EXPLORING TRANSCRIPTION PROCESSES WHEN CHILDREN WITH AND WITHOUT READING AND WRITING DIFFICULTIES PRODUCE WRITTEN TEXTS USING SPEECH-TO-TEXT

SANNA KRAFT[1], VIBEKE RØNNEBERG[2], JOHN RACK[3], FEDRIK THURFJELL[4] & ÅSA WENGELIN[1]

1. University of Gothenburg
2. University of Stavanger
3. Linnaeus University
4. Habiliteringens resurscenter, Stockholm

Abstract
Fluent transcription is hard to establish for children with reading and writing difficulties, due to problems with spelling. It has been proposed that composing by speech-to-text (STT) could facilitate their transcription, by circumventing the spelling process. To investigate this, transcription and error correction processes, and their relation to production rate (text length/time on task) was investigated in a sample of Swedish 10–13 year olds with and without reading and writing difficulties using STT to write expository texts. We determined effects of individual abilities: working memory, spelling, decoding, and the ability to interact with the STT tool under optimal conditions (STT success rate) on burst length, burst accuracy and production rate. Production rate was predicted by working memory capacity, by how long bursts the children produced, and by how accurate those bursts were. Further, burst accuracy was only predicted by a child's STT success rate (in a test), but none of the other individual abilities. Dictating more than one word at a time and combining STT and keyboard use were identified as two useful strategies that can be taught in STT instruction. The results indicate that composing text using STT is a cognitively complex process placing heavy demands on working memory, and that STT success rate (that is, the combined effect of the technical capability of the STT tool and the participants output) is crucial to gain a fluent transcription without unnecessary disruptions.

Keywords: reading and writing difficulties, writing processes, speech recognition, speech-to-text, children

Kraft, S., Rønneberg , V., Rack, J. Thurfjell, F., & Wengelin, Å (2023). Exploring transcription processes when children with and without reading and writing difficulties produce written text using speech-to-text. L1-Educational Studies in Language and Literature, 22, 1-28. https://doi.org/10.21248/l1esll.2023.23.1.427
Corresponding author: Sanna Kraft, Department of Swedish, Multilingualism, Language Technology, University of Gothenburg, Box 200, 40530 Göteborg, email: sanna.kraft@svenska.gu.se

1. INTRODUCTION

Being able to express oneself in written text is fundamental for academic success, not least because writing underpins the main method of assessment throughout education: students are expected to show their knowledge through writing. However, many children struggle with writing, and for those with reading and writing difficulties, spelling is a particular challenge and is reported as the most prominent difficulty in both primary and secondary education (Mortimore and Crozier, 2006). Further, difficulties with spelling often persist (Sumner and Connelly, 2020), even when reading skills have improved (Berninger, 2006). The effects of spelling difficulties on writing go beyond a high frequency of spelling errors; students who struggle with spelling also write less fluently (Torrance et al, 2016; Wengelin, 2007). If spelling processes are non-automatised, they are likely to demand more attention, causing disruption in the writing process. If, for example, writers are forced to consider spelling in the middle of a word they may even forget what they were planning to write next. These disruptions make text production more of a struggle and less fluent (Sumner et al., 2013; Torrance et al., 2016; Wengelin, 2007). This lack of fluency can hinder students' formulation processes and influence the final text, resulting in, for example, shorter texts (Beers et al., 2017) and/or generally lower text quality (Connelly et al., 2005). Such struggling and poor writing performance may in turn lead to low motivation (Camacho et al., 2021) and self-efficacy.

Recently, digital tools that offer alternative transcription methods to facilitate writing, such as speech-to-text (STT), have become more widely available. STT has the potential to support students who struggle with spelling since it offers the possibility to dictate words instead of having to spell them (MacArthur, 2009). However, we still have limited knowledge about the effects of STT on children's writing processes. Dictating (or "composing") with STT may involve other challenges, causing other types of dysfluencies. First, since composing with STT is carried out in the spoken modality, the underlying processing demands are somewhat different from those underpinning both handwriting and typing. Second, STT systems are not perfect. They may "mishear" words, meaning that the STT system transcribes a different word to the one the writer pronounced or attempted to pronounce. This results in a need for the writer to reread, check and revise the text. In order to understand whether and how STT can facilitate writing for children who struggle with spelling we need to gain knowledge about how these children handle and eventually master the dictation process and to understand what factors influence the STT writing process. STT may be of limited use if one type of word-level disruption is just replaced by another. To our knowledge, no previous study has investigated the transcription process of children dictating (but see Leijten, 2007, regarding STT writing processes in adults).The purpose of our study is therefore to address this knowledge gap by examining how children interact with an STT system, and to investigate whether

their strategies and possible dysfluencies during the process influence writing fluency overall. We further examine whether their independent abilities of spelling, decoding, and working memory function affect their behaviour and fluency.

1.1 The importance of automatised and fluent transcription skills

Several researchers have highlighted the importance of automatised transcription in writing (see, e.g., Hayes, 2012; McCutchen, 1996; Graham & Santangelo, 2014). In the theoretical model referred to as the *simple view of writing* (Berninger et al., 2002), transcription skills and executive functions together form the foundation that enables the development of text generation. During writing the various processes involved interact in an environment where working memory is limited. If lower-level processes are not automatised, they will make greater demands of working memory, leaving less available for other processes, and therefore undermining children's writing performance. In other words, if a writer has difficulties with transcription due to spelling deficits, he or she may not be able to devote resources to processes such as lexical decision-making, text generation or revision in an efficient way, leading to poorer texts (McCutchen, 1996).

In children who manifest reading and writing difficulties, the transcription process during composing is normally not automatised to the same extent as in their peers when writing by hand (Sumner, 2013) or by keyboard (Wengelin, 2007). This is in part because their spelling difficulties force them to have a local focus on spelling during composition (Sumner et al., 2013), which creates disruptions that are likely to affect the fluency of the transcription process. Furthermore, they must deal not only with actual spelling errors, but also with factors associated with them, such as worries about spelling or the overall motivation to write. Because of this, these children often continue to struggle with transcription throughout their schooling (Sumner et al., 2013; Wengelin, 2007).

