in Articles 3 and 4. It was ensured that the two groups of schools were equivalent with regard to factors such as the mean class size and socioeconomic status of the students. Inclusion criteria for being part of the *DigiHand* project were the use of Nynorsk

as a written standard and more than ten students in the class. When I selected five handwriting schools and five typing schools from the overall *DigiHand* sample, the schools (classes) were pair-matched for class size. I also consulted official statistics from Statistics Norway (SSB, 2021) for the municipalities in which the typing and handwriting schools were located. These statistics indicated no meaningful differences in family income, adult educational attainment or educational expenditure per pupil between conditions. As mentioned in Section 3.1, schools are financed by the state, which means that schools that provide children with personal digital devices do not have more financial resources than those that do not. Children usually go to their local state schools, and all education offered by state schools is free. When starting school, children have normally attended kindergarten and the preparation and language stimulation children receive will be fairly equal across kindergartens. There will have been some variation related to both schools and students, but there is reason to believe that the only systematic difference between the two groups of schools (conditions) was the form of the writing instruction with teaching writing either by handwriting or by keyboarding on a digital tablet.

3.3.1 Conditions

3.3.1.1 The handwriting condition

Schools were selected for this condition if (a) the initial writing instruction was based on the use of pencil and paper and (b) the school did not provide children with personal digital devices. Writing on digital devices could occur (in computer labs or by sharing sets of tablets/computers), but it did not form a significant amount of the instruction.

3.3.1.2 The typing condition

The typing condition included schools that (a) based the initial writing instruction on the use of a digital tablet, postponing handwriting instruction and (b) provided all students with personal digital tablets to be used in writing (and some other) lessons.

When writing, the students produced text on an electronic touchscreen keyboard on personal digital tablets. A touchscreen keyboard differs from a physical keyboard in the tactile feedback they provide, and for experienced typists touchscreen keyboard might slow down typing speed and fluency (J. H. Kim et al., 2014). However, children learning to type do not use touch typing and need to see the keys they press, and the differences between the types of keyboards are therefore probably not that pronounced for beginning writers.

The teachers reported that students used the applications Skoleskrift (Ohlis, 2018), STL+ (Lingit, n.d.) or BookCreater (Book Creator, n.d.), integrated with Intowords (Vitec, n.d.). All of these applications offer text-to-speech technology. Teachers reported that students could use this functionality during instruction and assessment tasks and that, in general, their students used text-to-speech, although a few students sometimes chose not to use it. The text-to-speech functionality means that when writing a letter, the child heard the corresponding sound, when pressing the space button, the preceding word was read aloud and when typing a sentence terminator the preceding sentence (or group of words) was read aloud. Additionally, the whole text could be read on demand. The teachers also reported that other writing support, such as spell checking, was turned off during writing instruction and assessment tasks. Text-to-speech is assumed to provide students with much the same support as spellcheckers. It is also assumed to be easier for beginning writers to use because it does not require the same level of spelling and decoding skills as spellcheckers.

3.3.1.3 The mixed condition

Schools were included in the mixed condition if (a) initial writing instruction involved both handwriting instruction and the use of a digital tablet and (b) the school provided students with personal digital tablets to be used in writing (and some other) lessons.

When writing digitally, students in this condition produced text on an electronic touchscreen keyboard on personal digital tablets. Teachers reported that students used the applications Skoleskrift (Ohlis, 2018) or BookCreater (Book Creator, n.d.) integrated with Intowords (Vitec, n.d.). These applications offer text-to-speech, and the teachers reported that the students had the opportunity to use this functionality during writing instruction and assessment. The same settings for this functionality were used as in the typing condition (spoken feedback of both letters, words and sentences, and the whole text on demand). Several of the teachers also stated that they had not yet had time to train children in this functionality when they were assessed in November. Other writing support, such as a spellchecker, was disabled during writing instruction and assessment tasks.

3.3.2 Treatment fidelity

This work studies natural practices where the modalities play different roles in writing instruction, which was otherwise business-as-usual instruction. It had to be checked that the three conditions actually differed with regard to the modality used in writing instruction – that the typing condition involved instruction primarily by digital tablets, that the handwriting condition involved instruction more or less exclusively by hand, and that the instruction in the mixed condition was fairly balanced between handwriting and typing. First, inclusion in the *DigiHand* project and selection for each condition were based on the schools' policy related to the provision of personal digital devices. The schools reported their intended writing practice at the start of the year, and the availability of the equipment (personal digital devices or not) would further make it

possible to execute the practices. Second, fidelity treatment was checked by surveying all teachers about their writing instruction practice in both the first and second semesters of the first grade. Teachers were asked to report how they introduced new letters and how much time they typically would spend on different writing-related activities, for example writing letters and words and discussing what makes a good text.

Responses from the teacher questionnaire indicated some variation in the content of the writing instruction, as could be expected in a natural experiment. However, teachers in the mixed condition generally reported that children did writing activities both by hand and by digital tablet. In the handwriting condition, teachers reported that activities with pencil and paper dominated, while in the typing condition, teachers reported that writing activities involving the use of digital tablets dominated. Articles 2 and 3 provide more details.

3.4 Writing tasks and procedures for assessment

3.4.1 Developing writing tasks

As is the case in most previous research on written composition in first graders, the narrative genre was chosen for the writing tasks. The narrative genre is also the most widely used genre in writing in primary-grade Norwegian classrooms (Graham et al., 2021; Håland et al., 2019).

An effort was made to make the writing tasks relevant for the children, as not only technical writing skills but also motivation are crucial for writing (Bruning & Horn, 2000). The topic of the tasks reflected situations that first graders could recognise, persons with whom they could identify and animals that could motivate writing. They were designed as pictures showing a scene inviting a story to be told, like a monkey on his way to steel a banana from a passer-by, a boy and a girl playing football and having lost their football in a tree. When writing, each student got a print of the picture on their desk – the idea being that

a colourful, delicate picture could inspire writing. In addition to the picture, students were provided with three words that denoted central persons or objects in the picture. If there were students with weakly developed transcription skills, they could copy these three words. All tasks are presented in Appendix 2.

In line with recent writing research emphasising that writing in school should be authentic (cf. Håland et al., 2019), a framing story of an audience was made. This audience was Teddy bear Elling, who loves stories ('Elling som elsker fortelling'), and a dialogue was created between Elling and the students – he wrote letters back to the students with pictures of him reading their stories, asking for more texts, etc.

Successful writing assignments often include preparatory work to provide pupils with common experiences and topic help (Kvistad & Smemo, 2015). Therefore, a short introductory plan was made to ensure that all students were reminded of the narrative genre through a brief explanation of what a story was and some examples of stories.

A pilot study was conducted in a first-grade classroom to test two of the writing tasks and the teacher's instructions. I was present and observed the lesson, and after the lesson, the teacher was invited to give feedback on the teaching plan. Minor adjustments were made to the instructions after the pilot; for example, the size of the print of the instructions was increased. Alternatives for handing in the texts were also discussed with the teacher.

3.4.2 Written composition assessment (procedures)

The teachers administered the writing assessment. They were given detailed and tight instructions to follow and received all necessary material, both by email and on paper by mail, prior to each assessment occasion. This included introductory material, tasks and – in the handwriting and mixed condition – paper on which to write, copied to all students. The teachers were encouraged to motivate the students to do

their best and not answer questions about spelling or what to write, other than repeating the general instruction.

Each writing task, including the introductory part, was conducted within a 45-minute lesson. Students who finished early were instructed to read a book quietly. The reason for giving the students so much time was to make sure all could finish their compositions.

In the handwriting condition, assessment tasks were written by hand; in the typing condition, assessment tasks were written on a digital tablet; and in the mixed condition, students were assessed in both modalities. When the assessment was done by handwriting, the children wrote with pencil on paper, and I provided the teachers with sheets of paper on which students were to write. The sheets were lined and of a thicker quality than ordinary copy paper (160 gsm). When the assessment was done using a digital tablet, students wrote in the application and with the writing support they usually used during instruction (cf. Section 3.3.1).

Two different writing prompts (Task A and B in Appendix 2) were used for assessment in the mixed condition, and these were counterbalanced across classes. Similarly, modality was counterbalanced across classes in this condition. Five different tasks (Task A, B, C, D and E in Appendix 2) were used for assessments in the handwriting and typing conditions, and these tasks were counterbalanced across time and condition.

3.5 Text analysis – foundational issues and measures applied

Before explaining how the texts were analysed, I give an account of the philosophical foundation on which this thesis rests. I have positioned this part here, as the philosophical ideas were decisive in how I analysed the texts.

3.5.1 Philosophical stance

All research is underpinned by a philosophical worldview. This philosophical foundation comprises ontological and epistemological assumptions that guide the researcher when conducting an enquiry (Creswell & Creswell, 2018, p. 5). Within the philosophy of social sciences, there is generally a distinction between naturalistic approaches, which hold that the social sciences should embrace the aims and methods of the natural sciences, and interpretivist approaches, which argue that the social sciences must develop their own ideals and methodology to account for the meaningful phenomena of the social world (Gorton, 2010; Rosenberg, 2016, p. 11–33). This thesis belongs to the former tradition. Cognitive writing research, arising from cognitive psychology, aims to explain writing performance and learning through cognitive processes by developing theories and models that can be tested empirically (MacArthur & Graham, 2016). These ideals are parallel to those found in the natural sciences, which, through empirical investigation, seek to uncover natural laws that can allow for explanation and prediction.

This thesis also positions itself within what Philips and Burbules (2000) call a postpositivist orientation. As the term *postpositivist* indicates, this philosophical orientation challenges aspects of traditional positivism, which has been influential in the philosophy of science underlying the natural sciences. The main criticism concerns the positivist foundationalist view of knowledge, which upholds that knowledge is built on unchangeable foundations. Postpositivism, on the other hand, sees knowledge as conjectural (Philips & Burbules, 2000, p. 26). Knowledge has to be soundly based, but it can still be fallible. As a subject matter is further investigated, the warrants on which the knowledge rests can be dismissed and knowledge can be updated (Philips & Burbules, 2000, p. 26). This nonfoundationalist view of science goes back to Popper (2009), who argued that science advances by the

formulation of conjectures that are subsequently subjected to tests that can falsify them.

According to a postpositivist worldview, objectivity is of crucial importance in scientific enquiry. This importance can be seen in relation to the philosophy's axiology, which concerns the role of values in research. Within the postpositivist approach, values should not influence the concepts or procedures that are used in the practice of research. To be more precise, the execution of research should not be influenced by external or epistemically irrelevant values, such as political or personal values (Philips & Burbules, 2000, ch. 3). However, internal and epistemically relevant values, such as the accuracy of measurements and pursuing truth, are necessary in the scientific enterprise (Philips & Burbules, 2000, ch. 3).

3.5.2 Measurement of meaningful texts

In line with the naturalistic, postpositivist worldview, the three empirical articles all apply a quantitative design. Even so, in the present thesis, the main source of data was students' texts. Text analysis normally entails an analysis not only of formal aspects of the texts but also of the content, and textual content has clearly qualitative characteristics. The content of a text consists of language – meaningful structures – and how one can best approach such content by quantification is worth consideration. In Article 1, I explore how a quantitative study of text quality based on a naturalistic, postpositivist worldview can handle meaningful structures in text. The point of departure is the division between naturalistic and interpretivist approaches to research and how philosophical assumptions have consequences for the way texts are analysed. I use the study reported in Article 2 as an example of how texts can be analysed in line with ideals in the naturalistic tradition.

I also argue in Article 1 that, on a fundamental level, text analysis involves understanding and interpretation. For example, features in texts cannot be coded – assigned different values within the categories of

analysis – before the analyst understands the content. Moreover, language structures carry meaning, and these meanings can be ambiguous; that is, they depend on context (Cruse, 2011). In texts written by young writers, there will typically be more ambiguity as the children are learning to commit text to paper/screen and to follow writing conventions. An analysis of such texts will imply judgments, for example, on the most fundamental level, whether the marks on paper are actually characters or not. Further, I discuss how rigorous work with operationalisations, precise coding rules and double coding can contribute to the systematic, objective and transparent analysis of texts.

The systematic and objective investigation of texts implemented in this thesis presupposed that a lot of thought had to be devoted to the analysis process. In the following sections, I state the rationale for the text analysis adopted in the thesis and present the text quality measures, both why they were selected and how they were operationalised.

3.5.3 Text quality assessment – rationale

I chose to assess text quality in first graders' narratives analytically. Many previous studies of first graders' written composition ability apply holistic scoring methods (Graham et al., 1997; Juel et al., 1986; Kim & Park, 2019; Torrance et al., 2021). A holistically based assessment is an overall rating of the quality of a text, for example, on a six-point scale, typically based on general level descriptions (Huot, 1990). In Article 2, texts were also rated holistically, but the main approach to assessing children's written composition performance and learning was text-analytic assessment. This approach was chosen because it provides more information about students' composition abilities and is more transparent than holistic methods. A holistic quality rating, reflecting the overall quality of the text, is based on balancing a set of criteria, and how the rater balances these criteria is implicit. When different text features are assessed separately and formally, on the other hand, it is clearer what is being measured.

An analytic assessment can be particularly useful when studying texts by beginning writers, which typically are short, incomplete and marked by orthographic errors. Holistic ratings of such texts will typically indicate low quality and not necessarily differentiate between texts, while analytic assessment can extract information about, for example, the number of words written, spelling accuracy, vocabulary and syntactic complexity, even though the texts are simple and short. As the ability to, for example, produce more words, more accurate spelling and more complex syntax increases, the text-analytic assessment can give information about how these skills are developing. In addition, as this thesis is a comparison of composing in different modalities, the text analytic approach is useful in investigating the potential modality effects on different aspects of the texts. For example, as students could use text-to-speech functionality, it is of interest to study whether, and in what way, this functionality influences the final product – whether the spelling is improved and also whether other higher-level features of text are affected.

3.5.4 Text quality measures – selection and operationalisation

When selecting the text quality measures, I aimed to find measures that captured text features that were expected to be present in children's early narrative compositions. I wanted to include measures on different levels, from the micro level, such as spacing, to the meso level, such as syntax, and the macro level, such as text organisation. The nature of children's texts – often short, simple, incomplete and with inaccuracies – was a constraining factor, and only features that could be operationalised in an explicit way were chosen.

Below, I present the text quality measures and challenges and I operationalise them by illustrating examples from the texts. First, I briefly describe how the texts were transcribed and how this first transcription was transformed into several new transcriptions from

which the text measures could be extracted. A more detailed account of all rules for transcription and coding can be found in the transcription and coding manual in Appendix 3.

3.5.4.1 Transcription and coding

The handwritten texts were transcribed, while the digital texts were copied and adjusted. All characters were standardised to lower-case letters. Drawings or inserted pictures of emoticons were not included in the transcription. The transcription maintained the spelling, spacing and punctuation done by the students. Inverted characters were accepted as long as they could not be mixed with other letters. Table 1 shows a few examples of (excerpts from) texts, and in the three first texts, there are instances of inverted letters. In text 2, there are examples of inverted <a>, which were all accepted as <a>; in the last word in text 3, there is an inverted <t> which was accepted as <t>, and in texts 1 and 2, there are examples of an inverted <b>, like *dallen* 'the ball', which were transcribed as <d>. To transcribe the space between words, there had to be a larger space between the words than between the characters in the words. For an illustration of two words transcribed without a space, see *nåer* [nå er] 'nowis' in text 3. To insert space within a word, there must be enough space to insert a character, see *is en* 'the ice' and *spi s* 'eat' in text 4. When facing unclear text, I discussed my transcription with the two Norwegian-speaking supervisors.

The subsequent coding of the texts, where the texts were scored for the text measures, was done by transforming the first transcription into several versions. The next transcriptions included stepwise corrections of the texts, including spacing errors, spelling errors, punctuation and markup of syntax and narrative structures. See Table 2 where text 1 from Table 1 is coded. In this process, the texts were blinded for modality, student, school and (for Article 3 and 4) point in time.

Table 1. Transcription of Texts

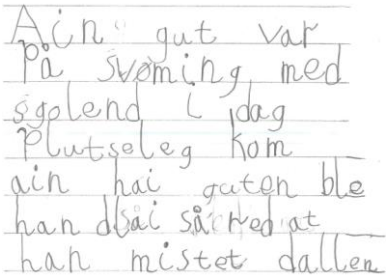
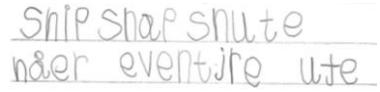
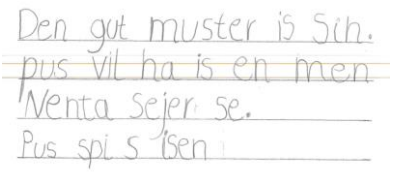
Original Text	Transcription	Spacing and spelling corrected, and translation
Text 1 	ain gut var på svøming med sgolend i dag plutseleg kom ain hai guten ble han dlai så red at han mistet dallen	ein gut var på svømming med skolen i dag plutseleg kom ein hai guten blei han blei så redd at han mista ballen 'a boy went swimming with school today suddenly there came a shark the boy was he was so afraid that he lost the ball'
Text 2 	be var en gang en gut som mista is klene og katen slika ben opp	det var en gang en gut som mista iskulene og katten slikka den opp 'once upon a time a boy who lost the ice balls and the cat licked it up'
Text 3 	snip snap snute nåer eventjre ute	snipp snapp snute nå er eventyret ute 'snip snap snout the story is out'
Text 4 	den gut muster is sin. pus vil ha is en men nenta sejer se. pus spi s isen	den gut mister is sin. pus vil ha isen men jenta seier se. pus spis isen 'that boy loses ice his. kitty wants to have the ice but the girl says look. kitty eat the ice'

Table 2. Coding of Texts

Trans-cription	Action	Measures Extracted	Text 1 Transcribed and Coded
T1	Character by character transcription		ain gut var på svøming med sgolend i dag plutseleg kom ain hai guten ble han dlai så red at han mistet dallen
T1.5	T1 parsed into words	Text length (words) Spacing accuracy	ain gut var på svøming med sgolend i dag plutseleg kom ain hai guten ble han dlai så red at han mistet dallen
T2	T1.5 copied and corrected for spelling	Spelling accuracy	ein gut var på svømming med skolen i dag plutseleg kom ein hai guten blei han blei så redd at han mista ballen
T2.5	T2 copied and homonyms marked up	Vocabulary sophistication	ein gut var på svømming med skolen i dag plutseleg kom ein hai guten blei han blei så redd at han mista ballen
T3	T2 copied and wrongly inserted terminators removed		ein gut var på svømming med skolen i dag plutseleg kom ein hai guten blei han blei så redd at han mista ballen
T3.5	T3 copied and missing terminators inserted	Terminator accuracy	ein gut var på svømming med skolen i dag. plutseleg kom ein hai. guten blei han blei så redd at han mista ballen.
T4	T3 copied and syntax marked up	Syntactic complexity and accuracy	[M ein gut var på svømming med skolen i dag] [M plutseleg kom ein hai] [MW guten blei han blei så redd] {S at han mista ballen}
T5	T4 copied and syntax markup altered to narrative markup	Basic narrative structures Advanced narrative structures Story grammar	O [D ein gut var på svømming med skolen i dag] C [EP plutseleg kom ein hai] R [EA guten blei han blei så redd] {EFP at han mista ballen}

Note: Translation can be found in Table 1. Scores: text length 23, space use accuracy 1.0, terminator accuracy 0.0, spelling accuracy 0.61, vocabulary mean age 6.9, syntax 4.5, event count 3, advanced narrative features 4, story grammar 2. See Section 3.5.4.5 and 3.5.4.6 for an explanation of the syntactic and narrative codes.

3.5.4.2 Text length

Text length was included as this measure has been found to correlate with text quality in primary-grade children (e.g. Berninger et al., 1992; Dockrell et al., 2015; Malvern et al., 2004). Composing a narrative presupposes the ability to produce a certain number of words, and the relation between quantity and quality is probably closer for beginning writers than for more experienced writers. Text length was measured as the number of words written. Errors of spelling and segmentation were ignored; for instance, *sijøpereinis* [kjøper ein is] 'bujysanice'[2] was counted as three words. To decide whether a letter string was a word, I looked for phonologically plausible spellings. The spoken dialects of the students were taken into account, and the co-text (surrounding words) was used as cues. Letter strings that could not be identified as words were coded as non-words (NW) and not counted; for instance, the following text was counted as six words: *gut muster is han blai lai. [NWqwetyuiopå]* 'boy drups ice he torned sad. [NWqwetyuiopå]'. Numbers that made sense in context were counted as words, for example, *3 små puser* '3 small kitties', which was counted as three words. Words that were repeated were counted as two or more words. For example, in *pusa sa mjau å gutten sa å kor søtt er du kjempe søtt søtt søtt søtts søtt* 'the kitty said meow and the boy said oh how sweat are you very sweat sweat sweats sweat', all repetitions of *søtt* 'sweat' were counted as separate words.

3.5.4.3 Transcription-related measures

Of the micro-level features, spelling has been studied extensively, but spacing and punctuation have also been assessed in other studies of first-grade children's written composition (Kim et al., 2013; Salas & Caravolas, 2019).

[2] I have tried to reproduce the linguistic errors in the translations in an attempt to include the readers in the complexity of analysing these texts. It has, though, not been possible to recreate the errors in an exact manner.

Even though the sampled schools used Nynorsk as the written standard, many of the texts contained word forms from Bokmål. The students would probably have been exposed to both Nynorsk and Bokmål before starting school, for example, from children's books. Additionally, both Nynorsk and Bokmål have many juxtaposed forms, officially correct forms of both lexemes and inflectional forms, some of which overlap in the two standards and others not. For example, *sjuk* and *syk* 'ill' are juxtaposed in Bokmål, while only *sjuk* is accepted in Nynorsk. In Nynorsk, *kome* and *komme* 'to come' are juxtaposed, while only *komme* is accepted in Bokmål. In Nynorsk, the official infinitive form of this verb can be either *kome*, *koma*, *komme* or *komma*. Part of the reason for the wide freedom of choice in norms is that the official language policy has aimed at having written standards that are close to the way people speak, both for democratic and pedagogic reasons, for example, to make it easy for children to learn to write. Children might have used their spoken dialect to sound out the words they were writing, and the spoken dialects in Western Norway are, in general, closer to Nynorsk than Bokmål. However, many dialectal forms are compatible with Bokmål, and many dialectal forms are not accepted in either of the written standards. In addition, the dialect of the capital and the surrounding area, a high-prestigious dialect, is very similar to Bokmål, and most children will have been exposed to this dialect, for example, through television. As the spellings in the texts were very unsystematic with regard to being from Nynorsk or Bokmål, I decided to accept both standards, although not in the same text. Each text was therefore spelling-corrected to either Nynorsk or Bokmål (following the official dictionaries *Nynorskordboka* and *Bokmålsordboka*), according to what would give the least number of errors. I chose Nynorsk when correction to either standard would give the same number of errors.

As the spelling measure captured the children's compositional spelling, the correction had to handle real word errors. For example, in a story about a boy who bought an ice cream, a student writes *han skole spise den* 'he was guing to eat it'. Isolated, *skole* means 'school', and the

student probably aimed at writing *skulle* 'be going to, shall'. Therefore, this word was treated as a misspelling. The coding rule was that when words did not make sense in context and were not based on a lexeme that would make sense in the context and there was a plausible spelling of a word that would make sense in the context, the word was corrected to the word that made sense in the context. The second part of the rule ('not based on a lexeme that would make sense in context') entails that inflected forms were not corrected to the inflected form implied by the context. Rather, this was picked up in the syntax measure. For example, in *han ete isen* 'he eat the ice', the context implies a present form of the verb, but the student has used the infinitive. This was not counted as a misspelling, but the clause would be marked as syntactically incorrect. There were, however, a few cases in which the spelling correction was also a correction of morphological errors. For example, when weak verbs were conjugated as strong verbs (or the opposite), these were corrected to the official forms; for example, *stjal* 'stole' was written as *stjelte* 'stealed', and as *stjelte* does not exist in the written standard, it was corrected to the official form. I treated this as a spelling error, although it might also be seen as a morphological error.

The spelling measures did not include errors in capitalisation. The use of lower-case and upper-case letters was generally rather unsystematic, with some texts written only in either of the two and some texts with a mixture. I chose to standardise all texts to lower-case letters, but if I had had more time, I would have liked to analyse the use of lower- and upper-case letters.

The texts were scored for spelling accuracy, which was the total number of correctly spelled words divided by the number of words written. This means that each word was scored dichotomously as either correctly spelled or not, irrespective of the number of characters misspelt or type of error (code error, misspelling of a regular word, etc.).

Spelling also had to be seen in relation to the space-use measure: Are, for instance, compounds that are split up, even though they officially are written as one word, spelling errors or segmentation errors?

I chose to treat all segmentation errors as spacing errors, and not spelling errors, both when simplexes were split, like *bu r* 'ca ge', when compounds were split, like *fot ball* 'football', and when space was missing between words, for example, *isenfalt* [isen falt] 'theicefell'. Punctuation was accepted as segmentation. For example, *ute.er.de.varmt* 'it.is.warm.outside' was scored as correctly segmented, while in the punctuation measure, these terminators were marked as wrongly inserted.

The sentence terminator measure reflected the ability to correctly use sentence terminators: to end a sentence with a terminator and not insert terminators within a clause (or a word). A challenge in this mark-up was never-ending sentences where several clauses were coordinated, for example *ei jente og en gutt spilte fotball men så spente jenten ballen oppi treet og så å så begynte greinen og knekke og så kom mamma med en stige og tok mamma stigen oppi treet og så klatret jenta ned med ballen* 'a girl and a boy played football but then the girl kicked the ball in the tree and then and then the branch started to break and then mama came with a ladder and took mama the ladder into the tree and then the girl climbed down with the ball'. I decided that only two main clauses (including any subordinate clause) could be coordinated within the same sentence (period), which again was terminated by a full stop. Thus, in this example text, three terminators were inserted.

3.5.4.4 Vocabulary sophistication

The vocabulary measure was specifically developed in this study. The aim was to measure the children's vocabulary sophistication as objectively as possible. Other analytic assessments of vocabulary imply, for instance, that the rater assesses whether the vocabulary is rich, expressive, mature or vivid (e.g. Wechsler, 2006). For Norwegian, no frequency lists with a focus on children's vocabulary exist. There is a lexical database of Norwegian content words containing, among other things, subjective age-of-acquisition ratings (Lind et al., 2015), but this corpus does not include all the words used in the text material analysed

in the thesis. The idea arose that a measure of the written age of acquisition could be developed. Thus, all lexical lemmas (content words) from the texts were extracted and rated by teachers and trainee students in a digital survey. The respondents were asked to judge when each of the words (lexemes/lemmas) typically would appear in children's texts. This reflects an assessment of how the words are perceived in terms of maturity. Each of these ratings is, of course, subjective, but in total, they represent an assessment of vocabulary sophistication that is more objective than my own rating would have been.

All lexical lemmas (nouns, verbs, adjectives) were extracted. Proper names and very specific compounds that would not make much sense out of context and therefore hard to rate were taken out. One example was the compound *eplekule* 'apple ball' from a text where the protagonist asked for different sorts of ice balls, including one that tasted of apple. I also had to manually go through the list of lemmas and identify homonyms, for example *grein*, which can mean both 'branch' and 'cried', and mark these up. In total, 845 lemmas were rated. The following are examples of the mean written age-of-acquisition scores: *is* 'ice' 5.33, *pus* 'cat' 5.80, *bil* 'car' 5.93, *redd* 'afraid' 6.37, *bukse* 'pants' 6.43, *sulten* 'hungry' 7.25, *parykk* 'hairpiece' 9.31, *gigantisk* 'gigantic' 10.62 and *tverrliggar* 'cross-bar 11.75'.

In hindsight, this measure, with words being rated in intervals of years, might be more useful when investigating the development of composition skills across several years.

3.5.4.5 Syntactic complexity and accuracy

The ability to create complex syntax is an advantage when composing narratives, as this makes it easier to structure events in hierarchical constructions (Berman & Slobin, 2013). The syntax of primary-grade children's compositions is typically assessed through measures of syntactic complexity (e.g. Kim et al., 2014; Wagner et al., 2011). In the present thesis, the syntax measure reflected complexity by differentiating between the use of main or (also of) subordinate clauses. For example,

the following story was credited with four main clauses (M) and one subordinate (S): *[M ole var i byen] [M og kjøpte is] [M men {S når han skulle ete isen då} datt isen] [M og pusur tok isen]*. '[M ole was in town] [M and bought ice] [M but {S when he was about to eat the ice then} the ice fell] [M and pusur (Garfield) took the ice]'. The main clause with a subordinate clause embedded could also have been expressed as two main clauses ([M he was about to eat the ice] [M but the ice fell]), but embedding a subordinate clause can be seen as more advanced. Typically, the amount of subordination increases with age both in the oral language and in the writing of primary-school children (Loban, 1976, pp. 35–39).

In addition, accuracy was included in the syntax measure, operationalised as whether the clauses contained syntactic errors (coded as [MW] if containing one or more errors). There could be many sorts of syntactic errors, of which the most important were the use of non-finite verb, such as *gutten merke ikke haien* 'the boy not notice the shark'; a lack of words/part of sentence, such as *guten sparka ballen treet* 'the boy kicked the ball tree'; the use of additional words, such as *ho skulle skal spele fotball* 'she should shall play football'; the wrong word order, such as *plutselig mannen sklei på bananen* 'suddenly slipped the man on the banana'; and the lack of concord/definitiveness *en dame går på vei* 'a lady walks on road'.

3.5.4.6 Narrative measures

Narrative measures have traditionally centred around structure, which is also reflected in the measures used in this thesis. Inspiration was found in Labov and Waltesky's (1967) canonical work on oral personal narratives, which identified obligatory and optional structural components of the macrostructure in oral narratives. However, as the texts studied were written by beginning writers, the measure of macrostructure had to be fairly basic, and I landed on a macrostructure of three stages: orientation, complication and resolution. The orientation gives information about participants, places, activities and/or time and

gives background knowledge to understand the complications of the story. The complication is a disruption in an expected course of events (the situation depicted in the orientation or if the orientation is missing, compared to a normal situation). The resolution answers the question of what finally happened as a consequence of the complication. To code a stage, there had to be at least two stages: either orientation and complication or complication and resolution. In Table 3, a few texts are coded for narrative measures. Texts 1 and 2 were analysed as having no global story structure, while texts 3 and 4 contained two and three stages, respectively. This story grammar measure primarily concerns structure, but it is also related to text coherence and to the use of main characters, which are also important aspects of narratives. For example, one of the coding rules for approving the orientation and complication was that at least one element (person, animal, object) from the introduction must follow in the complication.

