



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

Eivor Finset Spilling^{1,2} · Vibeke Rønneberg² · Wenke Mork Rogne¹ ·
Jens Roeser³ · Mark Torrance^{2,3}

Accepted: 27 May 2021 / Published online: 17 June 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

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 not 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 keyboard¹ 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 automated 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 (/<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 our 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 . 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 ($\Delta \text{elpd} = -1$, $SE=7$). 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 ($\widehat{\Delta elpd} = -26$, $SE=11$). 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 ($\widehat{\Delta elpd} = -15$, $SE=21$ 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)^aIndicates moderate or strong evidence in support of no moderating effect ($\text{BF}_0 > 5$). 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 isen 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	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	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 kulene datt ned frå isen smilte katten.	[M ein gut står i iskiosken] [MW ein gut jente står i kø] [MW så får den eine får isen med 6 kule] [C {N men akkurat da så skjedde noko} {EP så da kulene datt ned frå isen}] [R {EA smilte katten}]	[O {E ein gut står i iskiosken} {E ein gut jente gut står i kø} {E så får den eine far isen med 6 kule}] [M men akkurat da så skjedde noko] [M så {S da kulene datt ned frå isen} 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., & Strömbärgsson, 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.
- Beschchorner, 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.comedu.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 and visual arts* (pp. 12–44). American Ethnological Society.
- Lee, M. D., & Wagenaars, 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., Lundstræ, 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., & Lundetrae, 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](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.