Because composing with STT allows children to write without having to focus on spelling, this writing mode might enable this group to avoid disruptions due to their spelling difficulties, possibly freeing up more cognitive capacity to, for example, compose longer texts in the same amount of time (i.e., increase production rate). However, as mentioned in the introduction, there is a risk that the STT condition creates other challenges instead, and at this time little is known about how transcription processes are affected when composing with STT. When composing on a keyboard, Grabowski (2008) highlighted that the transcription process is dependent on several factors; including not only finding the keys and pressing them at short intervals, but also strategies used to handle transcription disruptions. For example, Grabowski mentions that some typists may be fast at transcribing, but slow at navigating and correcting errors, some can be slow typists but do not need to correct errors, and some can be fast typists who make many typographic errors that need to be corrected. For this reason, it is of importance to consider strategies the writer uses to solve problems related to disruptions during transcription when investigating

fluency. When composing by STT, such disruptions include those of detecting and correcting any errors produced by the tool (which involves reading to detect and writing to correct). For this reason, when investigating whether, and if so how, STT can facilitate the lower-level transcription process, it is very important to study not only STT transcription as such (in this study: burst length and accuracy) but also the strategies used by writers when the tool produces transcription errors (i.e., semantic errors).

Previous research indicates that the impact of transcription decreases as it develops and is automatised, such that disruptions arising from spelling a word, forming the letters by hand, or finding the keys on the keyboard, no longer limits text composition. In a model of direct and indirect effects in writing, Kim and Park (2019) showed that the transcription process (handwriting) had a direct effect on writing quality in first-graders but not in third-graders. This probably reflects that some aspects of transcription in beginning writers use up resources that otherwise could be available for higher-level processes. Similar findings were reported by Connelly et al. (2007), where transcription speed in handwriting accounted for a higher percentage of text-quality variance in younger children, indicating that gaining fluency in transcription is important to enable higher-level processes such as idea generation. Further, in a recent study, Rønneberg et al. (2022) found that in typically developing sixth-graders, with a mean age of 11:10 years, the ability to spell and type fluently (in a spelling test and in a timed keyboarding task) affected composition fluency but not text quality. The authors conclude that their result is inconsistent with previous claims that weakness in low-level transcription skills will hinder higher-level processing. However, they point out that one possible explanation for their results could be that the children in their study had already developed sufficient transcription skills. For children who struggle with transcription—both beginning writers and children with spelling difficulties—it is therefore of great importance to develop more efficient transcription skills.

As mentioned above, theories/models of writing and previous research have highlighted the importance of automatised transcription. Since the main reason for composing with STT for children with reading and writing difficulties is to avoid a dysfluent transcription, by circumventing the demands of spelling processes, and thereby enabling more cognitive capacity for other higher-level processes, a first step should be to investigate what factors affect fluency during composing with STT.

1.2 How STT can facilitate writing

Previous research into STT as a facilitating writing tool has yielded diverse—but overall positive—results regarding increased text length and/or text quality in people with various types of writing difficulties. It should be pointed out that the majority of the existing studies on STT were conducted almost two decades ago. Since that time, accuracy of speech-recognition technologies have improved considerably (Lu et al., 2020).

MacArthur and Cavalier (2004) investigated writing in secondary-school students with and without learning difficulties in three different conditions: writing by hand, dictation to a scribe (meaning that detecting and correcting errors was left out in this condition) and dictation with STT. They found that dictation, especially to a scribe, improved text quality in the group with learning difficulties, but not in the group of typically developing students. Neither measures of vocabulary or text length differed across conditions. Higgins and Raskind (1995) investigated writing in adults with difficulties in three different conditions: without support, dictation to a scribe and dictation with STT. Dictated texts were found to have a higher proportion of long words than texts written without support, and the dictating participants pointed out that they did not have to substitute hard-to-spell words with others when dictating. As regards younger students, Quinlan (2004) found that children between 11 and 14 years with difficulties produced longer texts when dictating than when handwriting. However, no differences in text quality were observed. A more recent study (Haug & Klein, 2018) showed that strategy instruction in two different modalities—handwriting and STT—influenced text quality, text length and argumentation in both modalities for children aged 10–11 years without difficulties. This indicates that STT was as good as handwriting for the children in this group, even though they clearly had much less experience of writing with STT.

Despite a number of positive effects, one risk highlighted by Sumner et al. (2013) is that texts produced in the spoken modality may become more "spoken" in character. Kraft et al. (2019) investigated differences in lexical aspects across expository accounts produced by children with spelling difficulties in three conditions: keyboard writing, STT and spoken presentation. Their results showed that texts written on a keyboard and those written using STT were similar in terms of lexical diversity, lexical density, text length and the proportion of long words, but that both of these types of texts differed from the spoken accounts in terms of lexical density and the proportion of long words. The authors concluded that children aged 10–12 years can apply written-language conventions while dictating.

Most studies investigating STT as a tool for composition—Kraft et al. (2019) being the exception—investigated texts produced using English-language speech-recognition systems (see also Svensson et al. 2019, on the effects of assistive technology intervention on different reading-related measures in Swedish, where STT was one available tool for the participants). There is thus clearly a need for more research to fully understand the effects of STT on text production across languages. Indeed, given the increased availability, sophistication and ease-of-use of speech-recognition technology, its potential usefulness when it comes to enhancing transcription fluency during composition in children (and adults, for that matter) with reading and writing difficulties is greater than ever. This calls for efforts to gain a better understanding of text production by means of STT in general, and greater insight into the STT transcription process in particular.