As the story grammar measure is rather coarse-grained, it was appropriate to include a more fine-grained measure of narrative structures on the local level. Here, inspiration was found in Martin and Rose (Martin & Rose, 2008), who analysed narratives through phases, i.e. discourse units, that can occur at any narrative stage (structural component on the macro level explained above), such as event, problem and solution. On the most basic level, a narrative can consist of two events that are linked in time (Labov & Waletsky, 1967), and therefore, the number of events functioned as a measure of basic narrative structures. For example, in Table 3, text 1 only contained descriptions, while text 3 illustrated a minimal story of two events. The more advanced narrative structures were *problem*, an undesired event or description; *solution*, which corresponds to a problem and restores order; *reaction*, an attitude, feeling or thought of a participant that was caused by a preceding event; *effect*, a consequence of a preceding event; and *comment from narrator*, when the narrator intruded to comment on events, give explanations, etc. See texts 4, 5 and 6 in Table 3 for examples.

Table 3. Examples of narrative coding

Example text	Translation	Coding key
Text 1: [D eg ser gut] [D eg ser is] [D eg ser pus]	[D I see boy] [D I see ice] [D I see kitty]	Story grammar O = Orientation C = Complication R = Resolution Basic and advanced narrative structures E = event P = problem S = solution F = effect A = reaction N = comment from narrator Other codes (not counted in the measures) D = description I = narrative formula
Text 2: [EP gut mistar is] [EF pusen ser opp] [EP å isen dett på bakken.] [E storesøster ventar i står i køen] [E lillebror ventar står i køen i iskiosken.]	[EP boy drops ice] [EF the kitty looks up] [EP and the ice falls on the ground.] [E big sister waits in stands in the line] [E little brother waits stands in the line in the ice kiosk.]	
Text 3: C [EP gutten mista isen i gulvet] R [EF pusen åt opp isen]	C [EP the boy lost the ice on the floor] R [EF the cat eat the ice up]	
Text 4: O [DI det var ein gong ein gut] [D som ville ha ein is]. [EF han gjekk til butikken]. [E han fekk fem iskuler]. C [EP men katten åt isen]. R [EA da blei guten sur]. [EFS guten fekk seg ny is]. [DS katten og guten vener].	O [DI once upon a time there was a boy] [D who wanted to have an ice]. [EF he went to the shop]. [E he got five ice balls]. C [EP but the cat ate the ice]. R [EA then the boy turned sour]. [EFS the boy got a new ice]. [DS the cat and the boy friends].	
Text 5: O [DI ein gong var det ein mann] [D som hadde ei ape] [D som elskar banan] [E og så kom det ein mann] [D og han hadde ein banan] [EF og da ropte apa banan!] C [EP apa tok bananen] R [EA og så blei mannen sint].	O [DI a time there was a man] [D who had a monkey] [D who loves banana] [E and then came a man] [D and he had a banana] [EF and then the monkey cried banana!] C [EP the monkey took the banana] R [EA and then the man got angry].	
Text 6: O [EP jenta klarer å skyte ballen opp i eit tre] [EF ho klatra opp i treet]. [D ho stod på ein grein]. C [N tro de] [N at ho datt ned] [EP ho datt ned]. R [EF ho brekte beinet].	O [EP the girl hits the ball into a tree] [EF she climbs up in the tree]. [D she stood on a branch]. C [N you think] [N that she fell?] [EP she fell down]. R [EF she broke a leg].	
Text 7: C [EP en gutt mister isen sin oppå en katt] R [EF å katten prompa]	C [EP a boy dropped his ice cream on a cat] R [EF and the cat farted]	

The coding of the three narrative features was checked through double coding of approximately 20% of the texts by the two Norwegian-speaking supervisors. Overall, there was very good inter-rater reliability, although not perfect. Text 7 in Table 3 is an example of a text that was coded differently by two raters. One of the raters interpreted the two clauses as two related events: The cat farted because the boy lost his ice cream on him – it could be that he was startled and therefore farted, or this reaction was a revenge against the boy who made him dirty/wet, or the reaction was a result of the cat eating the ice cream that was dropped on him. This rater coded two events, the first one including a problem and the second one including an effect. Further, the rater coded the story grammar with a complication and resolution on a global level. On the contrary, the other rater perceived these two clauses as two separate events without any causal relation. This rater coded two events, from which the first was also coded as a problem, and no story grammar. Both raters followed the coding rules; however, applying the rules when coding specific texts depends to some extent on background knowledge, such as possible and reasonable causal relationships. As discussed in Article 1, narrative coding involves understanding and interpreting complex linguistic structures in relation to coding rules. While, for example, the spelling of a text can be compared to a clear norm (e.g. a dictionary), the narrative coding must be done in line with a set of criteria, which in itself must be understood through language before it can be applied to language structures that also must be understood and interpreted.

3.5.4.7 Holistic quality

A rubric for holistic quality was developed for this purpose, as no other suitable rubric was found. The rubric was informed by the rating scale developed by Arrimada et al. (2018), which was used for assessing quality in first-grade compositions. The main challenge when developing the rubric was to make level descriptions that could differentiate between texts that were generally simple and short. The criteria centred around

progression of ideas and elaboration of details, story organisation, cohesion and vocabulary. The rubric can be found in Appendix 3 in Article 2. When the texts were rated, they were corrected for spelling, spacing and punctuation errors. Table 4 shows six texts rated, one for each level.

Table 4. Example texts rated for holistic quality

Score	Example Text	Translation
0	ball jente gut mål blad.	ball girl boy goal leaf.
1	jente og ein keeper jente sparka den oppi treet. ferdig.	girl and a keeper girl kicked it into the tree. done.
2	jente sparka ball oppi tre. jente klatra i treet. [NW grtre].	girl kicked ball into tree. girl climbed the tree. [NW grtre].
3	ho skulle skal spele fotball. så kom ballen datt i treet så ho måtte hente ballen i treet.	she was to shall play football. then the ball came fell in the tree so she had to fetch the ball in the tree.
4	dei sparka ball og så for ballen oppi eit tre. og så gjekk jenta for opp oppi treet og henta ballen oppi treet. snipp snapp snute so var eventyret ute.	they played ball and then the ball went into a tree. and then the girl went up into the tree and fetched the ball in the tree. snip snap snout the story is out.
5	det var ein fin dag. guten å jenta spela fotball men plutseleg hamna ballen i tre. jenta måtte klatre oppi tre. men plutseleg brakk greina. jenta heldt fast i greina. guten måtte klatre opp i tre å redde ho. å så snipp snapp snute så var eventyret ute.	it was a nice day. the boy and the girl played football but suddenly the ball fell in tree. the girl had to climb up in the tree. but suddenly the branch broke. the girl held on to the branch. the boy had to climb up in the tree and save her. and snip snap snout the story is out.

3.5.4.8 Final remark on the text quality measures

Language has embedded ambiguity, with words, expressions and grammatical structures potentially having multiple meanings (Cruse, 2011), but ambiguity is reduced when the linguistic structures can be interpreted in a context. In compositions by beginning writers, there is room for ambiguity because, one, there might be a limited amount of text available and, two, the texts typically have errors in spelling and syntactic structures and contain ellipses or incoherent parts. As the preceding discussion of the composition measures illustrates, I took into account the nature of the children's texts when selecting and operationalising the measures, for instance by including accuracy and not only complexity in the syntax measure and by applying a basic structure of story grammar. In all cases, detailed rules were made to systematically code the texts.

I considered including other measures, such as originality (briefly discussed in Article 1), and different measures of cohesion, which have been used in other studies measuring narrative quality in primary-grade students (e.g. Andersen et al., 2018; Jones & Pellegrini, 1996). With regard to cohesion, this was partly covered by the narrative measures. As the compositions were short, cohesion and coherence – captured by story grammar and advanced narrative features – would typically be concurrent. For example, if a text contained an advanced narrative feature, for example a *reaction*, where a preceding event caused a reaction in a participant, there would probably be cohesion between the clauses expressing this content. There might also have been other ways of operationalising the constructs that I chose to measure, although I think my operationalisations are reasonable. In sum, I believe that my measures capture central features that can be expected to be present in narratives by beginning writers.

3.6 Literacy related measures

At school entry, students were tested for a series of literacy-related skills: grapheme-to-phoneme mapping, phoneme isolation, phoneme blending, word reading, spelling and vocabulary. It was important to include measures that could capture variations in skills. As the participants had not received any formal literacy instruction, it could not be expected that most of them could read and spell single words, and measures of precursors of reading and spelling ability were therefore included. The literacy tests have been used in prior research on early reading and writing (Solheim et al., 2018; Sunde et al., 2020). The measures of phoneme isolation, phoneme blending and vocabulary have been shown to predict word reading and spelling ability in Norwegian first graders (Solheim et al., 2018). More details about the measures and the procedures related to testing are presented in Articles 2 and 4 and in Gamlem et al. (2020).

3.7 Statistical methods

3.7.1 Bayesian statistics

For the statistical analysis, the Bayesian framework (Kruschke, 2015; Lambert, 2018) was used. The use of Bayesian methods is growing in many fields, such as educational research and psychology, although the frequentist framework still dominates (Andrews & Baguley, 2013; König & van de Schoot, 2018). The Bayesian paradigm rests on probability theory in the making of models to understand the phenomena surrounding us. A Bayesian analysis centres around Baye's rule, which involves going from the effect (data) back to its cause (the mechanism that generated the data). Baye's rule is a formula that allows for combining prior beliefs in model parameters of interest with the data at hand to get a posterior – better and more informed – belief regarding the same model parameters. This posterior belief is expressed as a probability distribution that can be used to answer research questions

through probability statements, for example, the probability of a specific hypothesis given data.

3.7.2 Reasons for choosing the Bayesian approach

There are several advantages to the Bayesian framework for statistical analysis (Kruschke, 2015; Lambert, 2018; Nicenboim & Vasishth, 2016). For this thesis, the Bayesian approach was chosen in place of the frequentist approach for several reasons. The most important reason was that Bayesian statistics allows for the demonstration of evidence in favour of both the alternative hypothesis (H1) and the null hypothesis (H0). Frequentist methods measure whether the data are incompatible with H0. If the data are not compatible with H0, H0 is rejected. However, these classical methods can never support H0 directly.

In the case of modality effects on first graders' written composition, there is a possibility that the true effect is zero, and it is therefore of interest to obtain robust evidence in favour of H0. The Bayes factor (BF) is a way of quantifying statistical evidence in favour of the two hypotheses. More precisely, BFs can be seen as continuous degrees of evidence that can indicate evidence for H1 (and against H0), evidence for H0 (and against H1) or no evidence for either H1 or H0, which is inconclusive evidence (Dienes, 2016). A limitation of BFs is that they are sensitive to priors. For example, if a maximally uninformative prior is used, the Bayes factor is likely to favour H0 (Wagenmakers et al., 2010). The sensitivity of BFs to priors should always be kept in mind, but when weakly informative priors are used, as in this thesis, this sensitivity is less of a problem than when extreme priors are applied.

Another important reason for using Bayesian methods is that they offer flexible modelling. In Article 2, multivariate mixed-effects models were used to model writing performance. In Articles 3 and 4 these were extended to model growth, with non-linear effects. The initial plan for the PhD project was to use frequentist methods, and the analysis for the first article was in an early phase done through these, using the lme4

package (Bates et al., 2015), but convergence problems arose. Such problems often arise when fitting mixed-effects models with a maximal random-effects structure to sparse or moderate amounts of data (Nicenboim & Vasishth, 2016; Winter, 2019, p. 266). By contrast, the relatively complex mixed-effects models in this thesis could be fit without estimation problems using Bayesian methods.

In addition, Bayesian measures often have an intuitive interpretation. For example, Bayesian credible intervals can be interpreted as a probability (Lambert, 2018, pp. 129–131). A 95% credible interval gives the range of the parameter value of interest, and this can be interpreted as a 95% probability that the parameter value lies within the interval. By contrast, the frequentist confidence interval implies that, across an infinity of confidence intervals from hypothetical samples, the parameter value will be in the interval 95% of the time.

Below I will outline the Bayesian analysis and comment on choices done in the analyses of the empirical articles. The data and scripts for the statistical analyses can be found on OSF on the following sites:

Article 2: https://osf.io/q8z3u/

Article 3: https://osf.io/j7ne3/?view_only=bcb821eb97164cfb895e362c1691257d

Article 4: https://osf.io/5t934/?view_only=30ada9b1650c4ac9b706b6d4cf93fc22

3.7.3 The Bayesian data analysis

Bayesian analysis comprises several steps. First, and most importantly, a probability model (likelihood) for the data must be selected. There are numerous likelihood distributions available, and one must choose a distribution that imitates the data-generating processes under investigation to the highest possible extent (Lambert, 2018, p. 146). In the studies comprising this thesis, different likelihood distributions were selected for the outcome measures. For example, vocabulary sophistication (continuous data) was modelled as normally distributed,

whereas story grammar (ordinal data) was modelled as a sequential process because the highest achievable category – having a complete structure of orientation, complication and resolution – would only be possible if the lower categories were also present – for example, containing an orientation and a complication (Bürkner & Vuorre, 2019).

The next step in Bayesian analysis involves specifying a prior distribution for the model parameters. This prior distribution is the researcher's prior belief in the different parameters – which are seen as probable and improbable. Uninformative priors set all values of the parameters as equally likely. Weakly informative priors ascribe more weight to parameter values one considers more probable, and informative priors can maximise the impact of prior beliefs, for example, if the results from former analyses should be considered. The prior and the likelihood are weighted to produce the posterior. The choice of the prior is more influential when there is little data. With much data, the posterior distribution is less sensitive to the prior, and the likelihood will dominate in determining the posterior. In the three studies of this thesis, weakly informative priors were used, which means that extreme priors either way (i.e. maximally uninformative and strongly informative priors) were avoided. For example, in Articles 3 and 4, the prior for the effect of modality on the text measures was set to a student's t distribution with scale (SD) 2, the centre of the distribution at 0, and 2 degrees of freedom.[3] This means that the effect was set to be 0 by default, but it allowed for a substantial amount of variance in the width of the distribution and allowed for some extreme values by specifying relatively fat tails of the distribution. Through such weakly informative priors, the extent to which extremely large parameter values received some prior probability was limited, while any plausible positive or negative value (including 0) was favoured.

When the likelihood and priors are specified, the posterior distribution can be calculated. Usually, this is done by computers through

[3] All priors can be found in the analysis scripts on the OSF sites for each article.

algorithms that approximate the posterior distribution and produce posterior samples. In the present thesis, the Stan probabilistic programming language (Carpenter et al., 2017) accessed through the R brms package (Bürkner, 2018), was used, which implements Markov chain Monte Carlo algorithms. The algorithms produce chains of samples from the posterior distribution, and the result is a large matrix with all samples. In the present thesis, models were run with 10,000 iterations on three chains with 5,000 iterations of warm-up, yielding 15,000 total samples for each model.[4] From the samples, relevant information can be extracted. Because the posterior samples approximate the whole posterior distribution, we get probabilities of all possible parameter values, not just a point estimate (and standard error), as the frequentist analysis yields. It is therefore possible to estimate several probabilities of interest; for example, the probability of the parameter values being within an interval or the probability that it is larger than 0. Additional techniques can be applied to evaluate the fit of models and select between models, for example, leave-one-out cross-validation (LOO-CV) and Bayes factors, as used in this thesis.

3.7.4 Bayesian methods and philosophy of science

The use of Bayesian methods deserves a brief discussion in light of the philosophy of science. As stated in Section 3.5.1, objectivity in analysis is a core ideal within naturalistic and postpositivist approaches to research. Bayesian methods are often criticised for being subjective because the researcher has to choose probability models and priors. However, any statistical analysis requires assumptions and judgments to

[4] This is more than the standard 8,000 samples that is default in brms. This number is often considered enough to minimise the error of the approximation of the posterior distribution. More samples yield less error, but take more time. For example, for the analysis in Article 4, the intercept-only model took roughly 9 hours to run, while the more complex Model 7: (time + modality) x (literacy skills + age) + time : modality, took 117 hours, almost 5 days, to run on an ordinary computer (with the specifications 8 GB RAM and CPU Intel Core i5, 1.60 GHz, 1800Mhz, 4 cores, 8 logical processors).

be made. A frequentist linear regression analysis relies on assumptions about the predictor and outcome variables, and this technique is often applied without specific justification (Lambert, 2018, p. 26). The Bayesian approach forces the researcher to be aware of, for example, model assumptions. Also, as argued by Rouder et al. (2016), variation in research in general is not considered problematic; for example, researchers will operationalise concepts differently and choose different designs and samples, and variation due to model specification will usually be relatively modest for reasonable choices. Further, the model specifications made by the researcher in a Bayesian analysis will be informed by expertise, such as theory, previous findings and common sense, and specific choices can therefore be defended (Rouder et al., 2016). In addition, when the model specifications are transparent, the choices can be evaluated by others.

Further, as Gelman and Shalizi (2013) observed, Bayesian inference is sometimes viewed as fundamentally different from frequentist inference based on the deduction and falsification of the hypothesis. According to this view, Bayesian inference is inductive in that it learns from the particular to the general by computing the posterior probability given data. This view implies that in Bayesian analysis, knowledge is brought about not through falsification, but through the accumulation of evidence in favour of the hypotheses. However, Gelman and Shalizi (2013) argued that Bayesian statistics can be best understood from a hypothetico-deductive perspective. They emphasised that the posterior probability of models can be checked through different forms of posterior predictive checks, where the implications of the models are compared to (original or other) data. Models that predict poorly can be altered or abandoned – falsified – and by exploring models that do not fit data, learning can happen (Gelman & Shalizi, 2013). This view of Bayesian inference has again been debated (e.g. Borsboom & Haig, 2013; Kruschke, 2013; Morey et al., 2013), illustrating an ongoing discussion about Bayesian philosophy, how it relates to other

philosophical traditions and whether Bayesian methods can be executed practically without subscribing to Bayesian philosophy.

3.7.5 Mixed-effects modelling

Linear mixed-effects modelling was used in all three empirical articles. The data were nested in all cases. There were several students from each class and multiple texts from each student. We assume that students in the same class are more similar than students from different classes because they are taught by the same teacher, influenced by peers, live in the same area, etc. (Lambert, 2018, p. 425). Similarly, two texts written by an individual student will be more similar than two texts written by two different students. These non-independencies can be handled in mixed-effects (multilevel/hierarchical) modelling by allowing for random effects. In the three articles, there were random intercepts for both classrooms (schools) and for students nested within classrooms. This means that the models take into account the baseline differences for both students and classrooms. In addition, there were random effects for the slopes. In Article 2, students wrote both by hand and by keyboard. By including random by-class slopes, the model allowed classes to have different slopes for the effect of the modality. In Article 3 and 4, students wrote in one modality only but on several points in time. Therefore, there were random by-time slopes for students and classrooms for the effect of time; that is, the slopes were allowed to vary by the predictor time. Mixed-effects models can account for the random – non-systematic and unpredictable – influence on data on both student and school levels, while estimating the fixed effects – the effects of interest that are expected to have systematic and predictable effects on data (Winter, 2013).

The models that were applied were also multivariate models. The dependent variables all reflect some aspect of writing, and we can assume that they correlate. The Bayesian approach allows for different modelling distributions in the same multivariate model (cf. Section

3.7.3). A multivariate model determines the random effects across, and not only within, the dependent variables.

Articles 3 and 4 included a polynomial effect. The effect of time was likely to be non-linear. Time was therefore entered into the models as both a linear component and a quadratic component. This decision was based on a model comparison of models with no time effect, a linear time effect, a quadratic time effect and a cubic time effect. This method of comparing models (LOO-CV) penalises models with more parameters to prevent overfit.

3.8 Ethical considerations

Ethical considerations have been important throughout this work. When designing the project, it was important to plan a project that would take care of the participants and develop quality research. The ethical guidelines provided by the National Committee for Research Ethics in the Social Sciences and Humanities (NESH, 2016) have been followed, and I have strived to include ethical considerations in both major and minor decisions.

As this PhD project was part of a bigger research project, formal ethical aspects such as approval of the project from the Norwegian Centre for Research Data, NSD, were done through the mother-project, *DigiHand* (documentation of the approval can be found in Appendix 4). Required information about the research project and the right to withdraw and consent forms were distributed to school leaders and teachers and to students and their parents through the *DigiHand* project. See Gamlem et al. (2020) for a more ethical discussion about the project as a whole, including literacy testing of the children at the start of the school year.

This study relied to a large extent on the cooperation and contribution of both students and teachers. According to the NESH guidelines for research ethics (NESH, 2016) children are particularly entitled to protection. The students writing texts for this project spent

time on the activity, and they were also at the very start of their school career. Therefore, an effort was made to make the writing tasks profitable and meaningful for them, aiming for a situation from which both parties could benefit: the students taking part in meaningful activities and the researcher getting good data.

Further, writing stories is an activity in which all students could participate regardless of whether they had consented to take part in the project or not, so there would be no inconvenience of leaving the classroom, missing teaching, etc. for students not being part of the study. Students who did not feel like writing on the day the writing task was conducted were encouraged by the teachers to try to write something, but if they refused or did not manage to, this was, of course, accepted.

The writing tasks were also designed to be meaningful resources for the teachers. I wanted to give teachers good writing instruction tasks in line with recommended writing instruction practices (e.g. Gerde et al., 2012). Informal reports from the teachers and students indicated positive reactions to taking part in the writing tasks. For instance, one teacher said that (s)he had never given composition tasks to students of this age and (s)he was surprised that the students could produce text as early as after three months of schooling. Other researchers of the *DigiHand* group visiting schools reported that students asked about Teddy bear Elling and were eager to know more about him.

Collecting the texts implied regular contact with the teachers. I was inspired by Guillemin and Gillam (2004) and their concepts of 'ethics in practice' and 'microethics', which highlight that interacting with participants will not necessarily give serious ethical dilemmas, but that minor ethically important moments might come up where the researcher has to make decisions that have ethical implications. When a situation arises, the right answer can be clear but not necessarily obvious to the researcher in the moment. To ensure that the knee-jerk reaction when facing big and small ethical challenges is to care for the participants, the researcher should have a reflexive practice (Guillemin & Gillam, 2004). Constant reflection on goals and practice can prepare

the researcher for such situations. I reminded myself that the participants were free to withdraw from the project at any time and of the possible imbalance of power in which teachers might feel inferior to the researcher. My aim was to communicate with the teachers in a polite and professional way and to respond ethically adequate when situations arose. In line with the NESH guidelines for research ethics (NESH, 2016), which state that researchers are responsible for clarifying the roles of involved parties, detailed instructions were made to make it clear to the teachers what they were expected to do. At the same time, I emphasised flexibility where possible, for example, with regard to how they handed in the students' texts. My impression when communicating with the teachers was that they appreciated taking part in the project and felt that they were contributing to important research.

In April 2021, all teachers were invited to a webinar with a presentation of the (preliminary) findings from the *DigiHand* project, including results from my Article 2.

Other informants that contributed to my research were the respondents to the vocabulary survey, where the content words from the children's texts were rated for written age-of-acquisition. The respondents were given information about the research purpose and participation was voluntary.

With regard to the storage of information, all data were stored securely. Both the texts and the data from the literacy tests were anonymised. The coding lists with the identifying numbers and names were stored away from the data.

Finally, for each of the empirical articles, documents related to text coding (transcription and coding manual) and statistical analysis (data and R scripts) are available on the OSF. This increases transparency and makes it possible to reproduce the statistical analysis.

4 Findings

4.1 Overall findings

The overarching research question of this thesis asks whether modality – handwriting with pencil on paper or keyboarding on a digital tablet with text-to-speech functionality – affects first graders' written composition performance and written composition learning. The short answer to this question is no. Overall, this work found that modality does not affect written composition performance in first graders and that this was the case for all students, regardless of their literacy skills. Similarly, the thesis found that modality does not affect children's learning to compose text throughout the first grade and that this (lack of) effect did not vary with student's literacy skills.

In the following, I elaborate on these findings by, for each subordinate research question, extracting and presenting relevant findings from the three empirical articles. The four subordinate research questions are as follows:

1. Does modality affect written composition performance in first graders?
2. If modality has an effect on first-grade children's composition performance, is this dependent on children's literacy skills? For example, do children with weaker literacy skills produce compositions of better quality when composing in one or another modality?
3. Does modality affect first graders' learning of written composition?
4. If modality has an effect on first-grade children's written composition learning, is this dependent on children's literacy skills? For example, do children with better literacy skills learn written composition faster when learning to write in one or another modality?

4.2 Research question 1: Modality effects on performance

The first subordinate research question concerns the modality effects on written composition performance in first graders. The findings from Article 2 showed that children who received instruction in both handwriting and keyboarding from the start of school produced narratives of similar length and quality in both modalities when assessed in the first term in first grade (November). This finding suggests that modality does not have an effect per se on the written composition performance of children who have limited experience with either handwriting or typing.

Article 3 also informs this research question. Even though the focus of the article was on written composition learning, it also provides knowledge about written composition performance in first-grade classrooms of children learning to write either by hand or on a digital tablet. For the measures reflecting text length and narrative sophistication (syntax and narrative measures), there was no evidence of a main effect of modality. This finding means that average performance across the five test occasions was not better in either the handwriting or typing classrooms for these features. When looking at performance in transcription accuracy – spelling, spacing and terminator accuracy – there was some evidence that students in the typing condition performed at a higher level than students in the handwriting condition, particularly at the first assessment occasions in November and January. For spelling accuracy, there was evidence of a main effect of modality, which means that average performance across the five test occasions was better in the typing condition. As the performance in the typing condition was better solely in the features that were directly supported by text-to-speech technology, the best explanation for this is probably that students in the typing condition managed to take advantage of the text-to-speech functionality.

4.3 Research question 2: Interaction effects on performance

Research question 2 concerns potential interaction effects between modality and students' literacy skills on written composition performance, for example, whether students with better literacy skills produce texts of higher quality when composing by typing compared to when writing by hand. Article 2 found evidence in favour of there being no interaction effect between modality and children's literacy skills on written composition performance in first graders learning to write both by hand and by keyboard.

The present thesis did not investigate the interaction effects between modality and literacy skills on performance in the sample that received writing instruction by hand or keyboard because Article 4, where this could have been examined, focused on the effects of literacy skills and modality on learning to compose text.

4.4 Research question 3: Modality effects on learning to compose

Research question 3 asks whether modality affects first graders' learning of written composition. The findings from Article 3 show that overall, modality does not affect children's learning to compose narratives in first grade. Students learning to compose by hand and students learning to compose by keyboard in otherwise business-as-usual classrooms showed similar growth in performance of the measures of text length, syntactic complexity and accuracy, event count, advanced narrative structures and story grammar during their first year of schooling.

For the composition measures of spelling, spacing and terminator accuracy, performance over time differed in the two conditions. Students in the handwriting condition improved their ability to spell, segment words and punctuate their texts throughout the school year. Performance in the typing condition started at a higher level but showed little or no

growth. As mentioned above, the higher performance level in the typing condition was probably due to the support from text-to-speech provided by the digital tablets. It was not possible to decide whether students in the typing condition learned transcription accuracy in parallel to the handwriting group. It is possible that students in the typing condition learned in a similar manner as the handwriting group, but that this learning was masked by effects at assessment – that is, that students to a descending degree used the support provided by the text-to-speech functionality. At the same time, it cannot be ruled out that typing slowed the learning of transcription accuracy in narrative compositions and that better performance from the first to the last time point was due to support from text-to-speech.

4.5 Research question 4: Interaction effects on learning to compose

Research question 4 concerns whether the (potential) modality effects on learning to compose text vary with children's literacy skills. The findings from Article 4 suggest that there are no interaction effects between modality and literacy skills on first graders' written composition learning. This implies that when learning to compose text, children with weaker – or better – literacy skills do not particularly benefit from learning to write by hand or by a digital tablet.

5 Discussion

Overall, the present thesis found that modality did not affect written composition performance and learning to compose in first graders and that the absence of such effects was not dependent on children's literacy skills at school entry. In one area, however, a difference was found: for transcription accuracy, children in the handwriting condition showed some improvement during the first grade, while children in the typing condition started at a higher performance level but showed little or no growth.

From a theoretical viewpoint, an investigation of modality effects on early composing is justified because the two modalities differ in transcription demands. For beginning writers, both cognitive demands related to letter retrieval, motor demands related to outputting letters and more practical demands like learning to use rubber or technological spelling support influence their ability to compose text. According to the capacity-sharing hypothesis, writing involves several processes that draw on the same cognitive resources, and if some processes – transcription in the case of beginning writers – occupy a substantial amount of the resources, there will be fewer resources available for other higher-level processes (McCutchen, 1996; Torrance & Galbraith, 2006). It has been argued that typing can relieve the burden of transcription for beginning writers because motor actions are easier (Beschorner & Hutchison, 2013; Genlott & Grönlund, 2013; Trageton, 2003, pp. 80–81). It might also be that recognising letters on a keyboard is easier than retrieving letters from memory. Further, typing might involve spelling support, such as text-to-speech, which might have consequences for the production and the quality of the final product.

If it were the case that typing is less resource-demanding, we would expect children to perform better when they type. The finding of no modality effect on written composition performance might therefore indicate that the two modalities place demands on children to a similar

extent. In this early stage of literacy development, children might be constrained by inscription in both modalities. Cognitive letter retrieval when handwriting can be demanding, but so can cognitive processing while typing: When a beginner writer is not yet familiar with the keyboard, substantial cognitive resources may be used on visual scanning to find the letter key, with the other letters as potential distracters (Ouellette & Tims, 2014). Forming letters by hand is typically a slower process than pressing keys (Mangen & Velay, 2010), but for children learning to write, searching for the right key might also be time consuming (Ouellette & Tims, 2014).

Alternatively, it could be that for these very young writers, demands related to inscription do not largely impact their written composition performance. As mentioned in Section 2.2, it is often assumed that a lack of inscription fluency will disrupt the writing process in such a way that the quality of the final product is affected negatively (the process-disruption hypothesis). Empirically, meta-analyses of children from kindergarten to twelfth grade have found positive effects of handwriting fluency on text quality (Feng et al., 2019; Kent & Wanzek, 2016). Intervention studies have also shown increased text quality after training in handwriting for second graders (Alves et al., 2016) and typing for fourth, fifth and sixth graders (van Weerdenburg et al., 2019). In the present work, the participants were children learning to write, and this group of writers typically has weak inscription skills, which again may lead to disfluent production of text. However, the participants were given much time to complete their compositions, meaning they could take many pauses and frequently refresh their memories. For that reason, students might have compensated for lack of fluency when composing, and the potential effects of differences in inscription demands on compositional quality might have been less pronounced. It might also be, as Rønneberg et al. (2022) pointed out, that there is no causal relationship between compositional fluency and quality of the final text, and that a third factor might affect both transcription

ability and composition ability, explaining the effects of inscription fluency on text quality in previous studies.

Further, it is also possible that the writers were heavily constrained by spelling, so the potential benefit of easier inscription might be overridden by demands related to getting the spellings on the paper/screen.