1.3 The transcription process in text production using STT

It has been suggested that composing with STT can support children who struggle with writing (MacArthur, 2009). Because STT removes the burden of spelling it has the potential to enable a more fluent transcription which could free cognitive capacity for higher level processes. One characteristic feature of composing with STT, just as with other input tools, is that the writer produces linguistic content in intermittent word bursts which are interspersed with pauses (Kaufer et al., 1986). Those pauses enable the user to do various things. One such thing, on a higher level, could be planning linguistic content for the next burst. In the case of STT, however, the tool sometimes misinterprets the writer, or the writer sometimes does not express himself/herself clearly, and so the tool produces semantic errors that require detection and correction (Leijten, 2007). If the writer also exploits pauses for dealing with, or checking for, such errors, this could create unnecessary disruptions, which may place an additional load on working memory in terms of self-regulation and revision.

In this context, it should be noted that certain burst lengths may be generally preferable to others in terms of accuracy and error correction. In other words, the STT tool may be more likely to reproduce bursts of a certain length correctly, and certain burst lengths may also make it easier for the user to correct any errors made by the tool. If, for example, a user makes a nine-word burst and the STT tool misinterprets one of first few words, the cognitive demands stemming from error correction will probably be larger than if the burst to be corrected consists of only three words. When composing by hand, burst length increases with handwriting fluency (Alves et al., 2011) and translation fluency (see Hayes, 2009), but how burst length relates to fluency when composing with STT is as yet unexplored. Further, eye-movement studies of children with reading and writing difficulties have shown that they rarely re-read many words in the text-written-so-far (Johansson et al., 2008), tending instead to be “local” in their writing process. Because of this, there is a risk that errors far from the leading edge (cf. Lindgren et al., 2019) will remain undetected for the writer, and possibly even more for those with decoding difficulties. It follows that the burden on decoding when composing with STT may differ from typing, and is probably an important ability for evaluating and correcting errors in the burst transcribed. Therefore, decoding ability should be considered when investigating fluency in transcription processes when composing by STT.

To sum up, one potential gain from composing with STT is that the user is able to plan subsequent content while the tool is converting the previous burst into written text. If the tool interprets a burst accurately, this will facilitate transcription, and presumably reduce the load on working memory, since the user will not need to either spell or use the motor processes required in keyboarding or handwriting and—perhaps most importantly—will not have to correct any errors. However, a certain amount of monitoring will be demanded to check the text transcribed by the tool, regardless its accuracy. Further, the text-written-so-far can presumably be useful for further planning or possibly help the user continue according to plan by producing

the next burst, which may already have been planned and may have been kept in working memory during the production of the preceding burst. However, if the writer hesitates or pronounces words unclearly, or if the tool "mishears" and produces one or more unintended words, the user must engage in problem-solving by evaluating the output of the STT tool to detect any errors and then by editing to correct any such errors. This will interrupt the transcription process and might place a high load on working memory since both error-correction and returning to the plan will be demanding processes.

If the user is able to employ appropriate strategies to handle such transcription disruptions, this is less likely to affect transcription fluency negatively. For this reason, both burst length and burst accuracy are factors of the STT process that should be explored to investigate whether there is any relationship with transcription fluency during composing.

In this study, we will explore what strategies children use when interacting with the STT tool and how effective these are in terms of accuracy. We will also investigate how those strategies relate to decoding, spelling and working-memory function. This will provide valuable insights into the utility of the STT tool and may also have implications for the instructions to be given to STT-tool users.

2. PURPOSE

The overarching purpose of this study is twofold. First, it aims to examine how children with and without difficulties behave when composing texts, in particular how they interact with the STT system. Second, it aims to investigate what aspects of children's behaviour affects transcription fluency, and further to exploit those insights to identify transcription strategies that could be taught in STT instruction.

We investigate transcription processes by exploring burst length and burst accuracy, and strategies for error correction, by identifying what strategies the children use and how successful they are. Higher-level revisions are not analysed. In addition, we investigate whether the participants' working-memory capacity, spelling ability and decoding ability predict their behaviour during composing as well as whether their behaviour during composing predicts their fluency when composing, measured by production rate.

2.1 Research questions

1) What strategies for transcription and error correction do children use when they compose using STT, and how accurate are they?
2) How are spelling ability, decoding ability and working-memory function associated with burst length and the use of various error-correction strategies in children who compose using STT?
3) How are transcription and error-correction strategies related to fluency in the STT composing process?

3. METHOD

3.1 Participants

28 participants, aged 10–13 years, were recruited from seven schools in southwest Sweden, see Table 1. All participants had experienced their entire schooling in Sweden. The only exclusionary criteria applied were a diagnosis of intellectual disability or autism. The participants were divided into two groups—with and without reading and writing difficulties—by classroom teachers and special educators who especially were asked to encourage children with reading and writing difficulties to participate. Group belonging was confirmed based on measures of spelling ability and decoding ability. All but three participants remained in their initial group. This resulted in one group with reading and writing difficulties (n = 16) and one group without such difficulties (n = 12), see table 1. The thresholds used for difficulty were stanine 3 or below on the spelling test or percentile 22 or below on the decoding tests of both words and nonwords, since these scores indicate difficulties according to the standardised, normed tests. Assessment of all background measures and the performance of all composition tasks took place individually at each participant's school, in all cases these were administered by the main author.

The STT composition process data from the 28 participants used in the present study has been retrieved from a broader set of data collected as part of a research project on STT and text production financed by the Marcus and Amalia Wallenberg Foundation (Ref. No. 2014–0122). The subset in question includes all participants for whom there were complete STT process data. The study has received ethical approval from the Swedish Ethical Review Authority (Ref. No. 702–17) and written assent/consent was collected from the participants and their caregivers.

3.2 Independent variables

Measures relating to spelling ability, decoding ability and working-memory capacity were assessed using standardised tests (spelling and decoding ability) or tests previously used in research (working-memory capacity). These measures were used to investigate the potential impact of the respective variables on burst length and accuracy as well as on the choice of error-correction strategy and the accuracy of error correction. Table 1 shows descriptive statistics relating to these measures for the two groups of children investigated—one group with spelling and decoding difficulties (hereafter *Spell*) and one reference group without such difficulties (hereafter *Ref*).