Children writing on a digital tablet could use text-to-speech technology. Such an aid might support children in spelling words more accurately, but it does not necessarily make the spelling process easier (I return to this below). Moreover, tools supporting writing will impose some burden on the writer by requiring training in using it (MacArthur, 2000). Until children have learned to use text-to-speech, this functionality might not entail any advantage. As observed in this work, the children in the typing condition produced texts that, on average, were better spelled than children in the handwriting condition, while students in the mixed condition did not write texts with better spelling accuracy when typing compared to handwriting when being assessed at an early time point in first grade. Both students in the typing condition and students in the mixed condition had the opportunity to use text-to-speech when typing. However, children in the mixed condition were probably less experienced in using this support than children in the typing condition. In the mixed condition, the writing instruction was approximately evenly divided between handwriting and typing, while in the typing condition, all time was devoted to learning to write on the keyboard. Students in the typing classes who would not need to practise handwriting would therefore have more time to become familiar with and learn to take advantage of the text-to-speech functionality than students in the mixed condition.

This thesis also investigates whether potential modality effects on written composition performance depended on the children's literacy skills measured at school entry and found no evidence that this was the case. This finding might be seen as an indication that the modalities constrain the writing of first graders to a similar extent. If typing was

cognitively less demanding than handwriting, one could have observed that some students benefitted from typing. One could, for example, expect students with weaker letter knowledge skills to produce longer or better texts when writing by keyboard than by hand. Easier inscription demands in typing could directly support children with weak letter knowledge to output their ideas as written text. Furthermore, in line with capacity theory, it could be the case that students with well-developed literacy skills, such as good vocabulary skills, could produce texts of better quality when typing than when handwriting. If typing frees resources for higher-level processing, these students might be able to take advantage of their better literacy skills when typing, and for example generate more ideas and select more diverse or precise words when typing and less constrained by inscription. However, as suggested above, there might also be alternative explanations for why no differences in length or quality were found; for example, that spelling heavily constrained children's writing and overrode easier inscription demands.

In addition to investigating modality effects on performance, the present thesis examines whether one of the modalities better supports first graders in learning to compose text. The findings indicate that the modality does not affect the rate at which children learn to compose narratives in otherwise business-as-usual classrooms. It seems that neither modality provides any advantage – or disadvantage – for learning written composition during the first year of instruction. Again, this might suggest that the modalities put similar constraints on the writers and that no extra resources are freed to support learning higher-level processes in one of the modalities.

Alternatively, it might be that one of the modalities relieves the burden of transcription, but that the children did not manage to take advantage of this. It might be that learning to develop the quality of text above transcription, for example, to organise text on the macro level, requires a substantial amount of, or a specific type of, explicit instruction. For example, Arrimada et al. (2019) reported on a 10-week intervention in Spanish first-grade classrooms in which the intervention group

received instruction in self-regulating strategies for planning content and structure in composition, and the control group received instruction focusing on reading stories and discussing the content and enhancing motivation for composing stories. Students in the intervention group produced texts with better structure, coherence, overall quality and a larger number of narrative features after the intervention than students in the control group. The present thesis studies classrooms with writing instruction based on different modalities but otherwise business-as-usual instruction. I have limited information about how teachers in these classrooms instructed children to improve features related to higher-level processes. In the teacher questionnaire, teachers were asked to report how often students took part in activities like discussing what makes a good text and discussing the macro-organisation of text (confer Section 3.3.2 and Article 3, Section 2.2). Generally, the teacher responses indicated that this was not emphasised, either in the handwriting or in the typing classrooms. It might be that, for example, children who learned to write by typing had more resources available to focus on learning high-level processes but that they had not learned enough about what makes a good text to take advantage of these extra resources.

The present thesis also examines whether modality effects on learning to compose text vary with children's literacy skills. Article 4 did not find faster or slower growth rates in text length or composition quality in one or other modalities dependent on students' literacy test scores at school entry. On the contrary, the data showed evidence in favour of there being no interaction effects between modality and literacy skills on learning to compose narratives. It seems, therefore, that students did not experience benefits or detriments when composing in one or other modality depending on their literacy skills. Again, these findings might indicate that neither modality relieves the burden of inscription when learning to compose a text. If it were the case that handwriting is more resource demanding, students who start school with weaker literacy abilities, such as weak letter knowledge, could suffer when learning to compose by hand. These students would struggle more with inscription,

and if all their resources were directed towards forming letters, they would further struggle with learning the higher-level skills involved in composing. Similarly, if typing constrained inscription less, students with good literacy skills could, hypothetically, have exploited these skills and learned higher-level processes involved in composition faster when learning writing by typing than when learning by handwriting.

This would, however, not be the case if, as mentioned above, the demands of spelling override any effect of easier inscription demands. The students who learned writing by typing used text-to-speech technology, which probably supported their performance in orthographic accuracy. However, this did not lead to any improvement in higher-level features or production of longer compositions. This finding is in line with other studies indicating no gains for length or quality, but increased focus towards, and possibly improved, spelling accuracy when using text-to-speech support when composing text (Bjørkvold & Svanes, 2021; Borgh & Patrick Dickson, 1992; Dahlström & Boström, 2017; MacArthur, 1998, 1999). Thus, the present work provides no evidence that the text-to-speech supports children in producing text. Rather, it seems that this functionality may help students correct their orthographic errors once they have produced text. If text-to-speech leads to increased focus on transcription accuracy this might override potential effects from easier inscription demands when keyboarding. If pressing keys, compared to forming letters by hand, frees resources, these resources might have been directed towards spelling accurately rather than on learning other higher-level processes involved in composing.

As mentioned in the introduction, a few previous studies on learning low-level skills involved in writing have found benefits for handwriting training compared to typing training in children who have not yet started school (Longcamp et al., 2005; Kiefer et al., 2015; Mayer et al, 2020). For example, Longcamp et al. (2005) reported that children aged 4–5 improved their letter recognition performance after handwriting training, while typing training resulted in no such learning. If handwriting is important for letter learning, it could be that first graders

with poor letter knowledge at school entry would benefit from learning to write by hand compared to using a keyboard. However, Article 3 did not find any evidence supporting this.

Looking closer at Longcamp and colleagues' (2005) study, two aspects are worth noting. First, the training consisted of children writing by hand or typing twelve letters over three weeks with one half-hour session weekly where children in each session copied each letter twice. This is far less than children would do during regular writing instruction. Second, the training consisted of copying words with the letters to learn without focusing on the meaning of the words, neither the name nor the sound of the letters. This is also different from regular classroom training, in which letters are typically rehearsed by making children explicitly aware of the letter form and the corresponding sound. Duiser et al. (2022) recently replicated Longcamp and colleagues' (2005) study with a few adjustments. Preschool children aged 5 were exposed to a three-week programme with more minutes of training than in Longcamp et al. (2005), and the training was more similar to ordinary classroom tuition. In addition, a third condition was included where children learned letters without writing, but through regular alphabet tuition, for example, through visual presentation of the letter and pronouncing the letter and words containing the letter. Duiser et al. (2022) found that children learned to recognise letters to a similar extent regardless of the training being by hand, touch typing or alphabet tuition. This study suggests that there are several routes to learning letters. The present thesis did not directly investigate students' letter recognition abilities. Nevertheless, the findings of Article 3 showed that all students achieved sufficient letter recognition ability to produce written compositions. It might be that learning letters by handwriting requires fewer repetitions than learning by typing. However, in a natural classroom setting, we can expect that children learning to write by typing will also learn letters, although it might be by different mechanisms, for example, more repetitions or explicit tuition of the sound and the physical expression of the letters.

5.1 Limitations and recommendations for future research

The research reported in this thesis investigated natural practices, and schools were not randomly allocated to conditions. Therefore, it is necessary to discuss the internal validity of the empirical studies. In Article 2, all students learned to write in both modalities, and their written composition performance was assessed in both modalities. Consequently, the writing assessment in the two modalities and the writing tasks were counterbalanced across the classes. Reports from the teacher questionnaire indicated that writing instruction in the eight classrooms entailed writing activities by both hand and digital tablet. Therefore, the internal validity of this study should be good.

The studies reported in Articles 3 and 4 compare the written composition learning of students learning to write either by hand or on a digital tablet. It was important to make sure that there were no confounding variables explaining the findings of the lack of modality effects. As reported in Section 3.3, there were no systematic differences between the conditions related to mean class size, socioeconomic status of students, students' literacy abilities due to prior literacy stimulation in kindergartens or funding of the schools. Further, it was necessary to establish that the writing instruction in the handwriting condition was actually based on pencil and paper, and that the writing instruction in the typing condition was based on a digital tablet (cf. Section 3.3.2). Evidence for this comes from what access children had to digital devices and from the teacher questionnaire in which the teachers reported their writing instruction practices. There are limitations to self-reports, and there is no guarantee that the self-reports are true accounts of what happened in the classroom. However, due to practical limits, it was not possible to apply direct measures, such as video observations of a substantial number of the lessons in each condition. However, as pointed out in the discussion in Article 3, we could have increased the internal

validity by making teachers report their classroom practices more regularly.

The present thesis explores modality effects in natural classrooms, and this yields a high degree of ecological validity. This is an advantage for communicating findings and implications to the practice field. However, it would also be valuable to know more about the mechanisms involved when children learn to compose text either by hand or on a digital tablet. As discussed in Article 3, modality might have affected classroom practice, which again might have affected children's learning to compose text. Modality might also have had individual cognitive effects on children's learning and on performance during the assessment. However, the data do not give information about the underlying mechanisms that were at play. Future research could investigate to what extent basing initial writing instruction on a digital device affects the content of the writing instruction, for example, with a shift in the type of writing activities.

This work could not determine children's development of compositional spelling in the typing condition. Because students used text-to-speech, it is not clear whether typing slowed children's learning of transcription accuracy or whether it was masked by the support provided by the text-to-speech functionality. Previous studies of training in different modalities have either found advantages for handwriting (Cunningham & Stanovich, 1990; Mayer et al., 2020) or no modality effect (Kiefer et al., 2015; Ouelette & Tims, 2014) on learning to write real words accurately or learning to spell non-words, but none of these studies have investigated typing with the use of functionality like text-to-speech. Future research could explore the effects of typing, and specifically the use of text-to-speech support, on children's spelling development, both their spelling ability in general and their compositional spelling. It could for example be that using text-to-speech can lead to increased awareness of the relation between sounds and letters. It could also be that children to a lower extent have incentive to

develop their own ability to spell when having the opportunity to lean on the technological support when composing.

This work studied children who were at the very start of their formal writing development. Findings from the empirical articles do not provide information about written composition performance and learning in subsequent years. It is possible that the growth rates of children learning to compose in one or other modality might develop differentially as they attend older grades. In particular, research is needed on the transition from learning to write in one modality to another. In Norway, this transition usually happens around the second grade for children who have received either handwriting-first or typing-first instruction. It might be that what children learn in their first school year has consequences for how they continue to learn in later years. In addition, it might be easier to go from learning to write in one modality to learning in the other, compared to learning in the opposite order.

The present thesis does not study motivation related to written composition performance and learning in different modalities, but affect and motivation are relevant components for understanding the development of writing (cf. DIEW; Kim & Park, 2019). Future research could investigate how learning to write in a specific modality influences children's attitudes towards writing, and if it is motivationally harder to go from learning to write by typing to handwriting or vice versa (cf. previous paragraph).

With regard to the questions of whether modality effects on written composition performance and learning depend on students' literacy skills, a few aspects should be noted. In Article 2, when comparing models through LOO-CV, the model that included only literacy skills as the main effect did not improve model fit relative to the intercept-only model. One can therefore question whether these measures captured the constructs of interest. However, from previous research, these measures predicted word reading and spelling ability in first graders (Solheim et al., 2018; cf. Section 3.6). In addition, the same

literacy measures predicted average writing performance in Article 4, in which more texts were included.

Relatedly, it is worth noting that in Article 4, investigating the interaction effects between modality and literacy skills on learning to compose, fairly complex models were fitted. LOO-CV, the model comparison technique used to determine the predictive accuracy of the models, penalises the addition of more parameters to prevent overfit. The poor predictive performance of the models including two- or three-way interaction effects, could be due to penalisation for many new parameters. It might be that there was insufficient data to properly test the predictive performance of the models. Future research retesting whether modality effects on learning to compose text depend on children's literacy skills could preferably include larger samples.

The findings of the present thesis must be interpreted in the context in which the studies were conducted. The participants were children who started formal instruction rather late, as there was no formal instruction in kindergarten, and the participants learned to write Norwegian, a semi-transparent language. The findings cannot necessarily be generalised to other educational contexts. For example, in other countries, children might start formal writing training earlier than in Norway, and the characteristics of languages, such as transparency, may interact with modality in the potential effects on written composition. Future research could execute similar studies of children's written composition performance and learning in other educational and language contexts.

It is also worth mentioning that children's writing performance in one genre cannot necessarily be generalised to other genres, and that the writing scores obtained in one genre therefore are not representative of children's writing proficiency as a whole (Bouwer et al., 2015). I am therefore cautious in claiming that the findings extend to the writing of other text types. However, for beginning writers, the narrative genre holds a unique position, and composing in first-grade classrooms typically takes the form of narrative compositions.

Relatedly, measuring writing skills requires multiple assignments, and more texts are generally needed from unexperienced writers to be able to generalise to their writing skills (van den Bergh et al., 2012). For future studies that might want to replicate the studies encompassing this thesis, it is worth considering whether it would be possible to collect even more texts from each participant.

5.2 Implications

5.2.1 Implications for theory and methodology

Previous studies of modality effects on written composition performance in children have typically sampled students older than first graders and students who are more experienced with handwriting than typing. These studies have indicated that students produce longer and higher-quality texts when writing by hand than by keyboard (Berninger et al., 2009; Connelly et al., 2007; Read, 2007; but see Dahlström & Boström, 2017). Article 2 of this thesis investigated modality effects on length and compositional quality in children who, from the start of formal writing instruction, learned to write both by hand and using a digital tablet, and the findings indicate that modality per se does not affect children's written composition performance. This finding was the case for all students, regardless of their literacy skills. Therefore, previous findings of better performance when children write by hand can probably be attributed to the participants being more experienced in this modality.

Arguments have been put forward about using keyboards in early writing because typing involves easier motor actions, which might make it easier for children to write longer and better compositions (Genlott & Grönlund, 2013; Trageton, 2003). The present work does not provide any evidence that producing text by hand (or keyboard) is more resource demanding in such a way that higher-level processing suffers, further resulting in shorter texts or lower text quality of the final product.

However, this finding was in a context where children could more or less spend the time they needed to complete their compositions, and their writing process was not directly investigated. The present thesis can therefore, to a limited extent, contribute to our understanding of the processes associated with composing texts in the two modalities. The findings of the thesis suggest that handwriting and keyboarding constrain children's processing to the same extent, but there are also alternative explanations discussed above (Section 5).

Previous studies comparing instruction based on handwriting or on a digital device have indicated that children have improved the quality of their compositions more when being trained in typing (e.g. Larter et al., 1987; Moore & Turner, 1988; Owston & Wideman, 1997). However, these studies are quite old, and they have not sampled students from the very start of the first grade. The present thesis, which compared groups that, from the start of formal instruction, were taught writing either by hand or by a digital tablet in otherwise business-as-usual classrooms, found that modality did not affect the rate at which children learned to compose text – with the caveat that the learning of transcription accuracy could not be determined for students writing by a digital tablet as they used text-to-speech support. In addition, the thesis found evidence in favour of there being no interaction effects between modality and children's literacy skills on learning to compose text. These findings might indicate that the modalities constrain the learning of written composition to the same extent and that no extra resources are freed to support the learning of higher-level processes in one or another modality. However, more research is needed to understand the mechanisms involved and thereby draw conclusions about the modality effects on processes related to learning to compose text.

With regard to implications for methodology, I think my work illustrates that even texts by beginning writers can be analysed analytically. Studies that measure the composition quality of first-grade writers typically apply holistic quality ratings (Abbott & Berninger, 1993; Juel et al., 1986; Kim & Park, 2019, Torrance et al., 2021), even

though some have looked at separate dimensions of children's texts (Kent et al., 2014; Kim et al., 2013, 2014). The thesis illustrates that with clear criteria, it is possible to rate features on different levels of language, even in texts that are simple and short and marked by errors. The text-analytic approach, in which features are assessed explicitly, is suitable for enhancing our knowledge about written composition performance and written composition learning in beginning writers.

5.2.2 Implications for practice

During the last few years, the debate about the use of digital devices in beginning writing instruction has been highly polarised (Arndt, 2016; Mangen & Balsvik, 2016; Wollscheid et al., 2016). It is my wish that the findings of this thesis can contribute to informing teachers, parents and policymakers. The results presented here indicate that first-grade writers with limited experience in either modality produce compositions of similar length and quality in both modalities. Thus, there does not seem to be any inherent advantage of either modality for producing text for school starters. The findings also indicate that the lack of modality effect on written composition performance does not depend on children's literacy skills. Rather, this indicates that teachers can let their whole class produce text in the same modality without worrying that students with weaker literacy skills will suffer disproportionally in their text production.

The thesis also indicates that first graders learn to compose narratives at the same rate in both modalities when the writing instruction is based on either handwriting or a digital tablet without any other fundamental changes in the instruction. This means that schools/teachers who teach children composition using a pencil on paper can continue doing so if they want, without concern that their students will learn to compose slower than students learning to write by digital tablet. On the other hand, schools/teachers who have purchased digital devices to be used in initial writing instruction can also keep using these devices and

be confident that their students will learn to compose at the same rate as children learning to write by hand.

The research reported in this thesis also suggests that the lack of modality effects on learning written composition do not depend on children's literacy skills. This again indicates that teachers can let their whole class learn composition in the same modality without worrying that students with, for example, weaker literacy skills will experience detriments in their written composition learning in the first year of school.

The data that I collected showed a large variation in the text quality of compositions written by first graders. Still, all students could take part in the composition tasks. Some students could produce a complete story, while others could only write a sentence or copy the three words they were given along with the picture. Some of the texts were not analysable according to my criteria, but children who wrote letter strings that were not conventional words could still participate in the writing activity. As described by Håland et al. (2022), providing children with the opportunity to compose full texts from the start of school is advantageous for supporting all children in learning to write, regardless of their abilities. While composing, the teacher can scaffold children where they are and give feedback on what the child needs, whether that be developing knowledge about correspondence between sound and letters, forming letters or generating relevant content. An early focus on composing full text is also in line with research-based recommendations for writing in primary grade classrooms, which, for example, state that students should be given daily opportunity to write and that there should be a balance between practicing componential skills and more communicative writing of connected text (Cutler & Graham, 2008; Gerde et al., 2012; Graham et al., 2012). Giving communicative composition tasks, such as composing for an audience like Teddy bear Elling, from an early time point in first grade is likely to provide children with good opportunities for developing composition skills, regardless of whether they learn to write by hand, by keyboard or in both modalities.

When using digital devices in initial writing instruction, many applications are available for learning letters and composing text. Many of these are commercial programmes, and teachers often have a limited chance to influence which applications are bought (Rogne et al., 2022). Research on game-based applications for learning word-reading shows the importance of teachers (adults) supporting students while using the applications (McTigue et al., 2020). Based on my research, I also recommend that teachers take the time to explore the applications for text composition and critically review them in light of pedagogical aspects. In particular, applications with text-to-speech functionality should be reviewed by teachers. Teachers should check the quality of the spoken output, specifically in relation to students' spoken dialect and the use of written standard (Bokmål or Nynorsk). Students will need training in using text-to-speech support, and teachers could use instruction in this functionality as an opportunity to develop children's metalinguistic awareness.

As noted in the Introduction, an important argument for equipping students with the personal digital devices to be used in schoolwork is to enhance learning (Islam & Grönlund, 2016; Ricoy & Sánchez-Martínez, 2020). It might therefore be somewhat disappointing that learning to write by typing does not entail benefits for learning written composition compared to traditional methods, as considerable sums of money have been spent on buying digital devices. However, in the present research, the sampled classrooms that applied digital tablets did not follow a specific pedagogical method centred around the use of digital devices. The classrooms exchanged pencil and paper with digital tablets, and to a certain extent, the writing instruction would have been adjusted due to digital tablets being the main tool for writing. For example, Article 3 indicated that modality might affect the focus of the instruction – a few teachers in the typing condition reported more time spent on story writing compared to the teachers in the handwriting condition.

However, the question remains whether all potential affordances of the digital modality were exploited in the classrooms that were studied in this work. I therefore recommend that teachers who instruct children to learn writing on a digital device, explore the possibilities that a digital tablet might give. For example, digital writing can make it easy to share and publish text and thereby find an audience, for example, peers and parents, or other classes or the general public. Digital writing might perhaps also support feedback practice. Research on teacher feedback on primary-grade children's texts suggests that the focus often is on the local level, such as spelling and grammar (Lucero et al., 2018; Lunsford & Lunsford, 2008; Rønneberg & Nilsen, 2022), though in a recent study Norwegian primary grade teachers report giving feedback both on orthography and content and structure (Håland et al., 2022). When commenting on children's digital texts, there is no need to comment on, for example, letter formation, and the scope for commenting on all levels of the text (Rønneberg & Nilsen, 2022). I would encourage teachers to explore the potential for supporting the development of higher-level features in text when children write digitally.

Finally, my work aimed to find effective ways of analysing narrative compositions by first graders. Generally, there was improvement in the text features assessed across the assessment occasions, indicating that these measures captured relevant aspects of the children's compositions. These measures can be further developed and inform the development of writing tests to diagnose the writing of beginning writers. These measures can also be adapted to first-grade teachers who want guidance on what to look for when assessing their student's texts.

5.3 Conclusion

Based on the research presented in this thesis, I conclude that modality – handwriting on paper or keyboarding on digital tablet with text-to-speech functionality – does not seem to have a profound impact on first-grade

children's written composition performance and written composition learning.

The thesis has found evidence that for first graders who are taught both handwriting and keyboarding from the start of their formal writing career, modality does not affect the length or quality of children's compositions. It has also provided evidence that neither modality offers any advantage for written composition performance depending on children's literacy skills. Thus, it does not seem that children with, for example, weaker literacy abilities will be supported more in producing compositions when writing in one or another modality.

Further, the thesis demonstrates that writing instruction based on either handwriting or typing on a digital tablet in otherwise business-as-usual classrooms does not affect the rate at which children learn to compose text in their first school year. Regardless of modality children developed their composition skills in terms of text length, syntactic complexity and accuracy, and narrative structures. Children writing by hand showed development in transcription accuracy, while for children writing by digital tablet it was not possible to determine whether they improved these skills, due to the use of text-to-speech. The thesis also found evidence that when first graders learn to compose text, children with weaker – or better – literacy skills do not particularly benefit from learning to write in one or another modality. These conclusions are based on research with Norwegian students who start formal literacy instruction in the first grade and who learn to write a semi-transparent language, and the conclusions might not be valid in other language and educational contexts.

I also reiterate that my research provides limited knowledge about the mechanisms at play when children learn to compose text by hand and on a digital tablet, and future studies exploring this further are warranted. Also, before giving clear recommendations to teachers and policymakers about the choice of modality in initial writing instruction, more studies are needed, particularly on potential transition effects in going from learning to write in one modality to another. Until this has

been established, it is of value to know that both instruction based on handwriting and instruction based on digital tablets can provide children, at least in a context similar to the one studied here, with similar opportunities for developing their written composition skills in their first year of school.

References

Abbott, R. D., & Berninger, V. W. (1993). Structural Equation Modeling of Relationships Among Developmental Skills and Writing Skills in Primary- and Intermediate-Grade Writers. *Journal of Educational Psychology*, *85*(3), 478–508. https://doi.org/10.1037/0022-0663.85.3.478

Alves, R. A., Limpo, T., Fidalgo, R., Carvalhais, L., Pereira, L. Á., & Castro, S. L. (2016). The impact of promoting transcription on early text production: Effects on bursts and pauses, levels of written language, and writing performance. *Journal of Educational Psychology*, *108*(5), 665–679. https://doi.org/10.1037/edu0000089

Andersen, S. C., Christensen, M. V., Nielsen, H. S., Thomsen, M. K., Østerbye, T., & Rowe, M. L. (2018). How reading and writing support each other across a school year in primary school children. *Contemporary Educational Psychology*, *55*(September), 129–138. https://doi.org/10.1016/j.cedpsych.2018.09.005

Andresen, B. B. (2017). The acquisition of literacy skills in 1:1 classrooms – the Danish case. *Education and Information Technologies*, *22*(2), 533–550. https://doi.org/10.1007/s10639-016-9488-7

Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology* *66*(1) 1–7. https://doi.org/10.1111/bmsp.12004

Arndt, P. A. (2016). Computer usage for learning how to read and write in primary school. *Trends in Neuroscience and Education*, *5*(3), 90–98. https://doi.org/10.1016/j.tine.2016.07.003

Arrimada, M., Torrance, M., & Fidalgo, R. (2018). Supporting first-grade writers who fail to learn: Multiple single-case evaluation of a Response to Intervention approach. *Reading and Writing*, *31*(4), 865–891. https://doi.org/10.1007/s11145-018-9817-x

Arrimada, M., Torrance, M., & Fidalgo, R. (2019). Effects of teaching planning strategies to first-grade writers. *British Journal of Educational Psychology*, *89*(4), 670–688. https://doi.org/10.1111/bjep.12251

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Berman, R. A., & Slobin, D. I. (Eds.) (2013). *Relating Events in Narrative. A Crosslinguistic Developmental Study*. Psychology Press. https://doi.org/10.4324/9780203773512

Berninger, V. W., Abbott, R. D., Augsburger, A., & Garcia, N. (2009). Comparison of pen and keyboard transcription modes in children with and without learning disabilities. *Learning Disability Quarterly*, *32*(3), 123–141. https://doi.org/10.2307/27740364

Berninger, V. W., & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 96–114). Guildford Press.

Berninger, V., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing*, *4*(3), 257–280. https://doi.org/10.1007/BF01027151

Beschorner, B., & Hutchison, A. (2013). iPads as a literacy teaching tool in early childhood. *International Journal of Education in Mathematics, Science and Technology*, *1*(1), 16–24.

Bjørkvold, T., & Svanes, I. K. (2021). Writing practices on tablets with speech synthesis in grade 1 and 2. *International Journal of Educational Research*, *107*, 101742. https://doi.org/10.1016/j.ijer.2021.101742

Book Creator. (n.d.). *Book Creator*. Retrieved January 3 from https://bookcreator.com/.

Bordens, K. S., & Abbott, B. B. (2005). *Research Design and Methods: A Process Approach* (6th edition). McGraw-Hill.

Borgh, K., & Patrick Dickson, W. (1992). The Effects on Children's Writing of Adding Speech Synthesis to a Word Processor. *Journal of Research on Computing in Education*, *24*(4), 533–544. https://doi.org/10.1080/08886504.1992.10782025

Borsboom, D., & Haig, B. D. (2013). How to practise Bayesian statistics outside the Bayesian church: What philosophy for Bayesian statistical modelling? *British Journal of Mathematical and Statistical Psychology 66(*1), 39–44. https://doi.org/10.1111/j.2044-8317.2012.02062.x

Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100. https://doi.org/10.1177/0265532214542994

Bruning, R., & Horn, C. (2000). Developing Motivation to Write. *Educational Psychologist*, *35*(1), 25–37. https://doi.org/10.1207/S15326985EP3501_4

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Charney, D. (1984). The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview. *Research in the Teaching of English*, *18*(1), 65–81.

Connelly, V., Gee, D., & Walsh, E. (2007). A comparison of keyboarded and handwritten compositions and the relationship with transcription speed. *British Journal of Educational Psychology*, *77*(2), 479–492. https://doi.org/10.1348/000709906X116768

Creswell, J. W., & Creswell, D. J. (2018). *Research design. Qualitative, quantitative and mixed methods approaches* (5th edition). Sage.

Crook, C., & Bennett, L. (2007). Does using a computer disturb the organization of children's writing? *British Journal of Developmental Psychology*, *25*(2), 313–321. https://doi.org/10.1348/026151006X143164

Cruse, A. (2011). *Meaning in language. An introduction to semantics and pragmatics* (3rd edition). Oxford University Press.

Cunningham, A., & Stanovich, K. (1990). Early spelling acquisition: writing beats the computer. *Journal of Educational Psychology*, *82*(1), 159–162.

Cutler, L., & Graham, S. (2008). Primary Grade Writing Instruction: A National Survey. *Journal of Educational Psychology*, *100*(4), 907–919. https://doi.org/10.1037/a0012656

Dahlström, H., & Boström, L. (2017). Pros and Cons: Handwriting versus digital writing. *Nordic Journal of Digital Literacy*, *12*(4), 143–161. https://doi.org/10.18261/ISSN.1891-943X-2017-04-04

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dinehart, L. H. (2015). Handwriting in early childhood education: Current research and future implications. *Journal of Early Childhood Literacy*, *15*(1), 97–118. https://doi.org/10.1177/1468798414522825

Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children's writing products: The role of curriculum based measures. *British Educational Research Journal*, *41*(4), 575–595. https://doi.org/10.1002/berj.3162

Duiser, I. H. F., Ledebt, A., van der Kamp, J., & Savelsbergh, G. J. P. (2022). Does learning to write and type make a difference in letter recognition and discrimination in primary school children? *Journal of Cognitive Psychology*, *34*(6), 691–702. https://doi.org/10.1080/20445911.2022.2060240

Dybdahl, C. S., Shaw, D. G., & Blahous, E. (1997). The impact of the computer on writing: No simple answers. *Computers in the Schools*, *13*(3–4), 41–53.

Feng, L., Lindner, A., Ji, X. R., & Malatesha Joshi, R. (2019). The roles of handwriting and keyboarding in writing: A meta-analytic review. *Reading and Writing*, *32*(1), 33–63. https://doi.org/10.1007/s11145-017-9749-x

Freeman, A. R., MacKinnon, Joyce, R., & Miller, L. T. (2005). Keyboarding for Students with Handwriting Problems. *Physical & Occupational Therapy in Pediatrics*, *25*(1–2), 119–147. https://doi.org/10.1080/J006v25n01

Froese, V. (1989). *Aspects of Holistic Scoring Validity*. [Paper presentation]. The Seventh Annual Conference on Writing Assessment, Montreal. https://files.eric.ed.gov/fulltext/ED305647.pdf

Galen, P. van. (1991). Handwriting: Issues for a psychomotor theory. *Human movement science, 10*(2–3), 165–191.

Gamlem, S. M., Rogne, W. M., Rønneberg, V., & Uppstad, P. H. (2020). Study protocol: DigiHand – the emergence of handwriting skills in digital classooms. *Nordic Journal of Literacy Research*, *6*(2), 25. https://doi.org/10.23865/njlr.v6.2115

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38. https://doi.org/10.1111/j.2044-8317.2011.02037.x

Genlott, A. A., & Grönlund, Å. (2013). Improving literacy skills through learning reading by writing: The iWTR method presented and

tested. *Computers and Education*, *67*, 98–104. https://doi.org/10.1016/j.compedu.2013.03.007

Genlott, A. A., & Grönlund, Å. (2016). Closing the gaps – Improving literacy and mathematics by ict-enhanced collaboration. *Computers and Education*, *99*, 68–80. https://doi.org/10.1016/j.compedu.2016.04.004

Gerde, H. K., Bingham, G. E., & Wasik, B. A. (2012). Writing in Early Childhood Classrooms: Guidance for Best Practices. *Early Childhood Education Journal*, *40*(6), 351–359. https://doi.org/10.1007/s10643-012-0531-z

Gorton, W. A. (2010). The Philosophy of Social Science. In *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/soc-sci/

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology*, *89*(1), 170–182. https://doi.org/10.1037/0022-0663.89.1.170

Graham, S., Bollinger, A., Olson, C. B., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching Elementary School Students to Be Effective Writers*. http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch.

Graham, S., Skar, G. B., & Falk, D. Y. (2021). Teaching writing in the primary grades in Norway: A national survey. *Reading and Writing*, 34 (2), 529–563 https://doi.org/10.1007/s11145-020-10080-y

Guillemin, M., & Gillam, L. (2004). Ethics, Reflexivity, and "Ethically Important Moments" in Research. *Qualitative Inquiry*, *10*(2), 261–280. https://doi.org/10.1177/1077800403262360

Håland, A., Hoem, T. F., & McTigue, E. M. (2019). Writing in First Grade: The Quantity and Quality of Practices in Norwegian Classrooms. *Early Childhood Education Journal*, *47*(1), 63–74. https://doi.org/10.1007/s10643-018-0908-8

Håland, A., Lundetræ, K., & Solheim Oddny Judith. (2022). Hvordan kan skriveopplæringen gi rom for et mangfold av skrivere? [How

can the writing instruction make room for a variety of writers?] In K.-A. Næss & H. Hofslundsengen (Eds.), *Skriveutvikling og skrivevansker* [*Writing development and writing disabilities*] (pp. 209–231). Cappelen Damm Akademisk.

Huot, B. (1990). The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends. *Review of Educational Research*, *60*(2), 237–263. https://doi.org/10.3102/00346543060002237

Islam, M. S., & Grönlund, Å. (2016). An international literature review of 1:1 computing in schools. *Journal of Educational Change*, *17*(2), 191–222. https://doi.org/10.1007/s10833-016-9271-y

Jones, I., & Pellegrini, A. D. (1996). The effects of social relationships, writing media, and microgenetic development on first-grade students' written narratives. *American Educational Research Journal*, *33*(3), 691–718.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*(4), 437–447. https://doi.org/10.1037//0022-0663.80.4.437

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of Literacy. A Longitudinal Study of Children in First and Second Grade. *Journal of Educational Psychology*, *78*(4), 243–255. https://doi.org/10.1037/0022-0663.78.4.243

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of writing research*, *1*(1), 1–26. https://doi.org/10.17239/jowr-2008.01.01.1

Kent, S., & Wanzek, J. (2016). The Relationship Between Component Skills and Writing Quality and Production Across Developmental Levels: A Meta-Analysis of the Last 25 Years. *Review of Educational Research*, *86*(2), 570–601. https://doi.org/10.3102/0034654315619491

Kent, S., Wanzek, J., Petscher, Y., al Otaiba, S., & Kim, Y. S. (2014). Writing fluency and quality in kindergarten and first grade: The role of attention, reading, transcription, and oral language. *Reading and*

Writing, *27*(7), 1163–1188. https://doi.org/10.1007/s11145-013-9480-1

Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., & Sachse, S. (2015). Handwriting or Typewriting? The Influence of Pen or Keyboard-Based Writing Training on Reading and Writing Performance in Preschool Children. *Advances in Cognitive Psychology*, *11*(4), 136–146. https://doi.org/10.5709/acp-0178-7

Kim, J. H., Aulck, L., Bartha, M. C., Harper, C. A., & Johnson, P. W. (2014). Differences in typing forces, muscle activity, comfort, and typing performance among virtual, notebook, and desktop keyboards. *Applied Ergonomics*, *45*(6), 1406–1413. https://doi.org/10.1016/j.apergo.2014.04.001

Kim, Y. S., al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences*, *21*(5), 517–525. https://doi.org/10.1016/j.lindif.2011.06.004

Kim, Y. S., al Otaiba, S., Sidler, J. F., & Gruelich, L. (2013). Language, literacy, attentional behaviors, and instructional quality predictors of written composition for first graders. *Early Childhood Research Quarterly*, *28*(3), 461–469. https://doi.org/10.1016/j.ecresq.2013.01.001

Kim, Y. S. G., & Park, S. H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing*, *32*(5), 1319–1343. https://doi.org/10.1007/s11145-018-9913-y

Kim, Y. S. G., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, *109*(1), 35–50. https://doi.org/10.1037/edu0000129

Kim, Y. S., Otaiba, S. al, Folsom, J. S., Greulich, L., & Puranike, C. (2014). Evaluating the dimensionality of first-grade written

composition. *Journal of Speech, Language, and Hearing Research*, *57*(1), 199–211. https://doi.org/10.1044/1092-4388(2013/12-0152)

Kim, Y.-S., al Otaiba, S., & Wanzek, J. (2015). Kindergarten predictors of third grade writing. *Learning and Individual Differences*, *37*, 27–37. https://doi.org/10.1016/j.lindif.2014.11.009

König, C., & van de Schoot, R. (2018). Bayesian statistics in educational research: A look at the current state of affairs. In *Educational Review*, *70*(4), 486–509. https://doi.org/10.1080/00131911.2017.1350636

Kruschke, J. K. (2013). Posterior predictive checks can and should be Bayesian: Comment on Gelman and Shalizi, "Philosophy and the practice of Bayesian statistics." *British Journal of Mathematical and Statistical Psychology*, *66*(1), 45–56. https://doi.org/10.1111/j.2044-8317.2012.02063.x

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd edition). Elsevier.

Kvistad, A. H., & Smemo, J. (2015). Den gode skriveoppgaven? En studie av fellestrekk ved vellykkede skriveoppgaver fra Normprosjektet. [The good writing task? A study of common features of successful writing tasks from the Norm project.] In H. Otnes (Ed.), *Å invitere elever til skriving: Ulike perspektiver på skriveoppgaver* [*Inviting students to write: Different perspectives on writing tasks*]. Fagbokforlaget.

Lambert, B. (2018). *A student's guide to Bayesian statistics*. Sage.

Larter, S., Braganca, R., & Rukavina, I. (1987). Writing with microcomputers in the elementary grades: Process, roles, attitudes, and products. *Education and Technology Series. MGS Publication Service.*

Lind, M., Simonsen, H. G., Hansen, P., Holm, E., & Mevik, B. H. (2015). Norwegian Words: A lexical database for clinicians and researchers. *Clinical Linguistics and Phonetics*, *29*(4), 276–290. https://doi.org/10.3109/02699206.2014.999952

Lingit. (n.d.). *Lingit STL+*. Accessed January 3 from https://lingit.no/programmer/lingit-stl/

Loban, W. (1976). *Language Development: Kindergarten through Grade Twelve. NCTE Committee on Research Report No. 18.* National Council of Teachers of English. https://files.eric.ed.gov/fulltext/ED128818.pdf

Longcamp, M., Zerbato-Poudou, M. T., & Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychologica*, *119*(1), 67–79. https://doi.org/10.1016/j.actpsy.2004.10.019

Lucero, M., Fernández, M. J., & Montanero, M. (2018). Teachers' written feedback comments on narrative texts in Elementary and Secondary Education. *Studies in Educational Evaluation*, *59*, 158–167. https://doi.org/10.1016/j.stueduc.2018.07.002

Lunsford, A. A., & Lunsford, K. J. (2008). "Mistakes Are a Fact of Life": A National Comparative Study. *College Composition and Communication*, *59*(4), 781–806.

MacArthur, C. A. (1998). Word Processing with Speech Synthesis and Word Prediction: Effects on the Dialogue Journal Writing of Students with Learning Disabilities. *Learning Disability Quarterly*, *21*(2), 151–166. https://doi.org/10.2307/1511342

MacArthur, C. A. (1999). Word Prediction for Students with Severe Spelling Problems. *Learning Disability Quarterly*, *22*(3), 158–172. https://doi.org/10.2307/1511283

MacArthur, C. A. (2000). New Tools for Writing. *Topics in Language Disorders*, *20*(4), 85–100. https://doi.org/10.1097/00011363-200020040-00008

MacArthur, C., & Graham, S. (2016). Writing research from a cognitive perspective. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd edition, pp. 24–40). Guildford.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development. Quantification and Assessment*. Palgrave Macmillan.

Mangen, A., & Balsvik, L. (2016). Pen or keyboard in beginning writing instruction? Some perspectives from embodied cognition. *Trends in Neuroscience and Education*, *5*(3), 99–106. https://doi.org/10.1016/j.tine.2016.06.003

Mangen, A., & Velay, J.-L. (2010). Digitizing literacy: Reflections on the haptics of writing. In M. H. Zadeh (Ed.), *Advances in haptics* (pp. 385–402). Rijeka.

Martin, J., & Rose, D. (2008). *Genre Relations. Mapping Culture*. Equinox.

Mayer, C., Wallner, S., Budde-Spengler, N., Braunert, S., Arndt, P. A., & Kiefer, M. (2020). Literacy Training of Kindergarten Children With Pencil, Keyboard or Tablet Stylus: The Influence of the Writing Tool on Reading and Writing Performance at the Letter and Word Level. *Frontiers in Psychology*, *10*(January), 1–17. https://doi.org/10.3389/fpsyg.2019.03054

McCutchen, D. (1996). A capacity theory of writing: working memory in composition. *Educational Psychology Review*, *8*(3), 299–325. https://doi.org/10.1007/BF01464076

McTigue, E. M., Solheim, O. J., Zimmer, W. K., & Uppstad, P. H. (2020). Critically Reviewing GraphoGame Across the World: Recommendations and Cautions for Research and Implementation of Computer-Assisted Instruction for Word-Reading Acquisition. *Reading Research Quarterly*, *55*(1), 45–73. https://doi.org/10.1002/rrq.256

Ministry of Education and Research. (2019). *Curriculum for Norwegian (NOR01-06). National Curriculum "Kunnskapsløftet" 2020*. https://www.udir.no/lk20/nor01-06.

Moore, M., & Turner, S. (1988). Evaluating the Use of Word Processors in Teaching Writing Composition. *Florida Journal of Educational Research*, *30*(1), 73–82.

Morey, R. D., Romeijn, J. W., & Rouder, J. N. (2013). The humble Bayesian: Model checking from a fully Bayesian perspective. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 68–75. https://doi.org/10.1111/j.2044-8317.2012.02067.x

NESH = National Committee for Research Ethics in the Social Sciences and Humanities (2016). *Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology* (4th edition). https://www.forskningsetikk.no/ressurser/publikasjoner/forskningsetiske-retningslinjer-for-samfunnsvitenskap-humaniora-juss-og-teologi/

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207

The Norwegian Directorate for Education and Training. (2017). *Framework Plan for Kindergartens*. https://www.udir.no/in-English/Framework-Plan-for-Kindergartens.

Ohlis, K. (2018, May 3). *Skoleskrift*. Retrieved January 3 from https://www.skolstil.se/no/

Ouellette, G., & Tims, T. (2014). The write way to spell: Printing vs. typing effects on orthographic learning. *Frontiers in Psychology*, *5*(February), 1–11. https://doi.org/10.3389/fpsyg.2014.00117

Owston, R. D., & Wideman, H. H. (1997). Word Processors and Children's Writing in a High-Computer-Access Setting. *Journal of Research on Computing in Education*, *30*(2), 202–220. https://doi.org/10.1080/08886504.1997.10782223

Perminger, F., Weiss, P. L. T., & Weintraub, N. (2004). Predicting Occupational Performance: Handwriting versus keyboarding. *The American Journal of Occupational Therapy*, *58*(2), 193–201.

Philips, D. C., & Burbules, C. (2000). *Postpositivism and educational research*. Rowman & Littlefield publishers.

Popper, K. (2009). Science: Conjectures and refutations. In T. McGrew, M. Alspector-Kelley, & F. Allhoff (Eds.), *The. philosophy of science: An historical anthology* (pp. 471–488). Wiley.

Read, J. C. (2007). A study of the usability of handwriting recognition for text entry by children. *Interacting with Computers*, *19*(1), 57–69. https://doi.org/10.1016/j.intcom.2006.08.009

Ricoy, M. C., & Sánchez-Martínez, C. (2020). A systematic review of tablet use in primary education. *Revista Espanola de Pedagogia*, *78(*276), 273–290. https://doi.org/10.22550/REP78-2-2020-04

Rijlaarsdam, G., & Couzijn, M. (2000). Writing and learning to write: A double challenge. In R. J. Simons (Ed.), *New learning* (pp. 157–190). Kluwer.

Rogne, W. M., Gamlem, S., & Kobberstad, L. R. (2022). Læreres bruk av digitale læringsressurser som skrivestøtte i begynneropplæringen. [Teachers' use of digital learning resources as writing support in the initial writing instruction.] In K.-A. B. Næss & H. Hofslundsengen (Eds.), *Skriveutvikling og skrivevansker* [*Writing development and writing disabilities*] (pp. 274–292). Cappelen Damm Akademisk.

Rønneberg, V., & Nilsen, C. (2022). Teachers' talk about giving feedback to young text writers, and about giving feedback on handwritten and typed texts. *Writing and Pedagogy*, *13*(1–3), 207–226.

Rønneberg, V., Torrance, M., Uppstad, P. H., & Johansson, C. (2022). The process-disruption hypothesis: How spelling and typing skill affects written composition process and product. *Psychological Research*, *86*(7), 2239–2255. https://doi.org/10.1007/s00426-021-01625-z

Rosenberg, A. (2016). *Philosophy of social science* (5th edition). Routledge.

Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2016). The Interplay between Subjectivity, Statistical Practice, and Psychological Science. *Collabra*, *2*(1). https://doi.org/10.1525/collabra.28

Salas, N., & Caravolas, M. (2019). Dimensionality of Early Writing in English and Spanish. *Journal of Literacy Research*, *51*(3), 272–292. https://doi.org/10.1177/1086296X19858146

Seymour, P. H. K., Aro, M., Erskine, M. J., & COST Action A8 network (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Solheim, O. J., Frijters, J. C., Lundetræ, K., & Uppstad, P. H. (2018). Effectiveness of an early reading intervention in a semi-transparent orthography: A group randomised controlled trial. *Learning and Instruction*, *58*, 65–79. https://doi.org/10.1016/j.learninstruc.2018.05.004

SSB = Statistisk sentralbyrå [Statistics Norway] (2021, March 26). *Kostra nøkkeltall*. [Key figures of municipalities.] Retrieved 13 June 2022 from https://www.ssb.no/kommunefakta/kostra

Sunde, K., Furnes, B., & Lundetræ, K. (2020). Does Introducing the Letters Faster Boost the Development of Children's Letter Knowledge, Word Reading and Spelling in the First Year of School? *Scientific Studies of Reading*, *24*(2), 141–158. https://doi.org/10.1080/10888438.2019.1615491

The Norwegian Directorate for Education and Training. (2020, June 25). *The Education Mirror 2019*. https://www.udir.no/in-English/Education-Mirror-2019/School/#children-in-Kindergarten

Torrance, M., Arrimada, M., & Gardner, S. (2021). Child-level factors affecting rate of learning to write in first grade. *British Journal of Educational Psychology*, *91*(2), 714–734. https://doi.org/10.1111/bjep.12390

Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 67–82). Guilford Publications.

Trageton, A. (2003). *Å skrive seg til lesing. IKT i småskolen* [*Writing for reading. ICT in primary school*]. Universitetsforlaget.

UiO = Universitet i Oslo (2021). *Digital dekning i Norges 100 største kommuner*. [Digital access in Norway's 100 largest municipalities.] Retrieved January 3 from https://www.uv.uio.no/forskning/satsinger/fiks/kunnskapsbase/digitalisering-i-skolen%20%28tidligere%20versjon%29/digital-dekning-i-norges-100-storste-kommuner/

van den Bergh, H., de Maeyer, S., van Weijen, D., & Tillema, M. (2012). Generalizability of text quality scores. In E. van Steendam, M. Tillema, G. Rijlaarsdam & H. van den Bergh (Eds.) *Measuring Writing: Recent Insights into Theory, Methodology and Practice* (pp. 23–32). Brill.

van Weerdenburg, M., Tesselhof, M., & van der Meijden, H. (2019). Touch-typing for better spelling and narrative-writing skills on the computer. *Journal of Computer Assisted Learning*, *35*(1), 143–152. https://doi.org/10.1111/jcal.12323

Vitec. (n.d.). *Vitec*. Retrieved 3 January from https://www.vitec-mv.com/No/Produkter/Intowords/

Wagenmakers, E.-J., Lodewyckx, T., Kurival, H., & Grasman, R. (2010). Bayesian Hypothesis Testing for Psychologists: A Tutorial on the Savage–Dickey Method. *Cognitive Psychology, 60*(3), 158–89.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, *24*(2), 203–220. https://doi.org/10.1007/s11145-010-9266-7

Wechsler, D. (2006). *Wechsler Individual Achievement Test – Second UK Edition (WIAT-II)*. Pearson.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. *Tutorial 2*, 1–22. http://arxiv.org/abs/1308.5499

Winter, B. (2019). *Statistics for linguists: An introduction using R.* Routledge.

Wollscheid, S., Sjaastad, J., & Tømte, C. (2016). The impact of digital devices vs. Pen(cil) and paper on primary school students' writing

skills – A research review. *Computers and Education, 95*, 19–35. https://doi.org/10.1016/j.compedu.2015.12.001

Article 1

Spilling, E. F. (2021). The measurement of text quality. In T. A. Haugen, S.-A. Myklebost, S. J. Helset, & E. Brunstad (Eds.), *Språk, tekst og medvit* [*Language, text, and consciousness*] (pp. 47–66). Cappelen Damm Akademisk. https://doi.org/10.23865/noasp.146

The Measurement of Text Quality

Eivor Finset Spilling

Abstract: This article discusses a type of study that is based on the naturalistic view of science, but where the object of inquiry is texts – man-made products of meaning. A specific study of texts written by beginning writers is used as a starting point for discussion. This study applies quantitative methods and measures meaningful structures and the quality of the texts through objective and systematic inquiry. This contrasts with the view of the interpretivist tradition, usually related to a more qualitative research approach, that emphasizes interpretation of texts. The following question is explored: How can a quantitative analysis of text quality handle meaningful structures in text? The article discusses the role of language and the researcher in the making of the analytic categories and in the coding of the texts. Further, the article highlights both the necessity of interpretation and understanding through language, and the procedures offered by quantitative methods to address this.

Keywords: interpretivism, naturalism, meaning, measurement, text quality

Introduction[1]

Texts are important objects of study in different research fields. Underlying assumptions about science will guide how texts are handled and how they are understood as an object of study. Two main approaches can be distinguished: In an interpretivist tradition the researcher typically seeks to understand and interpret the potential meaning in the text through a dialogic process (Gadamer, 2004). In a naturalist tradition, on the other hand, where aims and methods build on ideals from the natural sciences,

1 I would like to thank Vibeke Rønneberg, Wenke Mork Rogne and an anonymous reviewer for reading and commenting an earlier version of the article.

Citation of this chapter: Spilling, E. F. (2021). The Measurement of Text Quality. In T. A. Haugen, S.-A. Myklebost, S. J. Helset, & E. Brunstad (Eds.), *Språk, tekst og medvit* (pp. 47–66). Cappelen Damm Akademisk. https://doi.org/10.23865/noasp.146
License: CC-BY 4.0

text analysis will typically entail some kind of quantification of text features. In a quantitative study of texts, the combination of an object of inquiry that is a man-made product of meaning, and methods originally designed for empirical and objective investigation, can lead to a tension that is worth reflection. The aim of the article is to discuss this tension in relation to a text material where it is particularly pronounced, namely in a study of the texts of young writers.

As a point of departure for the discussion, a specific study of text quality of beginning writers' stories (Spilling et al., 2021) will be discussed. This study applies quantitative methods while investigating the written performance of first graders, both through analysis of different text features and through a score of holistic text quality. The analysis involves both counting of meaningful structures in the texts and judgments of the quality of the texts, and according to a naturalist ideal this should be done objectively. But is it possible to conduct such analyses without any interpretation? This issue is especially relevant when analysing texts written by very young writers, as they do not yet master all writing conventions. In such cases, the analysis can for instance include decisions about whether marks on a paper are meaningful text or scribbles that do not convey meaning. With this as a backdrop, the specific problem to be discussed in this article is: How can a quantitative analysis of text quality handle meaningful structures in text?

The first section of the paper will outline some general features of the naturalist and interpretivist traditions. The proceeding section concerns text quality within writing research, and briefly presents the example study with its measures and how this study relates to the different philosophical schools, specifically through its methods. Then the process of quantifying meaningful structures will be discussed in light of the different views of science represented by naturalism and interpretivism. It will be argued that interpretation and understanding through language is an important foundation for the analysis in studies of text quality influenced by naturalism. Further, it will be demonstrated that rigorous work with explicit coding rules and double rating in analysis of features that to a high degree demand interpretation, is the way that such studies secure transparency, objectivity and replicability.

Philosophical discussions within the social sciences[2]

Naturalism

The view that the social sciences should adopt the ideals from the natural sciences can be named naturalism (Gorton, 2010). The adherents will claim that the social sciences should have the same aims and methods as the natural sciences. This implies that the social sciences should be empirical, seek to find lawlike causal explanations and be value neutral (Gorton, 2010). Being empirical is in this context related to the possibility of testing a theory. The social phenomena studied must be operationalized in a way that makes them possible to measure. This can often imply that phenomena are reduced to smaller parts, like different variables that constitute hypotheses and research questions (Creswell & Creswell, 2018). Observations and experiments can then be conducted to test hypotheses, which again can contribute to the construction and testing of theories. From a naturalist view, theories should give causal explanations of social phenomena. These explanations should be as general as possible, ideally lawlike generalizations, explaining different kinds of phenomena. Value neutrality refers to the claim that scientific evidence cannot imply moral evaluations (Gorton, 2010). Further, external values like the interest of the researcher should not influence the data analysis and the testing of hypotheses and theories. This reflects the view that the object of study can be investigated objectively.

Philosophical worldviews, like naturalism, represent some broad philosophical assumptions that guide the practice of research (Creswell & Creswell, 2018). Philosophical worldviews are interconnected with research designs and research methods, and one way of describing this relation is that these three components inform the overall research approach, which can be qualitative, quantitative or mixed methods (Creswell & Creswell, 2018). Frequently, the philosophical foundation will imply a specific research approach (Creswell & Creswell, 2018). When researchers bring with them naturalist assumptions about the

2 For practical reasons the term "social sciences" is used. This also includes what often is called the humanities.

world, they also tend to apply quantitative designs and quantitative research methods. Typical quantitative designs are experiments and surveys. Quantitative methods for collecting data usually imply specification of the information to be collected in advance, with for instance predetermined and closed-ended questioning, and quantitative analysis and interpretation are typically statistical (Creswell & Creswell, 2018). Different designs and methods are often related to a specific philosophical worldview because the philosophy postulates what warrantable knowledge is, and some methods are often viewed as more suitable than others to obtain this kind of knowledge (Bryman, 1984). According to a naturalist philosophy, knowledge should be objective and replicable, and it should concern the relationship among variables, and generally, quantitative methods are appropriate to provide this kind of knowledge.

Interpretivism

A completely different view of the social world and the social sciences than the one advocated by naturalism, is represented by interpretivism. This tradition points out that the social world consists of phenomena related to human beings, and that these phenomena always carry meaning (Gorton, 2010). According to this tradition, the social sciences should aim at understanding the meanings surrounding us and not strive for making causal explanations. When investigating the social world, attention should be directed to human actions – and to intentions and beliefs underlying these actions and the context out of which these actions arise (Gorton, 2010).

Interpretivism is related to the philosophical movement of phenomenology, which is concerned with how phenomena appear to us – how objects, activities and events appear to consciousness (Moran, 2002). To investigate how the world manifests itself to us is only possible through ourselves, through the first-person point of view. This contrasts with the naturalist ideal of an objective third-person perspective. The phenomenological tradition does not reject the existence of an objective world, but argues that our experience of what exists in the natural world is not an exact copy of this (Moran, 2002). Objects from the natural world are

not seen as isolated objective elements. On the contrary, they are perceived through our consciousness, through intentionality, a directedness towards the objects (Moran, 2002). For instance, when reading a book, the reader does not experience the book from the outside. Rather, the reader intuitively knows how to handle the book as an object and directs her attention to its content. When reading, the reader experiences both the book object and the meaning of the text through her perception and cognitive abilities. Thus, in the phenomenological tradition there is not a strict division between subjects and objects as we find it in naturalism.

Another important philosophical tradition within interpretivism is hermeneutics, which is concerned with interpretation and understanding. Gadamer (2004) seeks to clarify the conditions that enable human understanding, and he uses texts as a point of departure for his theory presented in *Truth and Method* (*Wahrheit und Methode*) from 1960. Understanding, Gadamer (2004) claims, always happens from a point of view – a *horizon*. The horizon of an interpreter is decided by her prejudices, the conscious and unconscious attitudes, beliefs and knowledge that she brings with her. The horizon is not static, rather it is constantly in the process of being formed, and tradition is important in this shaping. Gadamer (2004) writes:

> Every encounter with tradition that takes place within historical consciousness involves the experience of a tension between the text and the present. The hermeneutic task consists in not covering up this tension by attempting a naive assimilation of the two but in consciously bringing it out. (p. 305)

To explore the tension, the interpreter has to project the horizon of the text – try to find out what the text claims and take a stand on these claims. The interpreter should enter into a dialogue with the text where initial prejudices are questioned, and where true prejudices are sought for. Then the horizons of the interpreter and the text can fuse, and understanding about the subject matter may be achieved (Gadamer, 2004). This fusion implies that the horizon of the interpreter has been altered and expanded, and that the interpreter understands better than before.

The hermeneutic circle, which describes the interaction between the parts and the whole, plays an important role in the process of

understanding. In traditional hermeneutics concerned with interpreting and searching for the truth of authoritative, e.g. religious, texts, the hermeneutic circle refers to the relation between the parts and the whole of a text. In Gadamer's theory (2004) the whole is not restricted to a text, but points to the horizon, in which the understanding takes place. He emphasises that "all understanding inevitably involves some prejudice" (Gadamer, 2004, p. 272). Our intellectual basis, our horizon, will guide a preliminary understanding of the parts, and the meaning of the parts will again affect the understanding of the whole, and this interaction between the parts and the whole will continue until we experience that they constitute a coherent unity of meaning. Language is also a central part of Gadamer's work on understanding: "[*L*]*anguage is the universal medium in which understanding occurs*" (Gadamer, 2004, p. 390, emphasis in original). It is language that enables the interpreter to experience the unity of meaning, that is, that enables the text to speak in a way that it makes sense to the interpreter. The inner dialogue takes place in language – we need language to investigate texts, as well as all other objects representing human activity.

Interpretivism, incorporating insights from phenomenology and hermeneutics, is a tradition that typically applies qualitative designs and methods. The relation between interpretivism and a qualitative research approach, just as the association between naturalism and quantitative research, is a tendency and not a strict relationship (Bryman, 1984). In studies of social phenomena, observation and interviews are often used, because they can give rich data (Bryman, 1984). Hermeneutic text analysis is one among many approaches to text analysis. In general, qualitative text analysis is more inductive, nonstatistical and exploratory compared to quantitative text analysis (Roberts, 2000).

Quantitative studies of text quality

Cognitive writing research and text quality

Texts are used as object of study within different research fields. One of these fields is writing research, where one important tradition is cognitive writing research. This discipline arose out of cognitive psychology,

and with insights and methods from this field the cognitive processes of writing could be investigated (MacArthur & Graham, 2016). Pioneering works are for example Hayes and Flower's (1980) model of writing as problem-solving and Bereiter and Scardamalia's (1987) work on the development of writing. The general aim of cognitive writing research is to understand writing from a cognitive point of view – to develop models that explain writing performance, learning and development (MacArthur & Graham, 2016). There is a focus on empirical research and on finding general tendencies of writing performance and learning, which clearly can be traced back to a naturalist view of science. Within this tradition, texts are an important source of data, and one way of using them is through systematic text analysis that yields a measure of text quality (also called writing quality). This measure can for instance be used to give information about writing ability, either of single students or specific populations like first-grade students, it can serve as a factor deciding to what extent a writing intervention has succeeded, or it can shed light on product or process characteristics of writing (Grabowski et al., 2014; Van Steendam et al., 2012).

A single approved and established conception of text quality does not exist. According to Van Steendam et al. (2012, p. ix), the measurement of text quality is a neglected issue in many studies of writing research: "[D]efinitions of writing quality may be absent or unclear, and operationalizations of writing quality may suffer from measurement problems." Text quality can be operationalized in many different ways. Holistic scoring of text quality entails that a text receives a single score, e.g. on a six point scale, that reflects the rater's general impression of the quality of the text (Huot, 1990b). Usually there are some benchmark texts or guidelines on which to base the assessment, and typically these focus on the content, like structure and thematic progression, but surface features, like handwriting and spelling, can also be part of it. Another approach to assessing texts is analytic scoring. Here the rater "give[s] scores to individual, identifiable traits, and these scores are tallied to provide the rating for the paper" (Huot, 1990b, p. 238). The traits that are assessed can vary; many studies of texts written by beginning writers concern text length and spelling, while others include content features. The fact that text quality

is understood in different ways can have different causes. The research questions and the genres that are investigated will for instance influence how text quality is operationalized. Contrasting conceptions of text quality can also be a result of the nature of texts – that they are man-made products with complex meaning potential.

An example study of narratives written by first graders

The study of Spilling et al. (2021) is an example of a quantitative study where text quality plays an important role, and it will be used as an example and point of departure for discussion throughout this article. This study is part of the DigiHand project (Gamlem et al., 2020), which investigates beginning writing instruction with and without the use of digital tablets. The study of Spilling et al. (2021) investigates how writing modality affects text quality in stories written by first graders and whether literacy-related skills moderate this potential effect. Texts were collected from eight different schools in the western part of Norway after students had gone to school for three months. 102 first graders wrote two narrative texts each, one by hand and one by keyboard. The two writing tasks consisted of two different picture prompts, one showing a boy about to drop his ice cream on a cat, and one of a girl about to fall down from a tree. The students were, for both tasks, instructed to write a story about what was happening in the picture. The resulting 204 texts were analysed to decide if modality affected the quality of the texts.