To control for the participants' ability to interact with the STT tool, the research group designed a special test in which the participants used STT to transcribe 18 predetermined sentences of different lengths (five to eight words long). This test assessed how accurately the tool was able to translate the participants' speech in a situation with as little planning or other higher-level processes involved as possible.

The sentences had been pre-recorded and each sentence was played back to the participants prior to production. In addition, the sentence was also available to them in writing. The number of words correctly produced by the tool was divided by the total number of words, yielding an *STT success rate*.

Table 1. Age, scores for background measures and STT success rate by group, means and standard deviations

Background measures	Text	Description of test	Spell (n=16)	Ref (n=12)	
			M (sd)	M (sd)	*p* =
Age			11.44 (0.95)	11.42 (0.73)	
Spelling ability	DLS 4-6[a]	Spelling words from sentences read aloud	-1.89 (1.01)	0.16 (0.61)	< .001***
Decoding words	LäSt[b]	Decoding real words from a word list	-1.17 (1.14)	0.13 (1.00)	.012*
Decoding nonwords	LäSt[b]	Decoding nonwords from a word list	-1.07 (0.69)	0.10 (1.20)	< .001***
Verbal working memory	CLPT[c]	Judging sentences for correctness while remembering words	0.89 (1.28)	0.65 (1.48)	.798
STT success rate		Producing pre-determined sentences using STT	77.4% (7.8)	82.6% (6.5)	.771

*[a]DLS 4–6 (Järpsten & Taube, 2010), [b]LäSt (Elwér et al., 2011), [c]CLPT (Gaulin & Campbell, 1994). Note: The raw scores have been converted to z-scores (M = 0, sd = 1) for ease of comparison. * = p < .05, *** = p < .001.*

Analysis of the background measures confirmed, in line with the selection criteria, that the *Spell* group scored significantly below the *Ref* group for spelling and decoding ($p < .05$) but the two groups did not differ on working memory and STT success rate.

3.3 Material

3.3.1 Text composition using STT

The task given to the participants was to write an expository text in Swedish using STT. The text was elicited by means of a short silent film clip that presented one of two moral dilemmas: cheating or stealing (Berman & Verhoeven, 2002). Order and topic were counterbalanced. The participants were given spoken instructions to reason about what a superhero would think of what happened in the film clip. They wrote in MS Word using an Apple computer (with the built-in STT system) and were able to pause the STT tool if they liked. The participants were informed that the spell-checker was turned off, and that they were not allowed to use text-to-speech to

listen to the text they had produced. The maximum composition time was announced as 30 minutes, but all participants were allowed to finish their text (time on task varied between 3.4 and 61.1 minutes, but only one participant exceeded 26.5 minutes). At the moment there is no keylogging program that has the possibility to log STT process data. Therefore, to enable analysis of the text-production process, the participants' activity was screen-recorded using Camtasia (TechSmith Corporation, 2018). This method makes it possible to analyse all on-screen events, including error-correction behaviours.

3.3.2 Identifying and annotating bursts in the composition process

The screen-recorded composition processes were annotated manually in ELAN (2019), which enables annotation in several tiers. Bursts, defined as chunks of words produced between two executive pauses, were extracted from each recording. The executive pauses were identified manually, since previous research has shown automatic identification of pauses in spoken language to be problematic (Fors, 2015). In this study, executive pauses were defined with reference to participants' use of intonation to highlight a pause between two chunks of words. Inter-rater reliability for the manually extracted bursts was calculated on 20% (six participants) of the material, which was analysed by two independent raters (raters 2 and 3) in addition to the first author (rater 1), yielding agreement rates of 90.5% between raters 1 and 2, 89.9% between raters 1 and 3, and 90.0% between raters 2 and 3. All three raters were in agreement for 88.7% of the bursts. Intra-rater agreement for rater 1 (who re-rated the relevant 20% of the material) was 91.2%.

Examples of annotations from the STT processes are presented in Tables 2 and 3. Each column represents one burst. The *Participant* row indicates what the participant dictated, *STT* shows what the STT tool transcribed, and *Translation* gives an English translation of the *Participant* row. *Burst type* indicates whether the burst represents initial composition (*comp.*, i.e., a burst consisting of content that is new in the text-composition process), a revision (*rev.rep.* for exact repetitions and *rev.change* for repetitions involving a change in burst length) or a deletion (*delete*). It should be noted that the examples below reflect only revision bursts that are related to failures on the part of the STT tool. Revisions not related to the STT tool were also annotated, but they are not analysed in this study. *Correct words* indicates the number of words in the burst and the number of words that the tool transcribed accurately. *F/NF* indicates whether the burst or the revision strategy used was functional (*F*) or non-functional (*NF*) in terms of accuracy. Finally, *Modality* shows the modality used to produce the burst (*keyboard* or *STT*; except for deletions, the examples only include cases of STT). Words in bold were not accurately transcribed by the tool.

Table 2. Annotation example 1

Burst number	**1**	**2**		**3**		**4**
Participant	när	**man**		**man**		man blir vuxen
STT	när	**namn**	~~namn~~	**mannen**	~~mannen~~	man blir vuxen
Translation	when	**you**	~~name~~	**you**	~~the man~~	you grow up
Burst type	comp.	comp.	delete	rev. rep.	delete	rev. change
Correct words	1/1	0/1		0/1		3/3
F/NF	F	NF		NF		F
Modality	STT	STT	keyboard	STT	keyboard	STT