The quality measures in this study had to be adapted to capture central parts of texts written by students who are learning to write. Texts written by very young writers are often simple, short and incomplete. Also, on surface level, these texts can be marked by spelling errors and poor handwriting which can make it hard to identify characters and words. The texts analysed in Spilling et al. (2021) were on average 16 words long (*SD* 11 words), and the longest text was 47 words.[3] Two examples are (all

3 Texts shorter than four words were not included in the analysis. The initial sample of the study was 140 students, and of these 38 students did not manage to produce texts of four or more words in both modalities, which gave a final sample of 102 students and 204 texts.

letters standardized to lower-case, but errors of spelling, spacing and punctuation kept in the Norwegian transcriptions):

Example text 1: *en gut står i is såsken en gut jente gut står i kå så får den eine får isn med 6 kule mn auratda så sed nåke så dakulene dat ne fra isen smilte katten*
"a boy stands in the ice kiosk a boy girl boy stand in line then one of them gets the ice with 6 ball but preciselythen so something happened so whenthe balls fell down from the ice the cat smiled"

Example text 2: *iskø t gut mista is pus etis*
"iceline t boy dropped ice cat eatsice"

All texts were assessed for holistic quality – a commonly used quality measure in studies of written composition. This quality measure reflected the overall quality of the texts, and each text received a score from 0 to 5 where 0 reflected low quality and 5 high quality: Example text 1 was given a score of 5, while example text 2 was scored 3. A rubric with general level descriptions of structure, progression of ideas, coherence and vocabulary was used as guidance in the scoring. In addition, the texts were analysed through a text-analytic approach comprising measures of text length, spelling accuracy, space use accuracy, punctuation (correct use of sentence terminators), vocabulary sophistication, syntax (clause construction) and narrative structure (both on global and local level). The scores were not combined in a single sum score; on the contrary each feature was investigated separately. This approach made it possible to identify potential modality effects on specific features of the texts.

It can be discussed how the text features relate to the quality of the texts. While the holistic text quality measure is a judgment of whether the text is perceived as a good or bad story, identifying the different text features is not automatically in itself a judgment of whether the text is of high or low quality. However, which features to include involves some judgment: In this case features that can be regarded as relevant for written storytelling were chosen. Typically, a good story will have few errors on the microlevel, appropriate vocabulary, varied syntactical structures and fulfill the norms of narrative structure. A measure like

spelling accuracy, either as a count of correctly spelled words or as a ratio of correctly and incorrectly spelled words, will clearly be related to quality, as an important convention for written language is to write according to the written standard: Thus, in general, the higher spelling accuracy, the better. A measure of syntax might be related to quality in a different way. In the example study, the syntax measure was based both on the type of clause, main or subordinated clause, and on the presence or absence of syntactical errors. The texts were given a score based on the number of clauses, where subordinated clauses gave more points that main clauses, and where syntactically correct clauses gave more points than clauses with one or more syntactical errors. Subordinated clauses can be useful to express complex relations, like causal relations that are often used in stories, and therefore one might expect the use of such clauses to affect quality. However, it might be that the quality will increase up to a certain score, but that after this threshold value is reached, the quality is not affected, or affected negatively. Text length is perhaps the variable in the example study that intuitively seems less related to quality. However, the number of words produced has been shown to correlate with text quality in texts by primary-grade children (e.g. Berninger et al., 1992; Dockrell et al., 2015; Malvern et al., 2004). To be able to write a story you need to be able to produce a certain number of words, and for beginning writers, who are learning to write, the production of words in itself is probably closer related to quality than for more experienced writers.

The example study is clearly shaped within the tradition of cognitive psychology – and thus naturalism. The writing of the texts was put under careful instructions to make the conditions of writing as similar as possible. The text analysis was standardized through predefined variables. All the texts were assessed according to a manual with formalized coding rules (this is available on Open Science Foundation: https://osf.io/q8z3u/), and the content of the texts was quantified. Further, statistical calculations were done to find general patterns of the writing performance of the students. An important assumption of the study is that there are objective properties of the object of study that can be investigated by the researcher, which is also in line with a naturalistic view of science. It will, however,

be demonstrated below that insights from interpretivism are also valid for this study.

The process of measuring meaningful structures

The necessity of interpreting and understanding texts

Usually, text analysis entails analysis of meaningful structures, even though it might be possible to conduct formal text analysis where only aspects of the text that do not carry meaning are analysed. The analysis of meaningful structures will always imply some qualitative judgment on a fundamental level. Identification of a meaningful structure, to decide if a text feature is what is being searched for, depends on point of view and criteria of classification, and it has to be judged whether a text feature can be placed in a category or not. Further, some meaningful structures will require more interpretation than others. In the example study meaningful structures on different levels of language were analysed: From the micro level with the features spelling, spacing and punctuation via the meso level with vocabulary and syntax to the macro level with measures of narrative structure. The analysis of the features on the micro level requires less interpretation than the features on the meso and the macro level, as the judgments on micro level can be compared to clear norms. In digital texts, text length and spelling can, at least in theory, be analysed automatically with programmes for character count, word count and spellchecking. However, in text written by beginning writers, the counting of such features is more challenging than in texts written by experienced writers. Beginning writers do not necessarily master the correct form of all letters as they for example do not know the standard form of the letter, or motorically are not able to make well-formed letters when handwriting. Further, they do not necessarily segment words with whitespace. Then several questions arise: Is a slightly bent stroke an <i>, or is it just a line? Is an inverted <p> a <p> or a <q>? Is a mark a line or is it a conventional hyphen-sign? Is a mark a full stop-sign, or just a scroll?

Further, on word- and sentence-level, there will be similar questions. Can a letter combination like *pusvhis* "catwthice" be understood as *pus vil ha is* "cat wants to have ice"? Or can we only identify the start and the end of the letter combination as separate words, *pus* "cat" and *is* "ice" so that this is not really a sentence?

At first glance this letter string does not necessarily make sense. However, the task to which this text was a response, shows a picture of a boy dropping his ice on a cat, and this can guide the analysis. When investigating the letter string, relevant words can be found both at the start and at the end of the string. A researcher who is familiar with beginning writers also knows that segmenting words must be learned, and that a usual strategy can be to only write the first letter of a word. Therefore, one way of analysing this letter string is to split it into more parts and interpret the string as a sentence of four words. The knowledge that the researcher brings with her in the interpretation, her horizon of understanding, is not a disadvantage. On the contrary, it makes it possible to make sense of the text. Thus, prior knowledge of the researcher and an analysis of wavering between parts and whole can make apparently meaningless text parts meaningful.

There is an important distinction between physical objects and meaningful objects: Objects from the natural world, and the laws that govern them, exist independently of human beings (Gorton, 2010). Physical objects can be investigated empirically, and in the natural sciences, knowledge often builds on experience and observation. A text can be observed on a superficial level, e.g. as characters on a sheet of paper or a screen. However, the nature of texts is different from other objects typically investigated in the natural sciences. What constitutes text is meaning. Words, phrases and sentences have an expression, a physical appearance, that can be observed. At the same time, what makes these units entities of language, is that the expression is combined with meaning. Haugen (2021) argues that linguistics, the discipline that studies language, has an intuitive basis that precedes the analytic investigation of and theorizing of language. All language users will have an intuitive understanding of whether a text, a language structure, is acceptable, of whether it makes sense or not. Without this intuitive understanding it is, according to

Haugen (2021), impossible to analyse language analytically. The meaning of a text cannot be observed. Rather, it must be understood.

This point is valid for all research that uses texts, also the study discussed in this article. To code a text – to assign specific values of the different variables to the text – the researcher is obliged to understand the text. For instance, when coding the syntax of a text, which in the example study encompassed type of clause (main or subordinated) and whether the clauses were error-free or contained one or more errors, the reader first has to understand the words of the text. Then, a judgment must be made to decide whether the words constitute clauses, and finally these clauses can be compared to the relevant categories of analysis. A requirement for deciding to which categories the clauses belong, is to understand the meaning of the clauses, and this can only be done through the first-person point of view, as pointed out in phenomenological approaches.

As illustrated above, all analysis of meaningful structures in text is bound to involve some interpretation and understanding. The next section concerns how quantitative studies address this through procedures for systematic, explicit and transparent coding and coding by various raters.

Categories, operationalizations and language

In a quantitative study, the measures are of utmost importance. The measures should be properly defined and properly executed, in order to make a precise description of the object of study (Cartwright & Rundhardt, 2014). To secure accurate measurement it is necessary to find explicitly defined categories, and these categories should be defined according to the purpose of the study (Cartwright & Rundhardt, 2014).

In the study of Spilling et al. (2021), where text quality was measured, the first step was to define essential components of text quality for texts written by beginning writers. This entailed both finding relevant text features that could be quantified, and making a rubric for assessing the quality holistically. The texts were stories written by beginning writers with use of different technologies, and literature on what can be expected with regard to the genre, age group and modality could be the point of

departure for finding relevant components. The rubric for holistic quality defined both what aspects the raters should consider in their assessment, and the different quality levels of these aspects, e.g. that story structure should be evaluated, and that a text of high quality has a complete global story structure, while a text of low quality has no traces of story organization.

The text analytic approach goes further in formalizing the text analysis. The holistic rubric guides the raters, but does not define central concepts, like story structure. Also, when giving a text a single score, the contribution of and the interplay between the different components are concealed. The text analytic approach aims at investigating the different features separately. Thus, rigorous work on operationalizations was done.

Some constructs can be operationalized without much controversy, like spelling. In the example study this was measured as the number of correctly spelled words, and correctly spelled words were defined according to the official Norwegian dictionaries *Bokmålsordboka* and *Nynorskordboka* that correspond to the two written standards of Norwegian. As the spelling measure was a judgment of the ability to spell when composing, more clarifications concerning the coding had to be done compared to coding of spelling of single words (dictation). Rules to handle homographs had to be made: In one text about the boy and the ice, the ice fell *på baken*. Isolated *baken* means "the seat, buttocks", but from the context one can assume that the student intended to write *bakken* "the ground". Further, when analysing compositional spelling, questions about how to separate spelling from grammar and segmentation arise: In example text 1, where the boy gets an ice of 6 balls, it actually says *6 kule* "6 ball" in singular. Is this a grammatical error or a spelling error? In the same example text, a compound was divided; *is såsken* [iskiosken] "the ice kiosk", and two simplexes were written as one word; *auratda* [akkurat da] "precicely then". Are these spelling errors, errors of spacing, or both? Also, rules concerning the relation between *Bokmål* and *Nynorsk* had to be made: Is it acceptable to use both *Nynorsk* and *Bokmål* words in the same text (e.g. *en* "one" – *Bokmål* –, and *gut* "boy" – *Nynorsk* – in example text 1)? Different solutions can be justified with regard to these issues, but in a specific study the criteria for coding the variables have to be clear. With clearly formalized definitions the chances are high that the analysis

of a spelling measure can be executed objectively, in the sense that several raters will code the texts in the same way.

Other constructs are more difficult to operationalize than spelling. A construct of narrative structure can be operationalized in several ways. Literature on story structure (e.g. Labov & Waletsky, 1967; Martin & Rose, 2008; Peterson & McCabe, 1983) shows different ways of analysing this: in different stages/phases, in episodes governed by the goals of the protagonists, in syntactic hierarchies etc. The way that the constructs are operationalized should be decided by the researcher. Is it possible to do this in an objective way? The overall theory of the project, and other similar studies, will provide some guidelines. The researcher should also consider the specific writing task and the context. In the study used as example in this article, relevant questions concerning task and context might be: What kind of instructions did the pupils get? What kind of implicit instructions are conveyed through the context of writing in a classroom? With these answers some operationalizations of (global) narrative structure will be more reasonable than others, e.g. the context of school will imply writing a text with an introduction, a main part and a conclusion, and with labels related to the story genre: orientation, complication and resolution.

In a study of texts by beginning writers, the nature of the texts also makes it challenging to make operationalizations. Young writers' texts will often be short and incomplete, so how to handle ellipses and incoherent parts of the narrative structure, has to be described in detail in the coding rules. Is it for instance possible to make an orientation (introduction) of just one word? Example text 2 can illustrate the question: *iskø t gut mista is pus etis* "iceline t boy dropped ice cat eatsice". The answer can be debated, but a specific study needs criteria that enable the researcher to code the texts systematically. Similarly, what counts as a resolution? Is *smilte katten* "the cat smiled" in example text 1 a resolution? Is this a satisfactory way of ending the story, actually a clever resolution letting the reader draw some conclusions on her own, or is it too vague to qualify for being a resolution? Again, clearly articulated rules that decide the coding are needed. Complex constructs like narrative structure require that the researcher makes several decisions in the operationalization process. When the researcher gives reasons for the choices and is open about

the process, other researchers can judge whether this was an acceptable way of operationalizing the construct. Transparency through strictly formalized operationalizations also makes the coding easier to execute and replicate.

Narrative structure is a feature that demands more interpretation than for example text length and spelling (cf. the previous section). This makes it hard to formalize rules for the coding, and challenging to code objectively. As stories can be formulated in numerous ways, some interpretation is inevitable when a rater scores a specific text according to the coding rules. At the same time, the analysis of narrative structure provides potential for analysing complex aspects of text quality. A measure like spelling gives valuable information, but does not capture all aspects of interest in stories written by children. In a study that claims to assess text quality in beginning writers' stories, a measure of narrative structure would clearly be a relevant measure. The example study discussed in this article sought to find valid measures of text quality – measures that truly reflect quality aspects of the texts, and therefore several measures were included, a holistic quality score and measures of text features on different levels of language.

Since studies show that raters vary in their assessment of text quality, it is recommended to use multiple raters (Bouwer et al., 2015; Huot, 1990a). In this way, the degree of agreement in the coding between the raters can be calculated statistically through the measure of interrater reliability (Bordens & Abbott, 2002). As long as the agreement is acceptable, usually .7 or better, there is reason to believe that the assessments have been done objectively enough. Acceptable agreement indicates that, in spite of some variation, there is consensus about most of the coding. Language makes it possible to make stories of all kinds with different layers of meaning, which makes it unrealistic to achieve 100% agreement in the coding of features like narrative structure. Nevertheless, by accepting some divergence in the coding, it is possible to analyse complex text features within a quantitative frame. Some of the advantages with quantitative research is that it can be used to identify causal and correlational relationships between variables, and also that findings can be generalized when some prerequisites, e.g. related to sampling, are fulfilled (Bordens & Abbott, 2002). If for instance a correlational relationship between modality and text quality in texts by beginning writers is established, it is possible to

predict how students perform when writing in different modalities. Generalized knowledge about factors affecting the writing performance of a specific population, like first graders, is valuable because it can be used in decision-making in the politics of education.

Another example of a construct that could be relevant for the quality of a story, is originality. This was considered for the example study, but not included in the final measures because of the difficulties with operationalization. One reasonable way of capturing this construct could be to compare the storyline of a text to what one could expect as obvious or standard solutions given the task, and operationalize originality as solutions that are not standard, but still relevant. The standard solutions again would have to be defined, e.g. as a composition of fixed narrative phases. This probably would capture instances that clearly would be thought of as original, both by researchers and laymen. However, would this operationalization cover all of the original stories? Also, a concept of originality will typically stretch, or break with, text conventions. This makes it hard to find the boundary between the very creative and the incoherent – for example if one element in the text is so creative that it does not connect to the other elements, or if the whole text is too far from what one would expect given the task and the rest of the context. As human beings we are able to intuitively judge if a story is creative or incoherent. We can also reflect on this decision, and make the inferences and judgments in our interpretation explicit. When making rules for coding, the researcher tries to formalize such interpretations. The operationalization of a construct like originality is difficult, because originality can appear in various forms in texts. Careful descriptions of a construct will increase the chances of an acceptable interrater agreement. However, careful descriptions might also exclude other instances of originality, that are only slightly different. In the process of finding accurate measures there will be a trade-off between different considerations (Cartwright & Runhardt, 2014). When being clear about what is being measured and how, limitations of a study can be illuminated.

In quantitative studies of text quality, measures that do not to any great extent demand interpretation can quite objectively be executed in the sense that different raters will code exactly the same way. However, in some cases analysis of text quality should include features that demand

more interpretation in order to make the analysis more interesting, more complete or more accurate. Interrater reliability can be used to check and document that the agreement of the coding is acceptable, which indicates that there might be a certain amount of interpretation, but not more than what is regarded tolerable. To ensure agreement among raters the language used in the guidelines for the assessment, like holistic rubrics and formalized coding rules used for analytic assessment, has to be precise. Language offers possibilities to describe and explain with complexity and nuance, but language can also be ambiguous. Thus, the researcher is obliged to reflect on language to make clear definitions. Precise guidelines for assessment give potential for assessing aspects of the texts that to a considerable extent demand interpretation. Studies that both apply valid measures of text quality and have acceptable interrater reliability, can provide valuable knowledge about writing performance that can also inform writing instruction practices.

Concluding remarks

This article has discussed the analysis of texts in a quantitative study of writing performance in light of the philosophical traditions of naturalism and interpretivism. A study that investigates texts by beginning writers was chosen as example, as the analysis of text features and quality in these texts is challenging to execute objectively in several ways. It was argued that insights from interpretivism are also valid for this text study based on naturalism. Firstly, on a fundamental level, all text analysis of meaningful structures requires interpretation. The only way to access meaning, e.g. to understand a text, is through the first-person point of view, though one's own sensory apparatus and mental abilities. The researcher is a human being that has to understand the text – make meaning out of it – to be able to quantify the meaningful structures in the text. Further, it was argued that in the process of defining and operationalizing constructs, language plays a crucial role. Rigorous work with language in the guidelines for assessment is decisive for achieving objectivity and transparency in the coding of the texts. In quantitative studies of text quality, the measurement of interrater reliability enables analyses of

features that to a high degree require interpretation, and secures that this is done within the limits of what is considered objective enough. Thus, knowledge about the general characteristics of the text quality in first grader's written stories, can be achieved.

Eivor Finset Spilling
Volda University College
Mailbox 500
NO-6101 Volda, Norway
eivor.finset.spilling@hivolda.no

Literature

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.

Berninger, V., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing: An Interdisciplinary Journal*, *4*(3), 257–280. https://doi.org/10.1007/BF01027151

Bordens, K. S., & Abbott, B. B. (2002). *Research design and methods: A process approach* (6th ed.). McGraw-Hill.

Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100. https://doi.org/10.1177/0265532214542994

Bryman, A. (1984). The debate about qualitative and qualitative research: A question of method or epistemology? *The British Journal of Sociology*, *35*(1), 75–92. https://doi.org/10.2307/590553

Cartwright, N., & Runhardt, R. (2014). Measurement. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: A new introduction* (pp. 265–287). Oxford University Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children's writing products: The role of curriculum based measures. *British Educational Research Journal*, *41*(4), 575–595. https://doi.org/10.1002/berj.3162

Gadamer, H. G. (2004). *Truth and method*. (J. Weinsheimer & D. G. Marshall, Trans.) (2nd rev. ed.) Continuum. (Original work published 1960)

Gamlem, S. M., Rogne, W. M., Rønneberg, V., & Uppstad, P. H. (2020). Study protocol: DigiHand – the emergence of handwriting skills in digital classrooms. *Nordic Journal of Literacy Research*, *6*(2), 25–41. https://doi.org/10.23865/njlr.v6.2115

Gorton, W. A. (2010). The philosophy of social science. In *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/soc-sci/

Grabowski, J., Becker-Mrotzek, M., Knopp, M., Jost, J., & Weinzierl, C. (2014). Comparing and combining different approaches to the assessment of text quality. In D. Knorr, C. Heine, & J. Engberg (Eds.), *Methods in writing process research* (pp. 147–165). Peter Lang.

Haugen, T. A. (2021). Om språkvitenskapen som forståelsesbasert, analytisk vitenskap: Et metateoretisk essay. In T. A. Haugen, S.-A. Myklebost, S. J. Helset, & E. Brunstad (Eds.), *Språk, tekst og medvit* (pp. 11–46). Cappelen Damm Akademisk.

Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. N. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Lawrence Erlbaum Associates.

Huot, B. (1990a). Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication*, *41*(2), 201–213. https://doi.org/10.2307/358160

Huot, B. (1990b). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, *60*(2), 237–263. https://doi.org/10.3102/00346543060002237

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts. Proceedings of the 1966 annual spring meeting of the American ethnological society* (pp. 12–44). University of Washington Press.

MacArthur, C. A., & Graham, S. (2016). Writing research from a cognitive perspective. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed.) (pp. 24–40). Guildford.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan.

Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture*. Equinox.

Moran, D. (2002). *Introduction to phenomenology*. Routledge.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. Plenum Press.

Roberts, C. W. (2000). A conceptual framework for quantitative text analysis. *Quality and Quantity*, *34*(3), 259–274. https://doi.org/10.1023/A:1004780007748

Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., & Torrance, M. (2021). Handwriting versus keyboarding: Does writing modality affect quality of narratives written by beginning writers? *Reading and Writing: An Interdisciplinary Journal*. https://doi.org/10.1007/s11145-021-10169-y

Van Steendam, E., Tillema, M., & Rijlaarsdam, G. (2012). Introduction. In E. Van Steendam, M. Tillema, G. Rijlaarsdam, & H. Van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practice* (pp. ix–xxi). Brill.

Article 2

Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., and Torrance, M. (2022). Handwriting versus keyboarding: Does writing modality affect quality of narratives written by beginning writers? *Reading and Writing*, *35*(1), 129–153. https://doi.org/10.1007/s11145-021-10169-y

Reading and Writing
https://doi.org/10.1007/s11145-021-10169-y

Handwriting versus keyboarding: Does writing modality affect quality of narratives written by beginning writers?

Eivor Finset Spilling[1,2] · Vibeke Rønneberg[2] · Wenke Mork Rogne[1] · Jens Roeser[3] · Mark Torrance[2,3]

Accepted: 27 May 2021
© The Author(s) 2021

Abstract
To date, there is no clear evidence to support choosing handwriting over keyboarding or vice versa as the modality children should use when they first learn to write. 102 Norwegian first-grade children from classrooms that used both electronic touchscreen keyboard on a digital tablet and pencil-and-paper for writing instruction wrote narratives in both modalities three months after starting school and were assessed on several literacy-related skills. The students' texts were then analysed for a range of text features, and were rated holistically. Data were analysed using Bayesian methods. These permitted evaluation both of evidence in favour of a difference between modalities and of evidence in favour of there being no difference. We found moderate to strong evidence in favour of no difference between modalities. We also found moderate to strong evidence against modality effects being moderated by students' literacy ability. Findings may be specific to students who are just starting to write, but suggest that for children at this stage of development writing performance is independent of modality.

Keywords Handwriting/typing · Written composition · First-grade writers · Text quality

✉ Eivor Finset Spilling
eivor.finset.spilling@hivolda.no

[1] Department of Language and Literature, Volda University College, Mailbox 500, 6101 Volda, Norway

[2] Norwegian Reading Centre, University of Stavanger, Stavanger, Norway

[3] Psychology Department, School of Social Sciences, Nottingham Trent University, Nottingham, UK

Published online: 17 June 2021

Introduction

In primary (elementary) school, in most educational contexts, students handwrite rather than type their texts. This is despite the fact that, again in most contexts, the vast majority of post-school writing is performed by some form of keyboarding. There are various reasons for this. Many classrooms lack the resources necessary to give all children access to typing. Teachers may believe that the motor skills required for handwriting developmentally precedes those required for keyboarding and therefore more time and effort is needed to gain competency when writing on a keyboard (Stevenson & Just, 2014). Shaping letters by hand may also be seen as in some way fundamental to letter learning (James & Engelhardt, 2012).

Recently, however, in a break with tradition, some schools have started to use computers and digital tablets in the writing instruction and text production, even from the start of first grade (Gamlem et al., 2020). There are several possible reasons why typewriting might benefit young writers. Writing on a keyboard gives an easier-to-read end product that looks like the texts that students see when they are given texts to read (MacArthur, 2000). Selecting letters on a keyboard may be less cognitively demanding, particularly in younger children (Beschorner & Hutchison, 2013; but see later discussion). Typing letters is possibly motorically easier and quicker than shaping them by hand (Genlott & Grönlund, 2013). Typing on a computer also makes possible additional real-time feedback and support (e.g., spell checking).

There is, as things stand, no clear research evidence to support choosing one output modality over the other as the modality that children should use when they first learn to write texts. It is also not clear whether writing modality affects all writers in the same way, or whether the particular pattern of literacy skills that a child brings to the start of school determines the relative success of writing with pen or with keyboard. Generally, there is a lack of knowledge about the effect, if any, that writing modality has on the quality of texts written by beginning writers. Our present aim, therefore, is to establish effects of writing modality on a range of surface and substantive features of the written product of children who are just beginning to learn how to write. Further, we aim to establish whether modality effects can be moderated by child-level literacy skills.

According to the Not-so-Simple view of writing, composing written narratives requires knowledge about the narrative genre and ability to generate relevant content, low-level skills that translate the ideas into sentences and words, and strategic (executive) functions that marshal this knowledge and skill as text is produced (Berninger & Winn, 2006). The importance of the low-level transcription skills for higher-order skills is highlighted in the Direct and Indirect Effects model of Writing (DIEW; Kim & Park, 2019). DIEW specifies hierarchical structural relations between components involved in writing, where low-order transcription skills are needed for higher-order skills. Whether composing text on a computer or writing by hand, most of the underlying cognitive processes are the same, but the transcription process is not. Handwriting and typing differ in the processing

necessary for the final steps involved in outputting letters on the page, both in the motor actions necessary for forming the letter and, importantly, in the processing necessary for letter-selection.

For beginning writers the fine-motor movements required for handwriting are demanding (Dinehart, 2015). Producing a specific letter by hand requires the ability to map knowledge of the letter shape onto specific fine-motor movements that effect the pen strokes that produce the letter. Typing, particularly for beginning typists, involves motor actions that, to a large extent, do no vary from letter to letter. Using a finger to press a key is less complicated than handwriting (Connelly et al., 2007), although as typing skills develop to involve more fingers the complexity of the motor movement increases (Freeman et al., 2005).

Handwriting and typing also differ in the processing responsible for letter selection that occurs immediately before motor planning—the stage that van Galen refers to as selection of allographs (van Galen, 1991). In both handwriting and typing, the writing process must involve mechanisms for selecting individual abstract letter representations, graphemes (Bonin et al., 2012; van Galen, 1991). When handwriting, this is then followed by retrieval of a related allograph—a representation of the actual shape that the letter will form—which can then drive selection of a motor plan for generating the pen movements that will output the letter onto the page (van Galen, 1991). Unlike handwriting, where grapheme selection must involve retrieval from the writer´s memory, typists have access to an external representation—the letters that appear on the keys—which potentially cue retrieval. This is clearly not necessary for more expert typists, and particularly those who have developed the ability to type without looking at the keys. But for beginning writers, only having to recognise the letter rather than having to retrieve it on the basis of internal cues may be a substantial benefit. Note, however, that moving from recall to recognition comes at the potential cost of also presenting the writer with, at minimum 28 letters (in Norwegian) that *do not* represent the correct grapheme. Tangentially, even if the grapheme is fully retrieved, finding the correct key in the keyboard can be demanding for beginning writers who might be used to the alphabetical sequence, which is not present in the keyboard.

There are, therefore, potential differences in the demands that handwriting and typing make on a child who is learning to write. This variation in demand for transcription is likely to have knock-on effects for outputting fluency and the quality of the text that the writer produces. There is good evidence of correlation between transcription ability and the overall quality of students' completed text (Alves et al., 2016; Graham et al., 1997). This is particularly the case for beginning writers (Kim & Park, 2019). Whether handwriting or typing, struggling with the low-level demands for letter selection or formation directly prevents the children from making their ideas available as text. If a child cannot form correct letters, then they cannot output the words necessary to communicate their meaning. It may also be the case that, within a resource-limited cognitive system (McCutchen, 1996; Torrance & Galbraith, 2006) devoting attention to letter selection and output, reduces attention given to substantive features—selecting and structuring ideas, forming correct syntax. This suggests that because transcription is different when writing by hand than when typing, writing modality can influence fluency in output, and therefore has the

potential to affect text quality. It might also be that because unexperienced writers are not automated in any modality, they will be constrained by transcription, regardless of the modality. Further, as we have discussed, if there is a modality effect, predicting which modality—handwriting or typing—will benefit beginning writers, and therefore which will promote the best-quality text, is not straightforward.

Previous research on modality effects supports the notion that writing modality can have an effect on output fluency, though results are inconclusive. In a sample of 2nd, 4th, and 6th graders Berninger et al. (2009) found that alphabet recall was more rapid when typing compared to writing by hand. In the same sample, however, writing by hand was associated with writing longer essays with faster word production rate than typing. In a sentence-copying task Connelly et al. (2007) found that children from reception class to Year 6 were more fluent when writing by hand, children produced more correct letters when handwriting than when typing. Both Berninger et al. (2009) and Connelly et al. (2007) attribute the benefits of handwriting in these studies to the fact that these children were more experienced in handwriting than in typing. However, Crook and Bennett (2007) found that even in a sample of 2nd graders who had extensive experience with using computers in class, children wrote more quickly by hand than by keyboard both when writing a well-practiced text (their name or a simple sentence) and when they copied a pangram. In a study of Spanish 1st and 2nd graders Jiménez and Hernández-Cabrera (2019) looked at the effects of spelling and handwriting or typing skill on sentence-production fluency, in separate models for handwriting and keyboarding. They found that, when typing sentences, both spelling and typing skills constrain total number of correctly typed words per minute. When handwriting, only spelling constrains total number of correctly written words per minute. Jiménez and Hernández-Cabrera (2019) suggest that a possible explanation for this is that the children have not automated their typing skills, unlike their handwriting skills. From these studies of modality effects on fluency, it cannot be concluded, whether effects can be explained by modality or by experience.

There are fewer studies that have explored the effect of modality on the quality of students' texts. Again, evidence is mixed. Read (2007) found that 7- and 8-year old UK-students wrote texts that were both longer and received higher teacher-scored ratings when writing by hand compared to when they were typing. Connelly et al. (2007) too found a similar pattern of results in a slightly older sample of children. The children produced higher-quality texts, based on analytic measures of ideas and development; organisation, unity and coherence, vocabulary, sentence structure and variety; grammar and usage; capitalisation and punctuation, when they were writing by hand. In both of these studies children had already received considerable writing instruction, and importantly, they were considerably more experienced in writing by hand than by keyboard. By contrast, in a small sample of 4th grade students who had received relatively extensive keyboarding training alongside learning to handwrite, Dahlström and Boström (2017) found higher linguistic accuracy when the students wrote by keyboard.

Clearly, and as might be expected, experience with a modality will increase the probability of writing well in that modality. It may also be that students' other literacy skills—letter knowledge or spelling ability, for example—interact with the

effects of modality, i.e. the benefit that a student experiences as a result of writing by hand (or by typing) will be dependent, in part, on the student's general literacy skills. Understanding possible differential effects of modality is important for practical reasons—assuming homogeneous effects across the whole classroom may leave some children struggling—but also because this sheds light on the underlying mechanisms that result in the benefits or detriments of a particular modality. There are a range of literacy factors that may, in principle, moderate the effect of modality on written product. There is evidence that vocabulary, grammatical knowledge, single-word reading and spelling affect productivity in kindergarten and first-grade children (Kent et al., 2014; Kim et al., 2011; Puranik & AlOtaiba, 2012). There is also evidence that these factors affect the quality of written composition in first grade (Abbott & Berninger, 1993; Berninger et al., 2002; Jiménez & Hernández-Cabrera, 2019; Kent et al., 2014; Kim et al., 2013). As we have discussed, typing and handwriting, potentially at least, differ in the demands they place on young writers' letter retrieval and other low-level processes. It is therefore at least plausible that literacy-related factors that predict overall performance will also moderate the effects of modality.