Table 3. Annotation example 2

Burst number	**1**	**2**	**3**	**4**	**5**	
Participant	man får	dåligt	samvete	eftersom	**att**	
STT	man får	dåligt	samvete	eftersom	**allt**	~~allt~~
Translation	you get	bad	con-science	because	**that**	~~every-thing~~
Burst type	comp.	comp.	comp.	comp.	comp.	delete
Correct words	2/2	1/1	1/1	1/1	0/1	
F/NF	F	F	F	F	NF	
Modality	STT	STT	STT	STT	STT	keyboard
Burst number	**6**	**7**	**8**	**9**	**10**	
Participant	att	man	har	fuskat	och inte **lärt sig** något	
STT	att	man	har	fuskat	och inte **verkligen** något	
Translation	that	you	have	cheated	and did not **learn** anything	
Burst type	rev.rep.	comp.	comp.	comp.	comp.	
Correct words	1/1	1/1	1/1	1/1	3/5	
F/NF	F	F	F	F	NF	
Modality	STT	STT	STT	STT	STT	
Burst number		**11**				
Participant		lärt sig något				
STT	~~verkligen något~~	lärt sig något				
Translation	~~really anything~~	learn any-thing				
Burst type	delete	rev.change				
Correct words		3/3				
F/NF		F				
Modality	keyboard	STT				

The STT process data were extracted from ELAN (ELAN, 2019) into R (R Core Team, 2019) for further analysis.

3.4 Data analysis

The data were analysed with regard to the following:

Production rate. As a measure of fluency during composition, an overall production rate was calculated for each participant as the number of words remaining in the final text divided by the time, in minutes, spent on producing the text. Production rate as a measure has the advantage that it indicates how much final text the participants produced in a certain amount of time, that is, how disruptions during composition affects the amount of text produced. The measure has previously been used in studies of children's writing and in burst analysis (Alves & Limpo, 2015), in studies of speech recognition (Leijten, 2007), and in studies of people with reading and writing difficulties (Wengelin et al., 2014), enabling comparison of the present study with earlier findings. Another possible way to investigate fluency could be to calculate all words transcribed during the process. However, for STT there is a great risk that this measure would misleadingly favor compositions with long bursts consisting of numerous STT-errors with no words kept in the final text, and therefore this measure was not applicable. Production rate thus has its limitations: it does not indicate how much text was produced and then deleted during the process, meaning that it does not reflect any high-level revision such as deletion, substitution, reorganisation and addition that may have occurred. Since previous research into revision has shown that children in the present age-group rarely conduct any higher-level revisions (Kraft, 2023; Chanquoy, 2001), production rate was judged to be the most feasible measure to investigate fluency.

Burst length and *burst accuracy.* Burst length was calculated in words. Burst accuracy was dependent on whether the tool "misheard", or if the writer hesitated or pronounced words unclearly. For burst accuracy, a proportional measure (the number of accurately produced words in a burst divided by the total number of words produced in that burst) was chosen instead of a binary measure of correct/incorrect for the whole burst. This was because a binary measure would misleadingly favour short bursts. The more words a burst contains, the greater the risk becomes that it will contain some inaccuracy. For example, a participant might dictate the sentence "my superhero would think that the teacher should be more observant" (burst length: 11 words) and the tool might produce "my superhero would think that a teacher should be more observant", with ten words correct but "a" instead of "the". With a binary measure, that burst would be classified as "inaccurate" although almost all of the words came out correct and it made a major contribution to the participant's production rate. If a participant produces ten 11-word bursts with a single error in each, the burst-accuracy rate will be zero even though 100 out of 110 words were correct. By contrast, if a participant dictates the eleven words of the above-mentioned sentence in one-word bursts and only five of them come out correct, the mean accuracy rate will be all of 45%. Such a result would obviously be misleading when it comes to which burst length is more functional in terms of its effect on the production rate.

Error-correction strategy and *error-correction accuracy.* Three strategies for transcription error correction (that is, correction of STT-errors) during composition were identified: (a) repeating the same wording with the same burst length, (b) repeating the same wording but changing the burst length, and (c) typing the word on the keyboard. An instance of use of a strategy was marked as functional if it yielded a correct outcome, that is, if the tool reproduced the burst correctly or if the typed word was correctly spelled. If the participant changed a word, phrase or sentence, this was not considered an error correction, and was excluded from error correction analysis.

4. RESULTS

4.1 Burst length and burst accuracy

Table 4 shows descriptive statistics by group for various burst lengths: the total and mean number of bursts of each length as well as the respective burst-accuracy rates.

The most frequent type of burst produced was the one-word burst: M = 32.75 (30.65) for *Spell* and M = 25.92 (18.83) for *Ref*. All 28 participants produced at least one such burst. The second-most frequent type of burst was the two-word burst: M = 9.00 (8.17) for *Spell* and M = 8.67 (5.69) for *Ref*. The bursts then continued to become less frequent as they grew longer. The longest single category, the ten-word burst, was produced by fewer than half of the participants (n = 12) and there were only 25 such bursts in the entire material. A final category, that of > 11-word bursts, includes all bursts containing at least 11 words. More than half of the participants (n = 17) produced at least one of these very long bursts; however, it should be noted that their overall accuracy rate was the lowest of any category: M = 0.54 (0.20) for *Spell* and M = 0.44 (0.24) for *Ref*. To show the mean variation in burst length, the mean-value for each participants mean-burst length (composition and revision bursts) was calculated; it was M = 2.65 (1.04), min-max = 1.36–5.53 for *Spell* and M = 2.60 (0.83), min-max = 1.79–4.34 for *Ref.*

The descriptive statistics also indicate that the least successful burst type (except for the > 11-word burst) in terms of its accuracy rate was the one-word burst: M = 0.60 for *Spell* and M = 0.60 for *Ref*. The burst type with the highest accuracy differed between the groups: for *Spell*, it was the three-word burst (M = 0.73) while for *Ref* it was the seven-word burst (M = 0.79). However, it should be noted that three out of twelve participants in the *Ref* group did not produce a single seven-word burst. At a general level, the three-word burst had an overall high accuracy rate in both groups, and all but two participants produced at least one such burst. This suggests that a useful strategy (in terms of length) was to use a burst length of around three words and—more importantly—not to use one-word bursts.

Table 4. Descriptive statistics by group for various burst lengths: total and mean number of bursts of each length as well as the respective burst-accuracy rate

	Spell (n= 16)				Ref (n = 12)				
Burst length	**part.**	**Bursts (min–max)**	**M (sd) bursts**	**M (sd) accuracy**	**part.**	**Bursts (min–max)**	**M (sd) bursts**	**M (sd) accuracy**	***p* = accuracy**
1 word	16	524 (1–102)	32.75 (30.65)	.60 (.21)	12	311 (4–62)	25.92 (18.83)	.60 (.18)	.83
2 words	15	144 (0–35)	9.00 (8.17)	.64 (.15)	12	104 (1–19)	8.67 (5.69)	.67 (.30)	.34
3 words	15	80 (0–15)	5.00 (4.60)	.73 (.20)	11	67 (0–11)	5.58 (3.45)	.73 (.24)	.92
4 words	15	68 (0–11)	4.25 (3.28)	.65 (.20)	9	40 (0–10)	3.33 (3.11)	.73 (.21)	.30
5 words	14	48 (0–7)	3.00 (2.16)	.63 (.25)	11	44 (0–9)	3.67 (2.96)	.72 (.26)	.28
6 words	14	43 (0–11)	2.69 (2.75)	.68 (.20)	10	39 (0–9)	3.25 (2.67)	.70 (.20)	.93
7 words	11	26 (0–6)	1.62 (1.67)	.69 (.28)	9	25 (0–6)	2.08 (1.83)	.79 (.15)	.59
8 words	10	18 (0–4)	1.12 (1.15)	.70 (.32)	7	17 (0–5)	1.42 (1.73)	.65 (.45)	.92
9 words	8	13 (0–3)	0.81 (1.05)	.63 (.35)	7	16 (0–6)	1.33 (1.78)	.74 (.27)	.65
10 words	7	15 (0–4)	0.94 (1.24)	.62 (.22)	5	10 (0–5)	0.83 (1.47)	.78 (.27)	.19
> 11 words	11	33 (0–8)	2.06 (2.29)	.54 (.20)	6	19 (0–9)	1.58 (2.75)	.44 (.24)	.45

part. = number of participants having at least one burst of a given length. The p-values pertain to between-group comparisons with regard to burst accuracy.

4.1.1 Predictors of burst length and burst accuracy

We used R (R Core Team, 2019) and lme4 (Bates et al., 2015) to perform a linear mixed-effects regression analysis to investigate whether the independent variables predicted the participants' behaviour (in terms of burst length) in their initial composition (that is, no revision bursts were included in the analysis). We entered the number of words in each burst as the outcome variable and subject as random intercept. Then we added working-memory function, spelling ability, and decoding ability as fixed effects. We also added the accuracy of the STT tool's transcription of the participants' speech in pre-determined sentences (STT success rate) as a fixed effect. The fixed effects were added one by one and compared with the intercept-only model, where only the random intercept for each subject was included. None of the factors added contributed to a better model ($\chi^2(1) = 0.08$, $p = .77$ for working memory; $\chi^2(1) = 0.01$, $p = .93$ for spelling ability; $\chi^2(1) = 0.41$, $p = .52$ for decoding

ability; and $\chi^2(1) = 1.62$, $p = .20$ for STT success rate, relative to an intercept-only model). This indicates that the lengths of bursts that the children produced were interdependent of their working-memory capacity, their spelling and decoding ability and their ability to make themselves understood to the STT software under optimal conditions.

We also conducted a linear mixed-effects regression analysis to investigate whether working-memory capacity, spelling ability and decoding ability predicted the accuracy rate for each burst length (revisions excluded), meaning that we examined the interaction between the child and the STT tool. We entered the accuracy rate for each burst length as the outcome variable and subject as random intercept. Then we added working-memory function, spelling ability and decoding ability as fixed effects. We also added the STT success rate as a fixed effect. The fixed effects were again added one by one and compared with the intercept-only model, where only the random intercept for each subject was included. None of working-memory capacity, spelling ability and decoding ability contributed to a better model ($\chi^2(1) = 1.10$, $p = .29$ for working memory; $\chi^2(1) = 0.01$, $p = .93$ for spelling ability; and $\chi^2(1) = 0.48$, $p = .49$ for decoding ability, relative to an intercept-only model), but the STT success rate did ($\chi^2(1) = 6.37$, $p = .012$). In other words, the interaction between the child and the STT tool under optimal conditions predicted the accuracy of the STT tool during text composition, but working-memory capacity, spelling ability and decoding ability did not.

4.2 Error-correction strategies and success rate

When the STT tool made errors, the participants (provided that they detected the errors) were forced to engage in problem-solving to correct those errors. In this context, we observed the following correction strategies: (a) repeating the previous burst, (b) repeating the same wording but changing the burst length, and (c) correcting the error by typing on the keyboard. Table 5 shows descriptive statistics by group for those error-correction strategies: the number of participants using each strategy, the total and mean number of instances of each strategy and the success rate for each strategy.

Simply repeating the previous burst proved to be the least successful strategy. In some cases, participants persevered in repeating their initial burst up to six times until they finally attained an accurate result. In these cases, many words that had actually been accurately produced by the tool were deleted; this is part of the reason why the average mean for all repeated bursts yielded a low success rate for both groups (30% for *Spell* and 44% for *Ref*). In other words, a participant who repeated the previous burst had to perform yet another correction in about 60% of cases. Choosing instead the strategy of keeping the wording but changing the burst length increased the accuracy rate to 53% for *Spell* and 54% for *Ref*. That is, changing the burst length was a successful strategy in a little over half of all cases. Finally —and perhaps surprisingly—correction by typing was by far the most successful error-

correction strategy, with a mean accuracy rate of 100% for *Spell* and 99% for *Ref*. This result will be discussed below (see section 5.1), but it should be noted even now that minor corrections to tense or number endings were included in this category.