The present study contributes to the limited literature investigating the effect of writing modality on compositional quality in beginning writers by addressing two questions:

1. Is written composition performance in very-beginning writers affected by whether they write by hand or by typing? As we discuss above, there are theoretical arguments on both sides of the handwriting versus typing debate, but as yet no empirical test.
2. Are modality effects moderated by a child's literacy-related abilities? Do the particular skills and abilities that a child brings to a composition task affect whether they perform better when handwriting or when typing?

To address these questions, we compared the compositional quality of texts written by hand to the quality of texts written on an electronic touchscreen keyboard[1] in a group of Norwegian first-grade students. Children were sampled within three months of start-of-school, and we cannot assume that these writers had developed automatised handwriting or typing skills. Importantly, we only sampled from classes where writing instruction involved both handwriting (pencil on paper) and typing (touch keyboard on digital tablet). This controlled for experience with each medium. Norwegian first-grade students are older than in many countries, like for example the

[1] An electronic touchscreen keyboard differs from a physical keyboard in the feedback they provide. A touchscreen keyboard provides limited tactile feedback as there is no traveling across keys (Kim et al., 2014). Moreover, on a physical keyboard writers can rest their fingers on the keys as some force is needed for activation, while this is impossible on a touchscreen keyboard as keys are activated by any physical contact with a finger (Kim et al., 2014). For experienced typists using a touchscreen keyboard has proven to slow down typing speed and accuracy (Kim et al., 2014). Very beginning writers do not, however, use touch typing, and need to see the keys they type. Therefore, we do not think there will be the same differences between a touchscreen keyboard and a physical keyboard for these writers.

UK, and start school with no formal literacy training. The students that we sampled were therefore developmentally relatively mature, but were genuinely novice writers who had received three months of writing instruction, roughly evenly divided between handwriting and keyboarding. Comparing the quality of handwritten and typed texts produced by this sample therefore provided a strong test of the impact of writing modality on the quality of very-beginning writers' texts.

Although our analysis of students' texts included a holistic quality rating—the approach adopted by nearly all of the studies that we cite above—our main focus was on text-analytic measures. This promotes transparency and replicability. It also permits evaluation of exactly what text features are affected by writing modality. The analytic methods that we used were specifically developed for describing the short, inaccurate and incomplete narratives that very early writers produce.

Method

Participants

We sampled first grade students from eight public schools (one class in each school) in the Western part of Norway. Of the 143 students in these classes three students were absent for testing. For our main analysis we omitted students who wrote fewer than four words for one or both of the handwritten or typed narrative writing tasks. We adopted a four-word threshold for two reasons. As part of the writing prompt students were given three words to use in their narratives. Writing four words indicated that students had added at least one word on their own. Four words is also sufficient to form a minimal narrative that fulfilled the writing brief (e.g. *Greina knakk. Jenta datt.* 'The branch broke. The girl fell'; *Isen datt. Pusen smiler.* 'The ice cream fell. The cat smiles'.). Our final sample consisted of 102 students, with a mean age of 6 years, 2 months ($SD = 3.5$ months) at the first data collection point. Data collection was carried out between September and November 2018. The ethical oversight agency in Norway, Norwegian Centre for Research Data, has approved the study, which is part of the DigiHand project (Gamlem et al., 2020), and it follows the ethical guidelines provided by the National Committee for Research Ethics in the Social Sciences and Humanities.

Educational context

Before starting school 97% of Norwegian three- to five-year-old children go to kindergarten (Norwegian Directorate for Education & Training, 2018). A survey completed by parents of children in our sample (88% response) indicated that all children had attended. Kindergarten does not however include any formal literacy instruction. Norwegian children start school in August the calendar year of their 6th birthday, and this is when formal teaching of letters starts. Although they have not had any formal instruction, most students can recognise and name a few letters when they start school (Sigmundsson et al., 2017). All of the students in our sample received

literacy instruction using digital tablets in parallel with handwriting instruction, and all students had a personal digital tablet provided by the school. The typing on this tablet was done by fingers on an electronic touchscreen keyboard where the font was lower case by default. In the survey completed by the parents, only nine parents reported that their child did not have access to a digital tablet at home.

We surveyed all teachers in participating classes about their writing instruction in a questionnaire. In all classes, students learned between two and three letters a week. Students took part in activities that involved writing letters by hand and finding letters on the tablet. Teachers reported that students wrote and drew short texts both by hand and by typing. Thus, the students were familiar with both handwriting and typewriting.

Written composition assessment

Writing tasks were teacher-administered, following instructions provided by the research team three months after the start of school. All classes completed the two writing tasks within the same week, and in two consecutive days. The students were introduced to a teddy bear who loves stories and who would be the audience for the students' texts. Students were then briefly introduced to the story genre. Examples of narratives were mentioned, and the students were given the following explanation: A narrative is a story about something happening, it can be something exciting, scary, sad or funny. They were then asked to write a story to a picture, answering the question: *What has happened, and what will happen next?* Two pictures were used as tasks: One picture showed a boy about to drop his ice cream, and the other picture showed a girl about to fall down from a tree. Students were given three important words corresponding to the pictures (*is* 'ice', *gut* 'boy', *pus* 'cat', and *jente* 'girl', *tre* 'tree', *ball* 'ball'). Tasks and modality were counterbalanced across classes. The students were allowed to spend 45 min on the task (including the introduction part). Students who finished their composition earlier, were instructed to read quietly in a book. As our aim was to investigate modality effects on measures of text quality, it was important to make sure that all students were given enough time to complete their composition. In order to support the students in the writing process, the teachers were instructed to encourage the students to do their best, but not to help them with for example spelling or punctuation. When writing digitally the students had the possibility to use speech synthesis where they could listen to the sounds, words, and sentences corresponding to what they wrote.

The handwritten texts were first transcribed according to a transcription manual (see Appendix 1 for example of the transcriptions). Inverted letters were corrected as long as it was clear what letter was intended (<b>/<d> substitutions were not corrected). If any characters were hard to identify a second rater was consulted. Spaces between words in the handwritten texts had to be bigger than the distance between the characters within the words to be recognised as space. All verbal text, including numbers, was transcribed, while drawings in the handwritten texts were kept out of the transcriptions and analyses. Similarly, graphical illustrations like pictures and emoticons in the digital texts were excluded.

After all texts were digitalized, texts were scored to give the following measures.

Text length

Text length was measured by counting the number of words written by the child. A word was defined as a character string which represented a phonologically plausible spelling of a Norwegian word that children might plausibly know. If a character string represented two or more plausible Norwegian words, spaces were inserted. Spaces were only inserted to create maximally long words. Character strings that were not possible to identify as a word, were coded as non-words and excluded in the text length measure.

Spacing accuracy

Spacing accuracy reflects the ability to produce text that is orthographically segmented into discrete words. Space use was counted and categorized as correct spaces, and incorrect spaces (missing spaces and overgeneralized spaces [separation of simplexes or compounds]). Punctuation was accepted as correct segmentation. Space use accuracy was scored as proportion of spaces used correctly.

Punctuation (correct use of sentence terminators)

All sentences terminators (period, question mark, exclamation mark, colon) were counted: correctly inserted terminator after sentence, and incorrectly terminators (wrongly inserted terminators, e.g. in the middle of a sentence, and missing terminators after sentence). The measure was made into a binominal variable: more correct terminators than incorrect ones, or the same number of right and wrong and more wrong than correct terminators. Of the 204 texts there were 57 texts that used one or more terminators, and of these only 14 texts had more correct than incorrect use of terminators. From this we conclude that the students in our sample had not yet learned writing conventions related to terminators. We therefore did not include this measure as an outcome variable in out analyses.

Spelling accuracy

Spelling accuracy was operationalised as the total number of correctly spelled words. Spelling is understood as a correct character string. Separation of compounds was not regarded a spelling error (rather it was measured as failure to segment correctly, cf. space use accuracy). The texts were corrected to one of the two written standards of Norwegian (Bokmål or Nynorsk) according to what would give the least number of errors.

Vocabulary sophistication

All lexical lemmas from the 280 texts were extracted, in total 270 types. A sample of 21 teachers and trainee teachers completed an online-survey in which they were

asked to respond, for each word, at what age they would expect that word to appear in children's writing, on a scale from 5 to 14 years. Our measure was, therefore, similar to age-of-acquisition (Carroll & White, 1973; Gilhooly & Logie, 1980) but with a focus on written rather than spoken acquisition.

Inter-rater reliability was relatively low, as is common in subjective ratings of age-of-acquisition (Barrow et al., 2019; mean pairwise inter-rater correlation = 0.48, $SD = 0.12$, Krippendorff's $\alpha = 0.29$). Taken together, however, ratings showed high internal consistency (Cronbach's $\alpha = 0.93$, 95% CI [0.93, 0.95]).

Our age-of-acquisition-in-writing score for each word was, therefore, the mean score across all individual ratings for that word, and our vocabulary-age score for each text was the mean of these scores across lexical-lemma types within the text.

Syntax (clause construction)

Each text received a syntax score based on number of clauses, what kind of clause (main or subordinated), and whether these clauses were syntactically correct or contained one or more syntactic errors. The calculation was done according to these rules: 1 point for every syntactically correct main clause, 0.5 point for every main clause with one or more syntactical error, 2 points for every syntactically correct subordinate clause, and 1.5 points for every subordinate clause with one or more syntactical error.

Story grammar

The global narrative structure of the texts was measured through a basic version of story grammar (Labov & Waletzky, 1967) comprising the three stages orientation, complication and resolution. A text was scored zero if it did not have any stages, one point if it contained two stages (introduction and complication or complication and resolution) or two points if all three stages of story grammar were present.

The first author coded all of the texts for story grammar, while the second and the third author coded 50 texts each. Pearson's r indicated good interrater reliability, with $r = 0.89$ and 0.88.

Basic narrative features (event count)

The basic story structure is the event, as a story usually is a chain of events linked in time (Labov & Waletzky, 1967). The number of events were counted in each text as a measure of use of simple story structures.

The first author coded all of the texts for events, while the second and the third author coded 50 texts each. Pearson's r indicated good interrater reliability with $r = 0.99$ and $r = 0.98$ for the two rater pairs, respectively.

Advanced narrative features

On local level, other narrative structures than the event are: problem, solution, reaction, effect, comment from narrator and title (Martin & Rose, 2008). These features

have in common that they relate to other features, meaning the students who apply these, were able to connect content on another basis than time, like for example causal relations. The number of each of these structures were identified and counted in the students' texts.

The first author coded all of the texts for local narrative structure, while the second and the third author coded 50 texts each. Pearson's r indicated good interrater reliability. The coding of each feature ranged from $r=0.82$ to 0.95, except for the coding of the feature solution with $r=0.60$. For this feature, raters discussed and resolved cases where ratings disagreed.

See Appendix 2 for explanation and examples of the coding of the advanced narrative features.

Holistic quality rating

Each text received a holistic score between 0 and 5, based on criteria described in a rubric (see Appendix 3). The texts were scored by the first author and the second author. Before scoring the texts, raters practiced on a set of 20 texts, and had in-depth discussions around these if there was disagreement. Pearson's r indicated extremely good interrater reliability, $r=0.99$.

Literacy-related measures

Students completed a series of literacy tests in their 2nd to 5th weeks of school. Students were tested individually by members of the DigiHand project team. All tasks, apart from the spelling test, were completed on a digital tablet. Testing sessions lasted for approximately 20 min, and testing was carried out in quiet room at the students' local school.

Grapheme-to-phoneme mapping

Students saw 24 letters from the Norwegian alphabet, in upper case, in random order, one at a time (Sunde et al., 2019). They were asked to give the sound of the letter. If they named the letter instead, they were then prompted for the sound. They were given one point for each correctly-sounded letter.

Phoneme isolation

Children's phonological segmentation ability was measured in a 10 item task in which students were asked to speak the first sound in each of 10 words (Solheim et al., 2018; Haaland et al., 2021). Words were common objects like ball '*ball'*, ost '*cheese'*, eple '*apple'*. The researcher would start with two practice trials saying: Dette er en ørn. Den første lyden i ørn er /ø/. Hva er den første lyden i ørn? *'This is an eagle. The first sound in eagle is /e/. What is the first sound in eagle?'* and then the student repeated the first sound. After two test trials the researcher only named

the items, and the student had to identify the first sound. The test was stopped if the child made two consecutive errors. Students were given one point for each correctly isolated phoneme.

Phoneme blending

Children were asked to blend a series of phonemes into a word (Solheim et al., 2018). The student was shown four images and a prerecorded voice named all pictures: (e.g., hus, mur, mus, pus—'*house, wall, mouse, cat*'). The student then heard a segmented version of one of the pictures (e.g., /p/, /u/, /s/) and was asked to point to the corresponding picture. They were given two practise trials and then eight ordinary trials of increasing difficulty. All words were regular words, consisting of three to six phonemes. The test was stopped after two consecutive errors. Maximum score was eight points.

Word reading

Single-word naming accuracy was measured by asking participants to read aloud 10 single words (Haaland et al., 2021; Solheim et al., 2018). The words were regular frequent Norwegian words. A word appeared on the screen and the student was asked to read the word. Words were presented with increasing difficulty. If the student gave the letter names or unblended phonemes, the researcher asked "Yes, which word is that?". The test was stopped after two consecutive errors. Students were given one point for each correctly read word.

Spelling

Children's spelling ability was assessed as ability to write single words from dictation with pencil on paper (Haaland et al., 2021; Solheim et al., 2018). The words were regular frequent Norwegian words, starting with two- and three-letter-words ending with five-letter-words. The researcher read a sentence and repeated the word that the student should write. One test task was modelled by the researcher, and the child was asked to write the same word. Then there were ten ordinary tasks of increasing difficulty. The test was stopped after two consecutive errors. Words were scored as correct or wrong. Recognisable attempts at shaping the correct letter were accepted. Inverted letters were accepted as long as it was not a lower case <d> or <b>. The distribution of this variable was positively skewed, with a large proportion of students scoring zero (54%). Therefore, the variable was dichotomized (0 = students who scored zero, one or two [67%], 1 = students who scored three or more).

Vocabulary

Children's productive vocabulary was assessed using a short version of the Norwegian Vocabulary test (Størksen et al., 2013), this short version has been used in previous research (Solheim et al., 2018). The students were presented with a picture

of an object on the tablet screen and asked to name the object. All students completed the 20 items. One follow-up question was allowed: If the student for example answered with a less precise name, like "bird" instead of the correct "ostrich", the researcher could ask "do you know what kind of bird?". Each correct answer gave one point.

Statistical analysis

We determined evidence for effects of modality (handwriting, typewriting) and the possible moderating effects of student literacy measures on students' text using Bayesian multivariate mixed effects models.[2] The multivariate approach permitted simultaneous modelling of effects on all text features, and the particular approach that we adopted, permitted different assumptions about the forms of the distributions of the various dependent measures.

We calculated Bayes factors (*BF*; e.g., Dienes, 2016; Wagenmakers et al., 2018) to establish evidence for effects, or for no effect. A Bayes factor of 2, for example, for the hypothesis that handwritten and typewritten texts do not differ in quality ($BF_0 = 2$), would mean that evidence from our modelling of our data that the true (population) effect is zero, is twice as strong as evidence that the population effect is not zero. By convention $BF > 5$ represents moderate evidence and $BF > 10$, strong evidence (e.g., Jeffreys, 1961; Lee & Wagenmakers, 2014). Bayes factors were calculated by the Savage-Dickey method (Dickey & Lientz, 1970).

All models included random intercepts for schools, and for children nested within schools, and with random by-school slopes. Intra-class correlations for random effects of school and of child are provided in Appendix 4. Models were fitted with vague priors for all effects (zero-centred Student's *t*-distribution with $SD = 10$ and 1 degree of freedom), due to the sensitivity of the Savage-Dickey method to choice of prior, and weakly-informative priors (e.g., McElreath, 2016, p. 35) for all other parameters.

Models were implemented in the Stan probabilistic programming language (Carpenter et al., 2017) accessed via the R brms package (Bürkner, 2018). They were run with 10,000 iterations on 3 chains with a warm-up of 5,000 iterations and no thinning. Model convergence was confirmed by the Rubin–Gelman statistic (Gelman & Rubin, 1992).

We report parameter estimates with their associated 95% probability intervals (95% PI; see for example Sorensen et al., 2016). These are sometimes also referred to as *credible intervals*.

[2] Data and scripts for statistical analysis are available via Open Science Foundation (https://osf.io/q8z3u/).

Results

As we indicated above, prior to analysis we removed from our sample students who produced texts with fewer than four words in one or both of the handwritten and tablet conditions. 121 (86%) of handwritten texts and 112 (80%) of typewritten texts contained four or more words. We did not find evidence that modality affected whether or not children wrote more than three words.[3]

Correlations among the various text measures can be found in Table 1.

Modality (handwriting vs. typing) effects

Summary statistics for text measures in the handwritten and typing conditions can be found in Table 2. These indicate little or no difference between the two conditions on any of the nine measures.

As is frequently the case when counting text features, several of the text measures were zero-inflated (the feature was absent in a disproportionately large number of texts). Count of advanced narrative features was strongly zero-inflated, and we therefore treated this measure as dichotomous (0 = text with zero or one advanced structure, 1 = text contains two or more advanced structures). Spacing accuracy was perfect in a substantial minority of students, and strongly negatively skewed. This variable was therefore also dichotomised (0 = contains errors, 1 = error-free). In both cases these were modelled with Bernoulli distributions. Event count and clause construction were also zero-inflated, to a lesser extent, and these were modelled with, respectively, zero-inflated Poisson and zero-inflated negative binomial distributions. Story grammar and holistic quality ratings were treated as ordinal scale and therefore modelled with sequential processes, count data (text length and spelling accuracy) were modelled with Poisson distributions and vocabulary age was treated as normally distributed.

Findings from the multivariate mixed effects model with modality as the fixed effect are given in Table 3. As can be seen, we found no support for a difference between handwritten and typewritten text on any of our measures. Our data gave moderate or strong evidence for no effect of modality on text length, spelling accuracy, syntax, or the extent to which the text showed basic narrative structure. There was some evidence in support of no effect on holistic quality rating, on whether or not texts showed story grammar, and on the presence or absence of features associated with advanced narrative structure. Evidence was inconclusive for spacing accuracy and vocabulary age, although in both cases lent towards no effect.

We also compared overall predictive performance (model fit) for this model (with modality as a main effect) with an intercept-only model. Leave-one-out cross

[3] Analysis was by Bayesian generalised mixed-effects model with Bernoulli link function on the ability to produce a text. We predicted text length, represented by a dichotomous dummy variable with 0 representing texts with fewer than 4 words and 1 representing longer texts, on the basis of modality (handwriting, typing). Findings indicated no strong evidence for a modality effect ($BF_1 = 0.68$). There was also no evidence of an effect of modality moderated by literacy skills: Adding interactions between literacy skill variables and modality did not improve model fit relative to a model with just main effects for modality and literacy skill.

Table 1 Bivariate correlations among text measures

	Length (words)	Spacing	Spelling	Vocabulary	Syntax	Story Grammar	Basic structures	Advanced structures
Spacing accuracy	0.09							
Spelling accuracy	0.93	0.22						
Vocabulary age	0.36	−0.12	0.30					
Syntax (clause construction)	0.88	0.07	0.81	0.42				
Story grammar	0.59	0.00	0.48	0.34	0.61			
Basic narrative structure (event count)	0.79	0.01	0.70	0.38	0.81	0.73		
Advanced narrative structure	0.69	0.04	0.62	0.29	0.71	0.69	0.80	
Holistic quality rating	0.75	0.01	0.66	0.43	0.76	0.77	0.82	0.69

Table 2 Text measures by modality

	Handwriting	Typing
Text length (words)	16, 12 (8.0, 21)	16, 12 (8.0, 22)
Spacing accuracy	0.55, 0.62 (0.33, 0.94)	0.59, 0.85 (0.34, 1.0)
Spelling accuracy	10, 8.0 (5.0, 12)	10, 8.0 (5.0, 13)
Vocabulary age	6.8, 6.8 (6.6, 7.1)	7.0, 7.0 (6.7, 7.2)
Syntax (clause construction)	2.7, 2.0 (0.50, 4.0)	3.1, 2.5 (1.0, 4.0)
Story grammar	0.72, 1.0 (0.00, 1.0)	0.79, 1.0 (0.00, 1.0)
Basic narrative structure (event count)	2.0, 2.0 (1.0, 3.0)	2.4, 2.0 (1.0, 3.0)
Advanced narrative structure	1.9, 1.0 (0.00, 3.0)	2.1, 2.0 (0.00, 3.0)
Holistic quality rating	1.9, 2.0 (1.0, 3.0)	2.0, 2.0 (1.0, 3.0)

Observed mean and median with inter-quartile range in parentheses

Table 3 Estimated effect of modality on text measures

	Estimate	BF_0	BF_1
Text length (words)	0.03 (−0.06, 0.12)	31.03	0.03
Spacing accuracy	0.76 (−0.34, 1.92)	1.36	0.74
Spelling accuracy	0.02 (−0.11, 0.14)	30.94	0.03
Vocabulary age	0.17 (0.03, 0.31)	1.35	0.74
Syntax (clause construction)	0.13 (−0.09, 0.34)	8.56	0.12
Story grammar	0.42 (−0.42, 1.19)	2.49	0.40
Basic narrative structure (event count)	0.12 (−0.19, 0.37)	8.36	0.12
Advanced narrative structure	0.15 (−1.20, 1.25)	3.40	0.29
Holistic quality rating	0.29 (−0.35, 0.93)	4.15	0.24

95% PI in parentheses

validation, using methods described by Vehtari et al. (2017) showed effectively no change in expected log predictive density when modality was added as a fixed effect to the model ($\Delta\widehat{elpd} = -1$, $SE = 7$). This indicates that, across all measures, adding modality to the model did not improve the model's ability to predict writing performance.

Are modality effects moderated by literacy skills?

We therefore failed to find evidence that either typewriting or handwriting provided benefit averaged across all students. However, it remains possible that some students benefitted while other students suffered under one or other of the modalities—leaving a mean difference around zero—and that this variation was dependent on the student's various literacy skills and abilities. Means and correlations among the literacy skills measures can be found in Table 4.

To explore this hypothesis, we used a second multivariate mixed effects model adding first the literacy-skill measures as predictors, and then the interaction

Table 4 Literacy skills measures: means and bivariate correlations

	Mean (SD)	Age	Segmentation	Blending	Spelling	Vocabulary	Word reading
Age (months)	75 (3.5)						
First-sound segmentation	6.9 (3.1)	0.06					
Blending	4.6 (2.7)	0.16	0.37				
Spelling	0.45 (0.50)	0.21	0.55	0.48			
Vocabulary	14 (2.9)	0.08	0.25	0.34	0.30		
Word reading	4.2 (3.9)	0.28	0.57	0.55	0.82	0.34	
Grapheme-phoneme mapping	12 (6.8)	0.17	0.62	0.43	0.68	0.30	0.79

between these factors and modality (handwriting, typing). Table 5 gives parameter estimates for these interaction effects and Bayes factors for no effect (BF_0). We found no evidence to support the hypothesis that literacy ability moderated the effect of modality on students writing performance. There was moderate or strong evidence in support of no modality-by-ability interaction ($BF_0 > 5$) for 51 out the 63 possible effects. BF_1 was 1.6 or lower for all possible effects.

Again, comparison of overall predictive performance of the final model indicated no improvement in predictive performance relative to an intercept-only model ($\Delta\widehat{elpd} = -26$, $SE = 11$). Although this was not the focus of our analysis, it should be noted that adding literacy skills on their own as main effects—the first stage in building the moderator model—also did not improve model fit ($\Delta\widehat{elpd} = -15$, $SE = 21$) relative to the intercept-only model).

Discussion

The aim of this study was to establish whether children who are just beginning to learn how to write, and who have received some training in written production both by handwriting and by typing, produce better text in one or other of these modalities. We operationalized "better text" in terms of both a holistic quality rating, and of measures based on analysis of orthographic, syntactic and ideational structure. The present study differs from previous research by sampling students receiving a balanced teaching of both handwriting and keyboarding at the very beginning of formal writing instruction.

Our findings were straightforward. We found no evidence that modality affects students' writing. The statistical methods that we used in this study permit us to go beyond just failing to find evidence for an effect however, and allow direct inferences about the null hypothesis (the hypothesis that the true difference between modalities is zero). We found that for four of our nine text measures (word count, spelling accuracy, successful clause construction, presence of basic narrative structure) our data provided moderate to strong evidence in favour of no effect. For other measures evidence for no effect of modality was stronger than for an effect, with holistic quality

Table 5 Student age and literacy ability as potential moderators of the effect of modality (handwriting, typing) on students' texts

	Age	First sounds segment	Blending	Speller	Vocabulary	Word reading	Grapheme-phoneme mapping
Text length (words)	−0.027, 7.4[a]	0.002, 120[a]	−0.044, 3.6	0.034, 16[a]	0.018, 66[a]	0.012, 83[a]	0.009, 140[a]
Spacing accuracy	0.041, 31[a]	0.194, 6.2[a]	−0.269, 4.7	0.050, 1.6	−0.061, 10[a]	0.516, 1.2	−0.240, 2.7
Spelling accuracy	−0.031, 8.4[a]	0.018, 65[a]	−0.045, 9.4[a]	−0.142, 8.4[a]	−0.001, 110[a]	0.032, 36[a]	0.015, 77[a]
Vocabulary age	−0.016, 77[a]	−0.000, 98[a]	0.032, 34[a]	−0.266, 4.0	0.031, 28[a]	−0.021, 54[a]	0.001, 160[a]
Syntax (clause construction)	−0.030, 31[a]	−0.035, 36[a]	−0.055, 20[a]	0.161, 5.8[a]	0.072, 7.7[a]	0.000, 40[a]	0.001, 84[a]
Story grammar	−0.036, 33[a]	−0.115, 10[a]	0.030, 17[a]	0.916, 1.3	0.100, 12[a]	−0.114, 10[a]	0.114, 6.1[a]
Basic narrative structure (event count)	−0.019, 65[a]	−0.026, 38[a]	0.006, 43[a]	0.169, 5.5[a]	0.020, 45[a]	−0.025, 33[a]	0.028, 44[a]
Advanced narrative structure	−0.055, 21[a]	−0.121, 8.4[a]	0.372, 2.4	1.04, 1.1	0.144, 8.7[a]	−0.265, 4.6	0.120, 10[a]
Holistic quality rating	−0.043, 25[a]	0.130, 9.5[a]	0.093, 12[a]	1.17, 0.92	0.103, 12[a]	−0.381, 0.63	0.082, 15[a]

Parameter Estimate and BF_0 (evidence for no effect)

[a]Indicates moderate or strong evidence in support of no moderating effect ($BF_0 > 5$). No value of BF_1 (the inverse of the BF_0 values shown here) exceeded 1.6

(BF_0=4.2), in particular, falling just short of the conventional moderate-evidence threshold.

This finding leaves open the possibility, however, that students who enter school with better literacy skills are aided in their narrative production more (or hindered less) by one or other modality. We found no evidence that this was the case, at least across the literacy skills that we assessed in this study. Our analysis of literacy skills as potential moderators of modality effects gave moderate to strong evidence in favour of no effect for a substantial majority of (putative) moderator effects, and in no case did we find evidence to support an effect.

Determining what can and cannot be concluded from these findings requires a clear understanding of the particular instructional and educational context in which our study was conducted. Two features of our sample are important. First, Norwegian literacy education starts later than in many other countries. Children in our sample did not start primary school until they were at least 5 years, 7 months, and there is no formal teaching of literacy prior to this: With a small number of possible exceptions, children in our sample will have had very little writing-specific training or practice prior to starting school. Most children enter school being able to write their name and perhaps being able to recognise (sound) some additional letters (mean of 10 for girls and 7 for boys; Sigmundsson et al., 2017). 54% of students in our sample failed to spell any high-frequency regular words correctly in the spelling test, administered at school entry. Therefore, although it is reasonable to assume that all will have entered school with an implicit understanding of narrative structure, for most students any ability to commit narrative to paper or screen will have developed in the three months between school entry and the point at which we sampled their narrative writing ability. Second, schools in the present study were specifically selected because they taught first-grade writing using a combination of handwriting on paper and typing on a digital tablet.

This specific population provides a particularly valid context in which to test theories about the direct effects of modality on the text produced by very early writers. If, for example, students entered school with much more extensive writing training, all with pen and paper—as for example is the case for first-grade students in Spain and the UK (Dockrell et al., 2016; Tolchinsky & Ríos, 2009)—then differences between modalities would be predicted purely on the basis of previous experience. This may explain why we found evidence against a modality effect in this study, while the only relevant previous studies have found better performance when children wrote by hand (Jiménez & Hernández-Cabrera, 2019, in first grade Spanish children's writing fluency, and Read, 2007, in text quality for slightly older children in the UK).

We believe, therefore, that our study provides the best test to date of the hypothesis that modality per se affects text quality in young writers. Our findings are not consistent with claims that handwriting (or typing) is fundamentally more resource demanding, diverting students' attention away from processing other features of their text. The lack of a modality effect on resource demands is further evidenced by our finding of no interaction between modality and students' literacy skills. Had it been the case that, for example, writing by keyboard reduces demands associated with letter retrieval or spelling, then we would expect students with weaker letter

retrieval or spelling skills to perform better in the typing condition. We found no evidence that this was the case. Findings from the modifier analysis in our study do, however, need to be treated with some caution, in light of the fact that we also did not find evidence of main effects of our literacy measures on the students' text.

Our study does not, however, permit conclusions about modality effects when prior training has strongly favoured handwriting (or typing, although this is currently very rare). It also has nothing to say about the potential effects of modality on children's learning. In the brief period prior to completing our assessment task students in our sample received writing instruction focusing on both handwriting and typing, and we made within-writer comparisons of modality effects. Had we compared groups of writers who received writing instruction with similar content but in different modalities, and tested within the trained modality, it is possible that a modality effect would have emerged. Similarly, our findings do not permit conclusions about either learning or performance of students as they progress through primary school. It may be that as students develop both in transcription and ideation skills their rate of learning and/or performance will become more modality-dependent.

What our findings do permit us to conclude is that students at the start of school who are given similar opportunity to practice writing by typing and by handwriting are likely to produce text of similar quality in either modality: There is no inherent or essential advantage afforded by one or other modality. On this basis we tentatively suggest that first-grade teachers should feel free to base their writing instruction on one or other, or both, of handwriting and typing, without concern that this will limit the quality of their students' text. However, research is needed to establish whether this remains true across students' primary years.