Table 5. Descriptive statistics by group for the three error-correction strategies: total and mean number of instances as well as success rate

	Spell (n=16)				Ref (n = 12)				
Strategy	**Part.**	**Bursts (min–max)**	**M (sd) bursts**	**M (sd) success**	**part.**	**Bursts (min–max)**	**M (sd) bursts**	**M (sd) success**	**p = (success)**
Repeating burst	15	122 (0–21)	7.62 (6.38)	.30 (.27)	12	88 (1–16)	7.33 (5.68)	.44 (.38)	.42
Changing length	15	239 (0–45)	14.94 (14.74)	.53 (.26)	12	112 (1–22)	9.33 (7.34)	.54 (.31)	.96
Keyboard	16	201 (1–34)	12.56 (9.48)	1 (0)	12	214 (4–42)	17.83 (11.04)	.99 (.02)	.11

Part. = number of participants using a strategy. Bursts = total number of bursts for each group. The p-values pertain to inter-group comparison with regard to success rate.

As regards differences between the two groups, the descriptive statistics revealed that, on average, the participants with difficulties performed more error corrections than their peers without such difficulties (although this difference was not statistically significant). This suggests that, contrary to our expectations, the members of the former group are capable of detecting (i.e., reading) errors produced by the tool to a high extent. As shown in Kraft (2023) the proportion STT errors left in final text were few: M = 3 (3), min–max = 0–11 for *Spell*, and 1(2), min–max = (0-8) for *Ref*.

4.2.1 Predictors of the choice of error-correction modality

We used R (R Core Team, 2019) and lme4 (Bates et al., 2015) to perform a generalised linear mixed-effects analysis to investigate the participants' behaviour during error correction—specifically, to determine whether working-memory capacity, spelling ability and decoding ability predicted a participant's choice of error-correction modality. As outcome variable, we entered modality (STT or Keyboard). As fixed effects, we entered working-memory function, spelling ability and decoding ability. We also added the STT success rate as a fixed effect. As random effects, we had intercepts for subjects. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. We added variables stepwise, comparing each subsequent model with a null model where only the random intercept for each subject was included; *p*-values were obtained by means of likelihood-ratio tests. The intercept-only model was found to be the best model, meaning that neither working memory, spelling ability or decoding ability nor the ability to make

oneself understood to the STT software under optimal conditions predicted the modality (STT or keyboard) a participant chose to use for error correction ($\chi^2(1) = 0.45$, $p = .50$ for working memory; $\chi^2(1) = 2.69$, $p = .10$ for spelling ability; $\chi^2(1) = 3.20$, $p = .07$ for decoding ability; and $\chi^2(1) = 0.00$, $p = 1.00$ for the STT success rate, relative to an intercept-only model). This indicates that the choice between STT and keyboard when error correcting was independent of their individual abilities.

We also conducted a generalised linear mixed-effects analysis to investigate error-correction functionality. As outcome variable we entered functionality (yes/no—that is, whether or not an instance of strategy use was functional in the sense that it yielded the correct word). As fixed effects, we entered working-memory capacity, spelling ability and decoding ability. We also added modality (STT or keyboard) and STT success rate as fixed effects. As random effects, we had intercepts for subjects. We added the variables stepwise, comparing each model with a null model where only the random intercept for each subject was included; *p*-values were obtained by means of likelihood-ratio tests. Modality turned out to be the only statistically significant predictor of error-correction functionality ($\chi^2(1) = 183.80$, $p < .001$, relative to an intercept-only model). None of working memory, spelling ability, decoding ability and STT success rate contributed to a better model for predicting error-correction functionality ($\chi^2(1) = 2.01$, $p = .16$ for working memory; $\chi^2(1) = 0.85$, $p = .36$ for spelling ability; $\chi^2(1) = 0.53$, $p = .47$ for decoding ability; and $\chi^2(1) = 1.76$, $p = .18$ for STT success rate, relative to the model with modality as fixed effect). This indicates that the choice between using STT or the keyboard to correct an error predicted whether the correction was successful or not.

4.3 Production rate

Descriptive statistics by group pertaining to production rate are presented in Table 6. The production rate was higher for *Ref* (9.02 words/minute) than for *Spell* (6.46 words/minute), but a Mann–Whitney *U* test showed that difference not to be statistically significant ($p = .133$).

Table 6. Production rate by groups: means and standard deviations

	Spell (n = 16)	**Ref (n = 12)**	
Measure	**M(sd)**	**M(sd)**	***P* =**
Production rate (text length/time on task)	6.46 (3.64)	9.02 (4.62)	.133

The p-value pertains to between-group comparison with regard to production rate.

4.3.1 Predictors of production rate

We used linear regression to investigate what predicted the production rate. In a first model, Model 1, we entered production rate as the outcome variable, and in a first step we added working-memory function, spelling ability and decoding ability. The overall regression was statistically significant ($R^2 = .20$, $F(3, 24) = 3.23$, $p = .03$),

and this model was significantly better than the null model ($F(6, 21) = 3.22$, $p = .04$). Working memory significantly predicted production rate ($\beta = 0.39$, $p = .03$). This means that Model 1 predicted 19.82% of the variance in the production rate.

In the next step, Model 2, we added measures from the composition process. Since the one-word bursts had a very low accuracy rate, we excluded them from the analysis. Further, since the variable of median burst length had a right-skewed distribution, a logarithm transformation was used prior to analysis. In addition, since accuracy by burst length was not normally distributed, a breakdown was made into two categories: high accuracy and low accuracy. The final measures added in Model 2 were (a) median burst length, (b) accuracy level (high or low accuracy) of composition bursts (i.e., revision bursts were not included) and (c) mean revision success rate (for all revisions, i.e., including the strategies of *repeating burst*, *changing length* and *correction by keyboard)*. Model 2 turned out to be significantly better than Model 1 ($F(3, 21) = 6.77$, $p = .002$) and explained 53.43% of the variance in the production rate—34 percentage points more than Model 1. The overall regression was statistically significant ($R^2 = .5343$, $F(6, 21) = 6.16$, $p < .001$). Further, it was found that the production rate was significantly predicted by three variables: median burst length ($\beta = 0.38$, $p = .02$), accuracy level (high or low) of composition bursts ($\beta = 0.40$, $p = .02$) and working-memory capacity ($\beta = 0.36$, $p = .02$). By contrast, neither the mean revision success rate nor spelling or decoding ability significantly predicted the production rate.