Appendices

Appendix 1: example of transcription and coding

Transcription 1 (raw text)	Transcription 1.5 (spacing corrected)	Transcription 2 (spelling corrected)	Transcription 3 (wrong punctuation removed)	Transcription 3.5 (missing punctuation inserted)	Transcription 4 (syntax markup)	Transcription 5 (narrative structure)
en gut står i is såsken en gut jente gut står i kå så får den eine får isn med 6 kule mn auratda så sed nåke så dakulene dat ne fra isen smilte katten	en gut står i issåsken en gut jente gut står i kå så får den eine får isn med 6 kule mn aurat da så sed nåke så da kulene dat ne fra isen smilte katten	ein gut står i iskiosken ein gut jente gut står i kø så får den eine får isen med 6 kule men akkurat da så skjedde noko så da kulene datt ned frå isen smilte katten	cin gut står i iskiosken cin gut jente gut står i kø så får den eine får iscn mcd 6 kule men akkurat da så skjedde noko så da kulene datt ned frå iscn smilte katten	cin gut står i iskiosken. cin gut jente gut står i kø. så får den eine får iscn mcd 6 kule men akkurat da så skjedde noko. så da kulene datt ned frå iscn smilte katten.	[M ein gut står i iskiosken] [MW ein gut jente gut står i kø] [MW så får den eine får isen med 6 kule] [M men akkurat da så skjedde noko] [M så {S da kulene datt ned frå iscn} smilte katten]	[O {E ein gut står i iskiosken} {E ein gut jente gut står i kø} {E så får den eine får isen med 6 kule}] [C {N men akkurat da så skjedde noko} {EP så da kulene datt ned frå iscn}] [R {EA smilte katten}]

Score text length 36, space use accuracy 0.86, terminator accuracy -1.00, correctly spelled words 24, vocabulary mean age 7.4, syntax (clause construction) 6, basic narrative structure 5, advanced narrative structure 3, story grammar 2, holistic quality score 5

Appendix 2: coding of advanced narrative structure

Feature	Explanation	Example (corresponding to example text below)	Number of students using one or more	
			Handwriting	Typing
Comment from the narrator	The narrator intrudes to comment on events or participants, give explanations or evaluations etc.	But why did the ice cream fall? The boy was clumsy	27	19
Problem	An undesired event or state	He lost the ice cream	62	67
Solution	Corresponds to (the consequences of) a problem, and restores order	The mother bought him a new ice cream	12	16
Effect	A consequence of a preceding event/description: a material outcome	The mother bought him a new ice cream	22	44

Feature	Explanation	Example (corresponding to example text below)	Number of students using one or more	
			Handwriting	Typing
Reaction	Attitude, feeling or thought of a participant, a consequence of a preceding event/description	He turned happy again	24	14
Title	The first words of the text, implying the content of the text	The boy and the ice cream	3	1

Example text The boy and the ice cream

Once upon a time there was a boy who bought an ice cream. He dropped the ice cream. But why did the ice cream fall? The boy was clumsy. The mother bought him a new ice cream. He turned happy again.

Each idea was coded as one or more features, for example "the mother bought him a new ice cream" was coded both as an effect and a solution

Appendix 3: rubric for holistic quality rating

Score	Criteria
0	There is no text or it's illegible, or the text is a list of words without clauses
1	The text consists of at least one clause, and often in combination with single words There are no traces of story organization, either because the text is too short, or because the text functions as simple description(s) Vocabulary is simple/immature/inaccurate/repetitive
2	The text is a simple attempt at a story with a little progression of ideas There is no global story organization, but the text can denote something happening in addition to description(s) The text contains at least two coherent clauses, but can also have elements that do not fit together or repetitions Vocabulary is in general simple and inaccurate words can appear
3	The text is a recognizable attempt at a story with some progression of ideas The text has some, but not complete, global story organization (e.g. lacks introduction or conclusion) OR the text has complete story global organization, but is very simple without details and with simple vocabulary The text contains coherent parts, but parts that do not fit together or repetitions can also appear Vocabulary is average for student´s age
4	The text can be recognized as a basic story with certain progression of ideas The text has complete global story structure, but without details or with irrelevant/repetitive details, OR the text has some, but not complete, global story organization, but with relevant details The text is mainly coherent Vocabulary is appropriate
5	The text can be recognized as a story with progression of ideas The text has complete global story structure and usually contains relevant details The text is coherent Vocabulary is appropriate and can also have one/a few words that are advanced, specific or vivid

Appendix 4: intra-class correlation coefficients for effects of school and student

Intra-class correlations from the final (main effects and interaction) model. Calculation based on Goldstein et al. (2002) Method D.

	School	Child
Text length (words)	0.003	0.047
Spacing accuracy	0.085	0.701
Spelling accuracy	0.002	0.046
Vocabulary age	0.002	0.007
Syntax (clause construction)	0.006	0.072
Story grammar	0.231	0.502
Basic narrative structure (event count)	0.023	0.076
Advanced narrative structure	0.357	0.649
Holistic quality rating	0.197	0.580

Acknowledgements This study is a part of the DigiHand Project (2018–2021) funded by the Research Council of Norway, research programme "FINNUT", Grant Number 273422. We are grateful to all members of the DigiHand team for valuable input and for participation in the data collection.

Funding Open access funding provided by Volda University College.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abbott, R., & Berninger, V. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. *Journal of Educational Psychology, 85*(3), 478–508. https://doi.org/10.1037/0022-0663.85.3.478

Alves, R. A., Limpo, T., Fidalgo, R., Carvalhais, L., Pereira, L. Á., & Castro, S. L. (2016). The impact of promoting transcription on early text production: Effects on bursts and pauses, levels of written language, and writing performance. *Journal of Educational Psychology, 108*(5), 665–679.

Barrow, C. W., Björkenstam, K. N., & Strömbergsson, S. (2019). Subjective ratings of age-of-acquisition: Exploring issues of validity and rater reliability. *Journal of Child Language, 46*, 199–213. https://doi.org/10.1017/S0305000918000363

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities, 35*(1), 39–56. https://doi.org/10.1177/002221940203500104

Berninger, V. W., Abbott, R. D., Augsburger, A., & Garcia, N. (2009). Comparison of pen and keyboard transcription modes in children with and without learning disabilities. *Learning Disability Quarterly, 32*(3), 123–141. https://doi.org/10.2307/27740364

Berninger, V. W., & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 96–114). Guildford Press.

Beschorner, B., & Hutchison, A. (2013). iPads as a literacy teaching tool in early childhood. *International Journal of Education in Mathematics, Science and Technology, 1*(1), 16–24.

Bonin, P., Roux, S., Barry, C., & Canell, L. (2012). Evidence for a limited-cascading account of written word naming. *Journal of Experimental Psychology: Learning Memory and Cognition, 38*(6), 1741–1758. https://doi.org/10.1037/a0028471

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal, 10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v076.i01

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*(1), 85–95. https://doi.org/10.1080/14640747308400325

Connelly, V., Gee, D., & Walsh, E. (2007). A comparison of keyboarded and handwritten compositions and the relationship with transcription speed. *British Journal of Educational Psychology, 77*(2), 479–492. https://doi.org/10.1348/000709906X116768

Crook, C., & Bennett, L. (2007). Does using a computer disturb the organization of children's writing? *British Journal of Developmental Psychology, 25*(2), 313–321. https://doi.org/10.1348/026151006X143164

Dahlström, H., & Boström, L. (2017). Pros and cons: Handwriting versus digital writing. *Nordic Journal of Digital Literacy, 12*(4), 143–161. https://doi.org/10.18261/ISSN.1891-943X-2017-04-04

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics, 41*(1), 214–226.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dinehart, L. H. (2015). Handwriting in early childhood education: Current research and future implications. *Journal of Early Childhood Literacy, 15*(1), 97–118. https://doi.org/10.1177/1468798414522825

Dockrell, J. E., Marshall, C. R., & Wyse, D. (2016). Teachers' reported practices for teaching writing in England. *Reading and Writing: An Interdisciplinary Journal, 29*(3), 409–434. https://doi.org/10.1007/s11145-015-9605-9

Freeman, A. R., MacKinnon, J. R., & Miller, L. T. (2005). Keyboarding for students with handwriting problems: A literature review. *Physical & Occupational Therapy in Pediatrics, 25*(1–2), 119–147. https://doi.org/10.1080/J006v25n01

Gamlem, S. M., Rogne, W. M., Rønneberg, V., & Uppstad, P. H. (2020). Study protocol: DigiHand: The emergence of handwriting skills in digital classrooms. *Nordic Journal of Literacy Research, 6*(2), 25–41.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472.

Genlott, A. A., & Grönlund, Å. (2013). Improving literacy skills through learning reading by writing: The iWTR method presented and tested. *Computers and Education: An International Journal, 67*, 98–104. https://doi.org/10.1016/j.compedu.2013.03.007

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427. https://doi.org/10.3758/BF03201693

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 1*(4), 223–231.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*(1), 170–182. https://doi.org/10.1037/0022-0663.89.1.170

Haaland, V. F., Rege, M., & Solheim, O. J. (2021). Do students learn more with an additional teacher in the classroom? Evidence from a field experiment. *Under review*.

James, K. H., & Engelhardt, L. (2012). The effects of handwriting experience on functional brain development in pre-literate children. *Trends in Neuroscience and Education, 1*(1), 32–42. https://doi.org/10.1016/j.tine.2012.08.001

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford University Press.

Jiménez, J. E., & Hernández-Cabrera, J. A. (2019). Transcription skills and written composition in Spanish beginning writers: Pen and keyboard modes. *Reading and Writing. An Interdisciplinary Journal, 32*(7), 1847–1879. https://doi.org/10.1007/s11145-018-9928-4

Kent, S., Wanzek, J., Petscher, Y., Al Otaiba, S., & Kim, Y. S. (2014). Writing fluency and quality in kindergarten and first grade: The role of attention, reading, transcription, and oral language. *Reading and Writing: An Interdisciplinary Journal, 27*(7), 1163–1188. https://doi.org/10.1007/s11145-013-9480-1

Kim, J. H., Aulck, L., Bartha, M. C., Harper, C. A., & Johnson, P. W. (2014). Differences in typing forces, muscle activity, comfort, and typing performance among virtual, notebook, and desktop keyboards. *Applied Ergonomics, 45*(6), 1406–1413. https://doi.org/10.1016/j.apergo.2014.04.001

Kim, Y. S. G., Al Otaiba, S., Sidler, J. F., & Gruelich, L. (2013). Language, literacy, attentional behaviors, and instructional quality predictors of written composition for first graders. *Early Childhood Research Quarterly, 28*(3), 461–469. https://doi.org/10.1016/j.ecresq.2013.01.001

Kim, Y. S. G., & Park, S. H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing. An Interdisciplinary Journal, 32*(5), 1319–1343. https://doi.org/10.1007/s11145-018-9913-y

Kim, Y. S. G., Wagner, R. K., Puranik, C. S., Greulich, L., Folsom, J. S., & Al Otaiba, S. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences, 21*(5), 517–525. https://doi.org/10.1016/j.lindif.2011.06.004

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal an visual arts* (pp. 12–44). American Ethnological Society.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

MacArthur, C. A. (2000). New tools for writing. *Topics in Language Disorders, 20*(4), 85–100. https://doi.org/10.1097/00011363-200020040-00008

Martin, J., & Rose, D. (2008). *Genre relations: Mapping culture*. Eqinox Publishing.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*(3), 299–325. https://doi.org/10.1007/BF01464076

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.

Norwegian Directorate for Education and Training. (2018). *The Norwegian Education Mirror*. Retrieved October 21, 2020 from https://www.udir.no/in-english/education-mirror-2019/school/#children-in-kindergarten.

Puranik, C. S., & AlOtaiba, S. (2012). Examining the contribution of handwriting and spelling to written expression in kindergarten children. *Reading and Writing. An Interdisciplinary Journal, 25*(7), 1523–1546. https://doi.org/10.1007/s11145-011-9331-x

Read, J. C. (2007). A study of the usability of handwriting recognition for text entry by children. *Interacting with Computers, 19*(1), 57–69. https://doi.org/10.1016/j.intcom.2006.08.009

Sigmundsson, H., Eriksen, A. D., Ofteland, G. S., & Haga, M. (2017). Letter-sound knowledge: Exploring gender differences in children when they start school regarding knowledge of large letters, small letters, sound large letters, and sound small letters. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2017.01539

Solheim, O. J., Frijters, J. C., Lundetræ, K., & Uppstad, P. H. (2018). Effectiveness of an early reading intervention in a semi-transparent orthography: A group randomised controlled trial. *Learning and Instruction, 58*, 65–79. https://doi.org/10.1016/j.learninstruc.2018.05.004

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology, 12*(3), 175–200. https://doi.org/10.20982/tqmp.12.3.p175

Stevenson, N. C., & Just, C. (2014). In early education, why teach handwriting before keyboarding? *Early Childhood Education Journal, 42*, 49–56. https://doi.org/10.1007/s10643-012-0565-2

Størksen, I., Ellingsen, I., Tvedt, M. S., & Idsø, E. M. C. (2013). Norsk vokabulartest (NVT) for barn i overgangen mellom barnehage og skole: Psykometrisk vurdering av en nettbrettbasert test. *Spesialpedagogikk Forskningsutgaven, 4*(13), 40–54.

Sunde, K., Furnes, B., & Lundetræ, K. (2019). Does introducing the letters faster boost the development of children's letter knowledge, word reading and spelling in the first year of school? *Scientific Studies of Reading*. *24*(2),141–158. https://doi.org/10.1080/10888438.2019.1615491

Tolchinsky, L., & Ríos, I. (2009). Qué dicen los maestros que hacen para enseñar a leer y a escribir? *Aula de Innovación Educativa, 179*, 24–28.

Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67–82). Guilford Publications.

van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science, 10*(2–3), 165–191. https://doi.org/10.1016/0167-9457(91)90003-G

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and {WAIC}. *Statistics and Computing, 27*(5), 1413–1432.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Smíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Article 3

Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., and Torrance, M. (2023). Writing by hand or by keyboard: Does modality affect rate of learning to compose text in first grade? Under review, 2nd revision, *Computers & Education*.

This article is not available in Brage. However, the article has been published in *Computers & Education* (open access) and is available at: https://www.sciencedirect.com/science/article/pii/S0360131523000325

Article 4

Spilling, E. F. (2023). Effects of literacy skills on learning to compose narratives: A comparison of children writing by hand and by keyboard. Manuscript prepared for submission.

This article is not available in Brage.

Appendices

Appendix 1 – Author contributions of the four articles

Article 1

Eivor Finset Spilling was responsible for conceptualising the research question, and for writing the article. Vibeke Rønneberg and Wenke Mork Rogne commented on the manuscript.

Article 2

Eivor Finset Spilling, Vibeke Rønneberg and Mark Torrance contributed to the study conception and design. Eivor Finset Spilling, Vibeke Rønneberg and Wenke Mork Rogne performed the data collection and the coding of the data. Statistical analysis was performed by Jens Roeser and Mark Torrance. The first draft of the manuscript was written by Eivor Finset Spilling, and all authors commented on previous versions of the manuscript and/or rewrote/revised parts of the text. All authors read and approved the final manuscript. Wenke Mork Rogne secured the funding for the study.

Article 3

Eivor Finset Spilling, Vibeke Rønneberg and Mark Torrance contributed to the conception of research goals and design. Eivor Finset Spilling was responsible for the data collection. Eivor Finset Spilling, Vibeke Rønneberg and Wenke Mork Rogne coded the data. Jens Roeser and Mark Torrance executed the statistical analysis. Eivor Finset Spilling

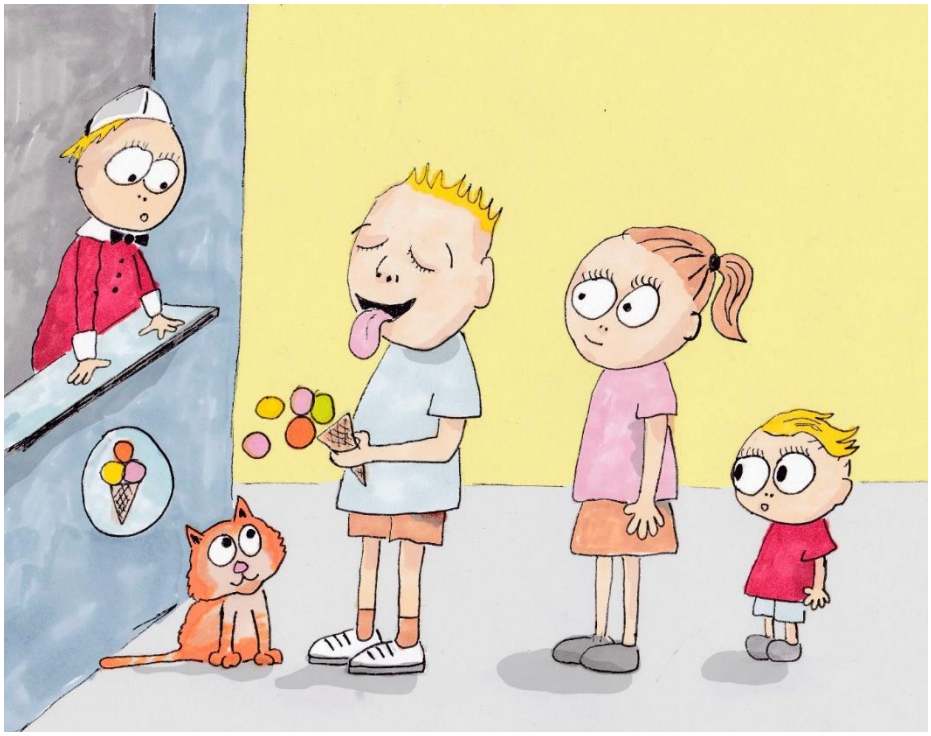
wrote the initial draft, and all authors commented on previous versions of the manuscript and/or rewrote/revised parts of the text. All authors read and approved the final manuscript. Wenke Mork Rogne secured the funding for the study.

Article 4

Eivor Finset Spilling, Vibeke Rønneberg and Mark Torrance contributed to the conceptualisation of research goals and the methodology of the study. Eivor Finset Spilling, Vibeke Rønneberg and Wenke Mork Rogne took part in the data collection and the coding of the data. Eivor Finset Spilling executed the statistical analysis, and Jens Roeser and Mark Torrance functioned as mentors for performing the statistical analysis. Eivor Finset Spilling wrote the article, and Vibeke Rønneberg, Wenke Mork Rogne and Mark Torrance commented on the manuscript. Wenke Mork Rogne secured the funding for the study.

Appendix 2 – Writing assessment tasks

Task A



gut is pus[5]

[5] Translation: *gut* ‘boy, *is* ‘ice’, *pus* ‘cat’

Task B

jente tre ball[6]

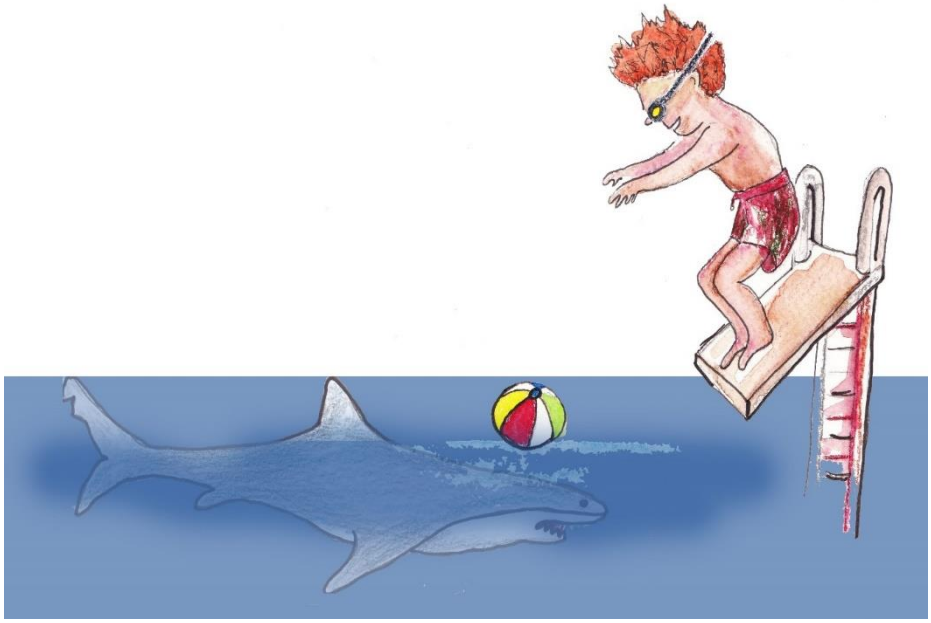
[6] Translation: *jente* ‘girl’, *tre* ‘tree’, *ball* ‘ball’

Task C

dame fugl bæsj [7]

[7] Translation: *dame* ‘lady’ *fugl* ‘bird’, *bæsj* ‘dirt’

Task D



hai ball gut[8]

[8] Translation: *hai* ‘shark’, *ball* ‘ball’, *gut* ‘boy’

Task E

mann banan ape[9]

[9] Translation: *mann* ‘man’, *banan* ‘banana’, *ape* ‘monkey’

Appendix 3 – Transcription and coding manual

Transcription and coding manual

This manual was used for transcribing and coding the texts for Article 2, 3 and 4.

The present version of the manual was used for Article 3 and 4. There were some changes in this second version compared to the first version used in Article 2. As Article 3 and 4 included more texts several new issues were raised, and new clarifications had to be made. The most important differences are: In the second version, an extra transcription was made for the vocabulary measure (because more words yielded some homonyms). Specifications regarding the coding of spelling, syntax and narrative structures were done. The first version included the rubric for holistic rating (only the texts analysed in Article 2 were rated for holistic quality).

Table of contents

OUTLINE OF THE DIFFERENT TRANSCRIPTIONS AND MEASURES……...……..……172

TRANSCRIPTION 1……..…173

TRANSCRIPTION 1.5……...176

Measure 1 and 2: Text length and spacing accuracy……..….176

TRANSCRIPTION 2……..... 177

Measure 3: Spelling accuracy……..... 178

TRANSCRIPTION 2.5……...…….……... 182

Measure 5: Vocabulary sophistication…….……182

TRANSCRIPTION 3 AND 3.5……...184

Measure 4: Terminator accuracy……..... 184

TRANSCRIPTION 4…….. 186

Measure 6: Syntactic complexity and accuracy……..… 186

TRANSCRIPTION 5…….. 190

Measure 7a: Story grammar……190

Measure 7b and c: Basic and advanced narrative features……194

References……...200

Outline of the different transcriptions and measures

Transcription 1	Transcription 1.5	Transcription 2	Transcription 2.5	Transcription 3	Transcription 3.5	Transcription 4	Transcription 5
Character by character transcription	T1 parsed into words	T1.5 is copied and the spelling is corrected	T2 is copied and homonyms are coded	T2 is copied and wrongly inserted sentence terminators are removed	T3 is copied and missing sentence terminators are inserted	T3 is copied and syntax is marked up	T4 is copied and the markup of the syntax is changed to narrative markup
	Measure 1: Text length – number of words Measure 2: Spacing accuracy	Measure 3: Spelling accuracy	Measure 5: Vocabulary sophistication		Measure 4: Terminator accuracy	Measure 6: Syntactic complexity and accuracy	Measure 7a–c (narrative structure): a) Story grammar b) Basic narrative features c) Advanced narrative features

TRANSCRIPTION 1

The transcription in general

The texts are transcribed in Excel. The schools and students are anonymised with numbers.

The texts are transcribed with goodwill. However, meaning is not interpreted at any cost – the interpretation must be reasonable. When facing unclear text, a second rater is consulted.

Transcription 1 is a character-by-character transcription. Thus, words can be misspelled, and spacing can be missing or erroneous.

All conventional characters – letters, numbers and punctuation marks – are transcribed.

All the letters are standardised to lower-case letters.

Names of the children with the function of identifying the writer, typically written on the top or the bottom of the sheet, is not transcribed. When a student has used his/her own name as part of the story / the text, this is transcribed with the following code: Each character is replaced by one q-character: Jo =qq, Maria = qqqqq.

Identifying characters: inverted letters etc.

Letters can have an unconventional form. However, they must be recognisable as a letter to be transcribed. In cases where it is hard to identify the letters two parallel strategies are used: 1. Letters in the same text are compared. Instances of the same letter and similar alternative letters are looked for. 2. The reasonable interpretation of the word/text is sought, and based on this, what the letter could be is suggested. Then it is decided if this can be justified with regard to the graphical form.

If the graphical form suggests one letter and the context another letter, the graphical form is the decisive factor.

Many of the letters are inverted, and these are transcribed as normal letters. However, the same rule applies here: In cases where an inverted letter can be mixed with another letter (<d> and <b>, <p> and <q>), the graphical form is decisive, not the letter implied by the context. For instance, *sqise is* 'eat ice', where the second letter of *spise* is inverted, is transcribed *sqise is* even though the student probably meant *spise is*.

Punctuation

What is regarded as punctuation marks is liberally decided, both with regard to form of the marks and their placement in the text. Full stop and colon can be tiny dots or bigger circles. Question marks and exclamation marks can be inverted.

Other punctuation, like comma, dash, hyphen and parentheses, and plus/minus sign, are transcribed as long as they are part of the text and have a function related to the letters, e.g. *she bought ice + banana*. A line consisting only of short horizontal lines are not transcribed as hyphen or dashes. If these kinds of signs are used systematically to separate words, they are not transcribed (but replaced with space, because the words are segmented, even though in an unconventional way).

Spacing

Spacing between letters is transcribed according to what is done in the text. If there is no space between the characters/words, there is no space character added in the transcription. If there is space – ordinary or long – one space character is transcribed. In order to transcribe space between words, there has to be bigger space between the words than between the characters in the words. The space after a word is compared to the spaces within the preceding word (the first word on a line with the following). If there are words with varying spacing between the letters, the space used as a reference point is the most used. To transcribe space within a word (wrong space), there has to be enough place to insert in a character

(with some space around), and the width of this character is decided by the context – the width of the characters used in the particular text. Extreme curves and lines are ignored.

NB: Spacing is not put into non-words. This means that non-word letter sequences that are separated by space/line brake are not transcribed with space.

Line brakes are transcribed as space (one space character). If the student masters the use of hyphenation at word-division, the word is transcribed without hyphen or space. If the student does not master this rule, e.g. writes a hyphen after every final word of the lines, space is transcribed for line break (and hyphen ignored).

Space is not put in before sentence terminators. Space is always put in after sentence terminators (full stop is accepted as segmentation of words).

When hyphen, dash etc. are used to separate words, this is transcribed as space. (The children have managed to segment the words even though they have done it in an unconventional way.)

Aspects of the text that are not transcribed

Text that is deleted is not transcribed. Characters erased or crossed out can sometimes be identified in the handwritten texts. Then this might be of help to identify the text that is not deleted. Deleted text is not possible to trace in the digital texts.

Text that is marked graphically, like bold print, bigger or smaller size of the characters etc. is not marked.

Drawings and symbols (e.g. emoticons) are not transcribed. Characters that are not letters, numbers or punctuation marks, and that do not have any function related to the other text, are not transcribed, like $ and *. Numbers are transcribed as long as they are part of the text, e.g. s*he*

bought 1 banana, but lists of numbers unrelated to the text is not transcribed, e.g. 1 2 3 00000000 3 2.

Blank lines are not marked. Broad spacing (spacing equal to the width of more than one character) is not differentiated from ordinary spacing (the width of one character).

TRANSCRIPTION 1.5

This transcription builds on transcription 1, but space use is corrected, and non-words are marked up. Measures of text length and spacing accuracy are extracted.

Measure 1 and 2: Text length and spacing accuracy

Operationalisation

Text length is defined as number of words written. A word is defined as a string which represents a phonologically plausible spelling (including the correct spelling) of a Norwegian word that children might plausibly know. Words cannot contain [.!?:] or space.

Space use accuracy is understood as ability to segment words with space (or sentence terminator).

Analytic procedure for transforming transcription 1 to transcription 1.5

1. Space insertion: Spaces are added to the text in order to create the maximum number of maximally long words. Spaces are inserted such that they bound words, but are not inserted to break up words into shorter words. Sentence terminating punctuation is not changed. Other punctuation, like hyphen, dash, equal sign and plus sign, are replaced by space.

2. Space deletion: Spaces are deleted ONLY IF they currently bound non-word strings AND ONLY IF they were not inserted in the previous step (i.e. were produced by the child) AND IF AND ONLY IF doing so creates new words. Spaces are also deleted to create maximally long words. Spaces are deleted within non-words.
3. Non-word markup: Strings that are bounded by space that are not words are marked as non-words: [NW]

The number of wrongly inserted spaces in compounds or simplexes (additional/overgeneralised spaces) are summed and noted in a separate column next to transcription 1.5.

Full stop within words: the full stop is moved and placed only after the word (if there already is a terminator after the word, the terminator is deleted). In addition, this is also noted in the column next to transcription 1.5 (regarded as an additional space like compounds that are divided).

NB: Here, white space between words and white space within words are placed on equal terms. The latter could, however, be interpreted as part of spelling (cf. spelling accuracy).

Analytic procedure for extracting the measures

Text length – functional words = count of strings bounded by spaces or sentence terminators in transcription 1.5, and not marked as non-words

Space use accuracy is modelled as number of correctly inserted spaces divided by the sum of correct, additional and missing spaces.

TRANSCRIPTION 2

This transcription is a copy of transcription 1.5, and in addition the misspellings are corrected. This transcription is used for the measure spelling accuracy.

Measure 3: Spelling accuracy

Operationaliation

The non-words are left out in this analysis, and only functional words are considered.

The norm for correctly spelled words can be found in Norwegian dictionaries, the Nynorsk dictionary and the Bokmål dictionary (https://ordbok.uib.no/). What is measured as spelling, is if all of the letters in the words are present in the right order (knowledge of phoneme-grapheme correspondence). Other conventions that can been regarded as part of the orthographic system, are isolated and treated as separate measures: punctuation and spacing (compounding errors are not included, as they are part of the measure of spacing accuracy). Capitalisation is not considered because many of the students have not learn both lower-case and higher-case letters.

Norwegian has to written standards, Nynorsk and Bokmål. The students in the sample are officially learning to write Nynorsk from the first grade (and Bokmål later). However, they are usually heavily influenced by Bokmål through society: books etc. The children are not necessarily conscious of learning Nynorsk, and there were no instructions to the teacher to mention Bokmål/Nynorsk when they were to write the texts. This means that many of the texts are written with forms that are not within the Nynorsk-norm, but the Bokmål-norm (and of course dialectal forms). Therefore, it is decided for each text if it is written in Nynorsk or Bokmål, and the words are corrected according to the corresponding written standard. Thus, a mixture of Bokmål and Nynorsk is not accepted, as the students write texts, and not single words. However, the less official rule about systematic use of equally official forms within each written standard is not applied (e.g. both *dansa*/*danset* ‘danced’ and *hjem*/*heim* ‘home’ will be accepted within the same Bokmål-text).

Both Nynorsk and Bokmål can be considered (semi-)shallow orthographies (Finish and German are more shallow, while Danish, French and English are more opaque) (Seymour, Aro & Erskine, 2003). To decide if the text is Bokmål or Nynorsk, the number of words signalling each of them are counted (a lot of words are common for both), and the standard with the most words is selected. In practice this means that the written standard that gives the least number of incorrectly spelled words is chosen. If the texts have the same number of errors in Nynorsk and Bokmål, the misspellings are considered and if these point to one of the standards, this is chosen (and if they do not, Nynorsk is chosen).

Analytic procedure for transforming transcription 1.5 to transcription 2 (rules of spelling correction)

Every functional word is corrected for misspellings according to the Nynorsk or the Bokmål dictionary, respectively. As few changes as possible are done with regard to insertion and deletion of characters to make the words correct (Norwegian has often various official forms of the same word).

To rule out **real word errors** the following rule is applied:

A word that does not make sense in context AND IS NOT based on a lemma that would make sense in context AND is a phonologically plausible or near plausible spelling of a word (or morphological form) that would make sense in context is corrected to the word/form that makes sense in context.

NB: As can be seen from the rule above, inflected forms are not corrected even though the context implies another inflected form. Examples: *han ete isen* 'he eat the ice' (the infinitive form is kept even though the context implies the present tense); *mannen ser ein apen* 'the man sees one the monkey' (the double definitiveness is kept).

In cases where a form cannot be found in the written standard, the form is corrected to the closest official form (the form that gives the least possible changes within the standard). Example: *jenta hnt is = jenta hent is* 'the girl gt ice = the girl get ice'

Loan words are accepted, e.g. *brun green vit* 'brown green white' is transcribed as *brun green kvit*. The Bokmål-word *spise* ´to eat´ is also accepted as a loan in Nynorsk-texts. (The alternative in Nynorsk is a different lexeme: *ete*, but *spise* has widespread use.) Words from other languages are corrected according to these languages, e.g. *hasta la bista = hasta la vista*.

Personal names are standardised according to the register of ssb: https://www.ssb.no/

Animal names are accepted as they are written (api, nemipus, puffi etc.)

As the spelling measure is compositional spelling and not spelling of single words (dictation), there are cases where it is hard to separate spelling from **grammar**. Generally, grammatical errors are not corrected. However, in some cases the morphology of words interferes with the spelling and it is inevitable to correct the morphology when the spelling is corrected.

In Norwegian 'the girl' usually is written *jenta*. In Bokmål, however, 'the girl' can be written *jenta* or *jenten*. I Nynorsk only *jenta* exists. A form *jenten* implies common gender, while *jenta* implies the feminine. In the transcriptions of Nynorsk-texts *jenten* will be corrected to *jenta*, because *jenten* does not exist in the official Nynorsk-norm. (Similar with mammaen/mammaa.) This can be seen as a correction of both spelling (finding the "irregular sound-letter combination"/choosing the correct form to write) and grammar (finding the correct inflectional suffix). In this study, this will be regarded a spelling error.

Another example is cases of wrong conjugation of verbs: weak forms instead of strong, and use of wrong weak suffix: *stjelte* 'stealed' instead of *stal* ' stole' , *håp**te*** 'jumped' instead of *hopp**a*.** These forms are corrected to the official ones, and regarded as spelling mistakes. The errors will not influence any other measure. In the syntax markup, however, syntactical errors are marked. (The forms *mast* and *must* are interpreted as spelling errors – forms that are erroneously conjugated as strong verbs in the past tense of the verb *miste* 'lose' and corrected to *mista* 'lost'.)

'A girl' is written *ei jente* in Nynorsk. Cases of *en jente* in Nynorsk-texts will be changed to *ein jente* – the spelling of the article is changed (*en* does not exist in Nynorsk – this form is *ein*), but the gender of the article is not corrected – the grammar will still be incorrect. This is incorrect syntax, and will be marked in the syntax markup. The same is the case with conjugated verbforms where the spelling reflects one form, but the syntax indicates another form: *han ete isen* 'he eat the ice'. The form *ete* is the infinitive, and gives a grammatically incorrect sentence. This is marked in the syntax markup. (In this clause, the present form *et*, probably was meant.)

Next to transcription 1.5 a column displays whether a text is corrected to Bokmål (b) or Nynorsk (n). If no letter/code, the text is within the norms of both written standards.

Analytic procedure for extracting the spelling measure

Transcription 1.5 is compared to transcription 2 and correctly spelled and incorrectly spelled words can be counted automatically. The non-words, marked up with [NW], are kept out. A count of correctly spelled words is extracted. Spelling accuracy is obtained by dividing number of correctly spelled words by text length (number of functional words).

TRANSCRIPTION 2.5

This transcription is a copy of transcription 2, and in addition homonyms are coded. This transcription is used for the measure vocabulary sophistication

Measure 5: Vocabulary sophistication

Operationalisation

Sophisticated vocabulary is here related to age of acquisition (Carroll & White, 1973; Gilhooly & Logie, 1980). Simple words are learned early, while more advanced words are learned later. Similarly, simple words will probably be used in written form earlier than more sophisticated words. As there are no data on written acquisition of Norwegian words, a digital survey was made, following Spilling et al. (2021):

All of the lexical words in the texts were extracted (nouns, verbs, adjectives). The words were grouped under their lemma (dictionary form) and put in a digital survey, with the instruction:

In this survey we want you to assess when children begin to use words in their written texts. Words that can be confused with others have the word class marked to clarify which word is meant: bakar 'baker' (noun). The words can be written incorrect, but they must be spelled correctly enough to understand which words is meant. Please judge each word in the list below and mark the age you think is most suitable. Some words are in Nynorsk and others are in Bokmål.

Please answer the following question: How old do you think children are when they start to use these words in the texts they write? (from 5 to 14 years).

Below is a picture of the instruction in Norwegian and the start of the list of the words to be assessed.

100%

I denne undersøkinga vil vi at du vurderer når barn startar å bruke ulike ord i dei skriftlege tekstane sine. Ord som kan forvekslast med andre ord, har markert ordklasse for å presisere kva for ord vi meiner: bakar (substantiv). Orda kan ha stavefeil, men dei må vere stava riktig nok til å forstå kva for ord som er meint. Vurder kvart ord i lista under og huk av for den alderen du meiner passar best. Nokre av orda er på nynorsk og nokre på bokmål.

Du skal svare på følgjande spørsmål: Kor gamle trur du barn er når dei startar å bruke desse orda i tekstane dei skriv?

	5 år	6 år	7 år	8 år	9 år	10 år	11 år	12 år	13 år	14 år
feire	○	○	○	○	○	○	○	○	○	○
kjøtt	○	○	○	○	○	○	○	○	○	○
øydelagt	○	○	○	○	○	○	○	○	○	○
stoppe	○	○	○	○	○	○	○	○	○	○
fiskegrateng	○	○	○	○	○	○	○	○	○	○
superhelt	○	○	○	○	○	○	○	○	○	○
sprekke	○	○	○	○	○	○	○	○	○	○

Analytic procedure

A list of all the lexical lemmas from the texts was extracted (845 lemmas). This was done by extracting all types from transcription 2, and then closed class words, proper names and words from other languages were sorted out. The lemmas consist of nouns, verbs (auxiliary verb excluded) and adjectives.

Homonyms were identified, and they were marked by changing the tokens to different corresponding lemmas in transcription 2.5.

As there were texts written both in Bokmål and in Nynorsk, some of the lemmas were in Bokmål and others in Nynorsk. Most lemmas can be found in both written standards, but often with a few differences in spelling, for example with different inflection suffixes, e.g. *bananer*

(Bokmål) and *bananar* (Nynorsk) 'bananas', *sjukehus* or *sykehus* (Bokmål) and *sjukehus* (Nynorsk) 'hosptial', *hel* (Bokmål) and *heil* (Nynorsk) 'whole'. Each lemma can therefore represent various different forms, not only with different inflections, but with slightly different letter combinations, for example the lemma *banan* representing the singular and plural form in both Bokmål and Nynorsk *banan, bananen, banarar, bananer, bananene, bananane*. When different roots are found within the same inflectional paradigm, different lemmas were used, for example the lemma *god* 'good' (adj.) that represents the forms *god*, *godt* and *gode*, and the lemma *betre* 'better' (adj.) that represents *bedre*, *betre*, *best*.

The lemmas were divided into three digital surveys distributed to 32 teachers/teacher students (N = 21, 16 and 16).

Mean score for all lemmas is extracted. This written age-of-acquisition score reflects the mean across all individual ratings for that word. Each text receives a mean score across lexical-lemma types within the text.

TRANSCRIPTION 3 AND 3.5

Transcription 3 is a copy of transcription 2. In addition, incorrect punctuation is removed. Transcription 3.5 is a copy of transcription 3 where missing punctuation is inserted.

Measure 4: Terminator accuracy

Operationalisation

Terminators encompass the sentence terminators full stop, colon, exclamation mark and question mark. Terminator accuracy is understood as ability to correctly use sentence terminators.

Confer the section of transcription 1 for details about what is transcribed as full stop, colon, question mark and exclamation mark.

Analytic procedure (rules of deletion and insertion of terminators)

In transcription 3 all sentences terminators that do not mark the end of a main clause, are removed. Confer the section about syntax for the rules of what is considered a main clause (e.g.: fragments consisting of an inflected verb are regarded clauses). A clause can start with a conjunction. Exclamation mark and question mark can be placed both at the end of a main clause and immediately after fragments, like exclamations.

Examples:

Transcription 2: jenta sparkar ballen i treet. klatrar opp. og jenta hentar. ballen 'the girl kicks the ball in the tree. climbs up. and the girl fetches. the ball'
Transcription 3: jenta sparkar ballen i treet. klatrar opp. og jenta hentar ballen 'the girl kicks the ball in the tree. climbs up. and the girl fetches the ball'

In transcription 3.5 missing sentence terminators are inserted. Every main clause should have a sentence terminator, unless it is coordinated with another main clause. Only two main clauses (with subordinate clauses) can be coordinated before a sentence terminator is set in (thus, never ending sentences are split up). All texts are terminated by a full stop. In clauses with verbs like say, scream, whisper etc. colon is inserted if this verb precedes the remark, e.g. *mamma sa: ambulansen kjem* 'mama said: the ambulance is coming'.

Transcription 3.5: jenta sparkar ballen i treet. klatrar opp. og jenta hentar ballen. 'the girl kicks the ball in the tree. climbs up. and the girl fetches the ball.'

Analytic procedure for extracting the punctuation measures

Number of correctly inserted terminators (correct terminators): Count of terminators [.?!:] in transcription 3.

Number of additional terminators: Count of terminators [.?!:] in transcription 2 minus correct terminators

Number of missing terminators: Count of required terminators (count of terminators [.?!:] in transcription 3.5) minus correct terminators

Terminator accuracy is modelled as count of correct terminators divided by the sum of correct, additional and missing terminators.

TRANSCRIPTION 4

This transcription builds on transcription 3. In addition, the text is marked up for syntax.

Measure 6: Syntactic complexity and accuracy

Operationalisation

The syntax measure reflects the student's ability to construct different kinds of clauses (main and subordinate) and whether the clauses are free of syntactical errors or not.

The syntax coding is done independently of student´s punctuation.

A **main clause** must have a subject and a verb (a finite verb). When two main clauses are coordinated with linking words, they are counted as two separate main clauses. In cases of coordination and omission of one of the parts (subject/verb), the clause will be counted as a main clause as long as the deletion follows regular writing rules. Imperative clauses are regarded as main clauses even though the subject is not expressed.

Clauses that wrongly lack subject are regarded as clauses, but are marked as ungrammatical.

Coordination with ellipses and pseudo-coordinations can be hard to classify as one or more clauses. Here, coordinated sentences are regarded as separate clauses as long as the verbs denote different processes:

Ho måtte dra heim og dusje og vaske håret. 'She had to go home and shower and wash her hair.' (three processes = three clauses)

Jenta er ute og går. 'The girl is out walking.' – literally 'is out and walks' (to processes = two clauses)

Han held på å ta ballen. 'He is about to take the ball.' (one process = one clause)

Han måtte forte seg å kjøpe ny is. 'He had to hurry to buy a new ice.' (one process = one clause)

Iterative cases are regarded as one process: *Ho leita og leita og leita og leita etter bamsen*. 'She looked and looked and looked and looked for the teddy.'

Cases of repetition within what can be regarded the same clause/sentence, are regarded as one, while completely identical clauses that are repeated, are regarded as two:

Han mista mista isen. 'He lost lost the ice.' = one clause

Det var ein gong ein gut. Det var ein gong ein gut. 'Once upon a time there was a boy. Once upon a time there was a boy.' = two clauses

A **subordinate clause** is part of another clause (nominal and adverbial clauses) or part of another part (adjectival clauses). A subordinate clause has a subject and a verb (in relative clauses the subject can be found in the subordinator *som* 'that').

Single words, lists of single words and phrases (e.g. *jente i tre* 'girl in tree') are regarded as **fragments** and are not coded.

If the clauses contain any **syntactical error**, they are marked as ungrammatical (e.g. wrong word order, non-finite verb as the only verb, a part of sentence other than the verb missing etc.). All kinds of syntactical errors, concerning level of phrases and clauses, are included. Only syntactical rules are concerned. Violations of general writing rules/recommendations are not marked as wrong expressions, like mixing *da* 'then' and *når* 'when' (they belong to the same part of speech, thus mixing them will not have any syntactical consequences). Syntactical constructions that are more usual in oral language are accepted as correct (e.g. extra position: *pusen den ville ha den isen* 'the cat he wanted to have that ice', *oppe i himmelen der er det blått* 'in the sky there it is blue'). Errors of *og*/*å* (the conjunction 'and' / the infinitive marker 'to', pronounced the same way) are not changed in the spelling, but affect the grammaticality of the clauses.

Morphological errors without syntactical consequences do not affect the coding of ungrammaticality. (In the clause: *Jenten er glad. Ho ler.* 'The girl is happy. She laughs', the genus of *jenten* is wrong according to the Nynorsk normal. This is does not have syntactical implications, so the clause is not marked as ungrammatical. However, *jenten* is written *jenta* in Nynorsk, and this is regarded a spelling error – confer the section about spelling.)

Analytic procedure

Every main clause is marked up with [M].

Every subordinate clause is marked up with {S}.

Clauses that are ungrammatical are marked up with: [MW], {SW}

Subordinate clauses that are not part of a main clause, e.g. function as a complement in a phrase, are marked up with {SQ}.

Every clause that is correctly terminated is also given a tag, P: [MP] / [MWP]. The terminator has to be placed immediately after the clause.

Fragments before, between and after clauses are not coded (and not part of the [] {}).

Linking words are regarded as part of the clauses.

Each text receives a syntax score according to these rules:

- 1 point for every syntactically correct main clause
- 0.5 point for every main clause with one or more syntactical errors
- 2 points for every syntactically correct subordinate clause
- 1.5 points for every subordinate clause with one or more syntactical errors

Examples

Norwegian	**English**	**Comment**
[MW mista isen] [M katten blir glad] [M den lille guten blir og glad] snipp snapp	[MW lost the ice] [M the cat is happy] [M the little boy is also happy] the end	Two main clauses, the incomplete clause *mista isen* 'lost the ice' lacks subject and is marked as ungrammatical
[M så {S da kulene datt ned frå isen} smilte katten]	[M so {S when the balls fell down from the ice} the cat smiled]	One main clause and one subordinate clause
[M dei står i ein kø] [M og ventar på ein is]	[M they are standing in a line] [M and waiting for an ice]	Two main clauses, the last lacking subject because the clauses are coordinated, but still grammatical because it follows

		Norwegian syntax rules
[M først så tok guten ein is] [M og så fekk lillebroren store auge] [MW og så tok katten]	[M first the boy took an ice] [M and then the little brother had big eyes] [MW and then the cat took]	Three main clauses, two correct and one ungrammatical
jente tre ball	jente tre ball	No sentence structure with a verb, and therefore no mark up
[MP ein gut står i mål.] eit tre [MWP ei jente er oppi tre.]	[MP a boy is the goalkeeper.] a tree [MWP a girl is up in tree.]	Fragments between clauses are not coded. Correct punctuation coded.
ein jente {SQW som leika med ei fotball] [M ho sparka fotballen oppi eit tre]	a girl {SQW that played with a (wrong concord) ball}] [M she kicked the ball in a tree]	The subordinate clause is describing a noun-phrase, not a clause

TRANSCRIPTION 5

This transcription is a copy of transcription 4. All syntax markers are removed/changed to narrative markers. The majority of the markers are concurrent: most of the syntactical units correspond to a narrative unit on local level. Fragments before, between and after clauses are not marked up syntactically, but are given a narrative marker. In addition, narrative markers on global level are added.

Three different measures of narrative structure are extracted: one on global level (story grammar) and two on more local level (basic and advanced narrative structures).

Measure 7a: Story grammar

Operationalisation

The measure story grammar builds on theory from Labov and Waletsky (1967) and Senje and Skjong (2005). Labov and Waletsky (1967) emphasise the complication and resolution in personal narratives. These correspond to the tasks given to the children where they are instructed to write about a problem displayed in a picture. Senje and Skjong (2005) is a text book for teacher students presenting didactics for analysis of student's stories in a Norwegian context. A point here is that students normally learn to write stories with the traditional structure with introduction, main part and conclusion.

A narrative is understood as a sequence of at least two events linked in time. The first event must have the potential to be something disruptive – to depict a shift in a normal state, and thus make the reader expect a resolution, something more happening telling what happens next, as a consequence of the first event. The first event must have the potential to make the story progress/add development to the story.

Thus, the first example is not regarded a narrative, but the second is:

1. The boy bought an ice cream. He went home.
2. The boy lost an ice cream. He got a new one.

Both examples have to events linked in time. However, only the second one has an element of something disruptive. Then, after this disruptive event, the situation is back to balance through a resolving event.

A narrative could also be the linking of a state and an event:

3. The boy was sad. Then he got an ice cream, and turned happy.

Usually, the texts will also have an introduction/orientation, especially since these are written texts. Just as there is a change/contrast between the complication and the resolution there will be a change/contrast/disruption between the orientation and the complication.

Analytic procedure

The typical structure of a narrative with the three stages orientation, complication and resolution is coded. There must be at least two stages to code a stage (orientation + complication / complication + resolution / orientation + complication + resolution). Each stage can only be coded once. If stages are coded, all of the text is put in stages.

Texts with no stages are coded zero, while texts with two stages are coded one, and texts with all three stages are coded two.

The stages are coded with the following tags:

[O] = Narrative orientation

- The orientation gives information about participants, places, activities and/or time.
- For something to be an orientation at least one element from the orientation must follow in the complication (element is understood in a wide sense: metonymic relationships are accepted: children – boy, ice cream line – ice cream).
- The orientation helps the reader to understand the rest of the text/gives background information to understand the complication of the story.
- The orientation lasts until the complication is introduced.
- A headline can function as the orientation (headline is coded within the orientation, as a phase)
- The orientation must be positioned before the complication.
- The orientation cannot be coded as the only stage.

[C] = Narrative complication

- The complication is a disruption (shift) in an expected course of events (the situation depicted in the orientation).

- If there is no orientation, the complication is something disruptive compared to a normal situation.
- The complication makes the reader expect a resolution that answers what finally happened to the relevant participants that experienced the disruption.
- The complication starts with the disruption and ends where the resolution starts.
- The complication-stage cannot be coded as the only stage.

[R]= Narrative resolution

- The resolution answers the question what finally happened as a consequence of the complication.
- The resolution must make the reader understand what happened to at least one of the main participants (persons, animals or objects) that experienced the complication.
- The coding of the resolution-stage starts with the answer to the question what finally happened as a consequence of the complication.
- The complication and the resolution must be logically ordered – the resolution can only come after the complication.
- The resolution-stage cannot be coded as the only stage.

[G] = Greeting to Elling (the teddy bear receiving the texts) or greeting from the student. This is on a level outside the story, like a frame for the story. The other stages can only be coded once for the stories, while G can be coded twice: both before and after the story.

Examples are provided together with the other narrative measures, see the next section.

Measure 7b and 7c: Basic and advanced narrative features

Operationalisation

Narrative structures on more local level are also coded. As a story usually is a chain of events linked in time (Labov & Waletsky, 1967), the basic story structure is the event. The measure of basic narrative features is number of events in each text.

Martin and Rose (2008) present a way of analysing narratives through phases, that is, discourse units, like setting, description, action, effect, reaction, problem and solution. These phases are used to move sequences forward and engage the readers. Some of these phases are more advanced in the sense that they relate to other phases/parts of the text: comment from narrator, problem, solution, effect and reaction. The number of these phases are used as a measure of advanced narrative features.

Analytic procedure

Each clause is coded as a phase (this follows the syntax markup). Fragments positioned before, between and after clauses are also coded as one phase. Non-words are not coded as phases, but they can be part of phases if they are positioned in the middle of a clause.

The phases are marked up with the following tags (more or less the categories of Martin & Rose, 2008, p. 82):

> {D} = Narrative phase: Description (provides content, gives the reader better understanding of the story, does not drive the plot/sequence of events forward, the verb used denotes a state)
>
> {E} = Narrative phase: Event (something that happens at a point in time, it has the potential to add to the sequence of events (a temporal sequence), the verb used denotes a process or an action like *gutten ble lei seg* 'the boy turned sad')

{N} = Narrative phase: Comment from the narrator concerning the story (anything that is not a description or event, the narrator intrudes the sequence of activities to comment on events or participants, give explanations, evaluations, extra emphasis on an event etc.)

{P} = Narrative phase: Problem (an undesired event or description/state)

{S} = Narrative phase: Solution (corresponds to a problem/the consequences of a problem, and restores order)

{F} = Narrative phase: Effect (a consequence of a preceding event or description: a material outcome, e.g. an event)

{A} = Narrative phase: Reaction (attitude, feeling or thought of a participant, a consequence of a preceding event/description)

{I} = Narrative formula (starting or ending formula, like *snipp snapp snute* 'and that is the end of the story', *slutt* 'the end'. A formula will always be a description.)

In cases where a subordinate clause is a complement in a phrase and not a clause, the phrase and the subordinate clause will be regarded one phase:

syntax markup: *a girl {SP who played with a football}*
narrative markup: *{D a girl who played with a football}*

In cases where a subordinate clause splits a main clause, there will be two phases, and the tag for the main clause that is cut in two, is only put in at the start: {**E and he** {D who was guarding the goal} **had to tell the parents**}

In some fragments, the content points to different phases, and these are separated:

mann banan ape lur = [D mann banan] [DF ape lur] '[D man banana] [DF monkey sly]'

A phase will have all the tags possible according to this hierarchy:

D	E	N
H I F A P S	H F A P S	
S S	S S	

At the first level, all phases must be either D, E or N. In addition, a phase can have a tag from the next level (and again, phases on the same level are mutually exclusive). If a phase is E, it can also be P, F, A, S or H. F and A can on the third level be S. NB: S can be F or A (in addition to E) or only E.

It can be several P-s in one text. When coding P, it is coded if the phase has P as its main function. Thus, the text: *she fell from the tree she got hurt* is coded as: {EP she fell from the tree} {EA she got hurt}. The last phase could also be looked upon as a problem, but here, the main function of the latter phase is to depict the consequence of the problem, and A is coded. However, in the following text there will be two problems: {EP the ball was kicked in the tree} {EF the girl climbed the tree} {EP the branch broke}. The branch breaking is not seen as an effect of the preceding problem, but as another problem. In some cases a phase can denote a clear consequence/effect, but at the same time introduce a new problem that is developed further, and then both F and P are coded: C [EP putseleg kom ein hai] R [EA guten blei han blei så redd] {EFP at han mista ballen} 'C [EP suddenly a shark came] R [EA the boy turned so afraid] {EFP that he lost the ball}.

In some cases, it is hard to decide where a phase (problem/solution) starts and ends. The rule is to only denote a phase at the actual problem/solution etc, not the introduction to it: C [EP fugl bæsj på dame] R [EF hjem til hus] [EFS vaska håret sitt hår] 'C [EP bird shit on lady] R {EF home to house} {EFS wash her hair}. In this case, going home is necessary to wash the hair, but the actual solution is the clause denoting the washing action.

P usually will be E, but there are also cases where P is D, like: {DP ballen satt fast} 'the ball was jammed'.

The phases are coded within the overall structure orientation, complication, resolution.

Different perspectives in the texts: Some texts are hard to code because it is possible to choose different perspectives sympathising with different characters, e.g. the man or the monkey, the boy or the cat. The rule is to be open for either perspective, but default will be the man/the boy, as we tend to identify with characters that are similar to ourselves. However, if the text is written in a way that emphasises the monkey/the cat, e.g. this character dominates the story, it might be the best solution to code after the animal´s perspective. Another consideration can be to choose the most advanced coding, e.g. as many different stages and phases as possible.

Examples of coding of stages and phases

Norwegian	**English**	**Comments**
[O {D ein gut står i iskiosken} {D ein gut jente gut står i kø} {E så får den eine får isen med 6 kule}] [C {N men akkurat da så	[O {E a boy stands in the ice kiosk} {E a boy girl boy stand in line} {E then one of them gets the ice with 6 ball}] [C {N but precisely then so	

skjedde noko} {EP så da kulene datt ned frå isen}] [R {EA smilte katten}]	something happened} {EP so when the balls fell down from the ice}] [R {EA the cat smiled}]	
[O {D det er haust} {E og ein jente og ein gut speler fotball}] [C {D og dei hadde det kjekt} {EP heilt til fotballen blei sparka opp i treet} {EP og greina knakk}.]	[O {D it is autumn} {E and a girl and a boy play soccer}] [C {D and they had a good time} {EP until the ball was kicked up in the tree} {EP and the branch broke}.]	The Es drive the main sequence of events forward. The Ds do not, on the contrary, they provide more information about the context
{D eg ser iskule.} {D eg ser} {E iskulene fall}	{D I see ice ball.} {D I see} {E the ice balls fell}	D, not N
[G hei!] {E ein gut et ein is} [G klem frå §§§§§§§§]	[G hi!] {E a boy eats an ice} [G hug from §§§§§§§§]	
[O {E guten et is}.] [C {EP guten mistar isen på golvet} {E **dama** {D som lagar isen} **ho såg det}**]	[O {E the boy eats ice}.] [C {EP the boy drops the ice on the floor} {E **the lady** {D who makes the ice} **she saw it}**]	A tag for a stage and a phase is only put once, even if the stage/phase is split into two – cf. the bold text.
[C {EP guten mista isen}] [R {EFS mora hans kjøpte ein ny til han}]	[C {EP the boy dropped his ice}] [R {EFS his mother bought him a new one}]	Solution is F
[C {EP guten mista isen sin}] [R {ES onkelen hans fra Amerika kom hjem}	[C {EP the boy dropped his ice}] [R {ES his uncle from America came home} {EF the boy forgot	S is not F or A

{EF gutten glemte isen} {DA og var glad}]	about the ice} {DA and was happy}]	
{E mamma seier} {E ambulansen kjem}	{E mummy says} {E the ambulance is coming}	Subordinate clause with verbs like *say, think, mean* etc.
{D han synest} {D isen ser god ut}	{D he thinks} {D the ice looks good}	Subordinate clause with verbs like *say, think, mean* etc.
[D jenta klatra ikkje i treet]	[D the girl did not climb the tree]	When the main verb is negated, the phase is coded D
[E jenta skulle klatre i treet]	[E the girl will climb the tree]	When the main verb is moderated by an auxiliary verb like *skulle* 'should, would' etc. the phase is coded E even though we do not know for sure that the event happened
[C {E hun står på en grein} {DP som knekker} {DP hun kan dette ned} {DP hun kan brekke fot handa}] [R {EFS den lille gutten henter pappa} {EFS som hjelper henne ned}]	[C {E she stands on a branch} {DP that is about to break} {DP she can fall down} {DP she can break leg hand} [R {EFS the little boy fetches dad} {EFS that helps her down}]	If the auxiliary verb *kunne* 'can' is used and several hypothetical alternatives are expressed, D is coded

References

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*(1), 85–95. https://doi.org/10.1080/14640747308400325

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427.https://doi.org/10.3758/BF03201693

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts. Proceedings of the 1966 annual spring meeting of the American ethnological society* (pp. 12–44). University of Washington Press.

Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture*. Equinox.

Senje, T., & Skjong, S. (2005). *Elevfortellinger: Analyse og vurdering*. Cappelen akademisk.

Seymour, P. H., Aro, M., Erskine, J. M., & Collaboration with COST Action A8 Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology*, *94*(2), 143–174.

Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., & Torrance, M. (2022). Handwriting versus keyboarding: Does writing modality affect quality of narratives written by beginning writers?. Reading and Writing, *35*(1), 129–153.

Appendix 4 – Approval from the Norwegian Centre for Research Data

NSD

Wenke Mork Rogne
Postboks 500
6101 VOLDA

Vår dato: 02.05.2018 Vår ref: 59799 / 3 / LAR Deres dato: Deres ref:

Tilrådning fra NSD Personvernombudet for forskning § 7-27

Personvernombudet for forskning viser til meldeskjema mottatt 13.03.2018 for prosjektet:

59799	*DigiHand*
Behandlingsansvarlig	*Høgskulen i Volda, ved institusjonens øverste leder*
Daglig ansvarlig	*Wenke Mork Rogne*

Vurdering

Etter gjennomgang av opplysningene i meldeskjemaet og øvrig dokumentasjon finner vi at prosjektet er unntatt konsesjonsplikt og at personopplysningene som blir samlet inn i dette prosjektet er regulert av § 7-27 i personopplysningsforskriften. På den neste siden er vår vurdering av prosjektopplegget slik det er meldt til oss. Du kan nå gå i gang med å behandle personopplysninger.

Vilkår for vår anbefaling

Vår anbefaling forutsetter at du gjennomfører prosjektet i tråd med:
- opplysningene gitt i meldeskjemaet og øvrig dokumentasjon
- vår prosjektvurdering, se side 2
- eventuell korrespondanse med oss

Meld fra hvis du gjør vesentlige endringer i prosjektet

Dersom prosjektet endrer seg, kan det være nødvendig å sende inn endringsmelding. På våre nettsider finner du svar på hvilke endringer du må melde, samt endringsskjema.

Opplysninger om prosjektet blir lagt ut på våre nettsider og i Meldingsarkivet

Vi har lagt ut opplysninger om prosjektet på nettsidene våre. Alle våre institusjoner har også tilgang til egne prosjekter i Meldingsarkivet.

Vi tar kontakt om status for behandling av personopplysninger ved prosjektslutt

Ved prosjektslutt 31.12.2021 vil vi ta kontakt for å avklare status for behandlingen av personopplysninger.

Se våre nettsider eller ta kontakt dersom du har spørsmål. Vi ønsker lykke til med prosjektet!

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

NSD – Norsk senter for forskningsdata AS Harald Hårfagres gate 29 Tel: +47-55 58 21 17 nsd@nsd.no Org.nr. 985 321 884
NSD – Norwegian Centre for Research Data NO-5007 Bergen, NORWAY Faks: +47-55 58 96 50 www.nsd.no

Vennlig hilsen

Marianne Høgetveit Myhren

Lasse André Raa

Kontaktperson: Lasse André Raa tlf: 55 58 20 59 / Lasse.Raa@nsd.no
Vedlegg: Prosjektvurdering