4.4 Exploration of the one-word burst

Since the one-word burst was the most common type of burst but also the least accurate one (except for bursts of 11 words or more), we wished to explore what influenced its success rate. One factor of potential relevance in this context is word length, which we chose to analyse. Considering that typing was an error-correction strategy commonly used when one-word bursts had failed, we also decided to compare the words found in one-word bursts with words typed on the keyboard, to investigate whether, and if so how, the two modalities might complement each other during STT composition.

The one-word bursts were divided into accurately and non-accurately transcribed ones. The typed words were categorised as correctly spelled or not.

Analysis for word length of the accurate and non-accurate one-word bursts yielded a statistically significant difference ($p < .001$); see Table 7. Specifically, the words accurately produced by the STT tool were significantly longer than the non-accurately produced ones. For typed words, the opposite pattern was found: the correctly spelled words were significantly shorter than those containing spelling errors.

We then proceeded to check whether this was really an effect of length rather than of frequency, given that short words are typically more frequent than long words (Harley, 2014). To do this, we compared the words found in the accurately

and non-accurately produced one-word bursts with a frequency corpus for Swedish: *The Stockholm Umeå Corpus* or SUC (Gustafson-Capková & Hartmann, 2006). The statistics we used pertained to SUC3, provided by Språkbanken (Borin et al., 2012); we chose it because it is large (over seven million words) and includes texts representing different genres and styles.

We used the 1,000 most frequent words in the corpus to create a high-frequency corpus. Punctuation, given names, and names of towns and cities were excluded. Then we compared all dictated one-word bursts and typed words with the high-frequency corpus. For the typed words, we could see the expected pattern: the correctly spelled (shorter) words were more likely to belong to the high-frequency corpus (74.2%) than the incorrectly spelled (longer) words (42.9%). That is, the participants spelled high-frequency words correctly more often than lower-frequency words. Interestingly, this held for the dictated one-word bursts as well. The proportion of words belonging to the high-frequency corpus was 75.7% for the accurately transcribed words and 70.6% for the non-accurately transcribed words. However, the difference was much bigger for typing. To sum up, correctly spelled typed words were shorter and more frequent than incorrectly spelled typed words. Accurately transcribed one-word bursts were longer—but also involved more frequent words—than non-accurately transcribed ones. However, a chi-square test showed that the percentage of words belonging to the 1,000 most frequent words in the SUC3 corpus did not significantly differ by accuracy for either STT (χ^2 (1, 808) = 2.21, p = .14) or typing (χ^2 (1, 104) = 1.56, p = .21).

These results indicate that success when dictating one word at a time is more likely for longer words, while using the keyboard is effective for short and relatively frequent words. Hence an appropriate combination of STT and keyboard use could potentially be useful during composition with STT.

Table 7. Word-length means and standard deviations for accurate and non-accurate one-word bursts and typed words

Modality	**Accurate**		**Non-accurate**		
	Bursts	**M (sd)**	**Bursts**	**M (sd)**	***p* =**
STT	424	4.62 (2.33)	384	3.97 (1.98)	< .001***
Keyboard	156	4.03 (2.00)	14	4.93 (1.64)	.036*

Comparisons were calculated using the Mann–Whitney U test. The p-values pertain to comparison regarding word length between the modalities. * = p < .05; *** = p < .001.

5. DISCUSSION

The aim of our study was to explore how children with and without spelling difficulties interact with the STT system when composing, and further to investigate whether their behaviour, in turn, affects their production rate. Our study is unique

in investigating the transcription process during children's composition with STT. We set out to answer three questions: (1) What strategies for transcription and error correction do children use when they compose using STT, and how accurate are they? (2) How are spelling ability, decoding ability and working-memory function associated with burst length and the use of various error-correction strategies in children who compose using STT? And (3) How are transcription and error-correction strategies related to fluency in the STT composing process? In the following, we will address each research question in turn.

5.1 Strategies for transcription and error correction

Looking first at burst length, we found that, overall, the most common type of burst produced was the one-word burst. This may seem a bit surprising, but a large part of the reason for this result is that, when the tool transcribed a word incorrectly, many writers tended to re-dictate that word in a one-word burst. Even words with non-problematic spelling, such as *en* ('a') and *i* ('in'), were dictated separately on such situations. This yielded a low STT accuracy rate, which was probably due to, at least in part, a lack of context. In addition, the children may sometimes have hyper-articulated in a way that would have worked better with a human listener than with the STT-tool. It should be noted that the one-word burst was in fact not only the most common burst type, but also the least successful one. In this context, it should be pointed out that the accurately transcribed one-word bursts were significantly longer than the inaccurately transcribed ones. Hence the STT tool is more useful for long words—and those were more likely than short words to be spelled incorrectly when participants used the keyboard. The participants' choice to use one-word bursts, instead of typing, even when correcting short words might also be due to the instructions they had received: they were told to compose a text using STT, even though they were allowed to use the keyboard if needed. A further reason could be their lack of experience with the STT tool.

Our analysis indicates that the most successful error-correction strategy was correction by typing. At first sight, it seems as though typing is far more useful than dictating for this purpose. However, several things need to be kept in mind to fully understand this result. First, the participants mostly used typing when STT had failed (often failed repeatedly). Second, all corrections performed using the keyboard were included, meaning that a large number of minor corrections to tense or other endings affected the result. Hence the general conclusion to be drawn is that it is important to combine STT with keyboard use in order to circumvent the obstacles caused by the STT tool.

Further, our results showed that both groups in the study engaged in error correction, meaning that even the children with difficulties checked (i.e., read) the text produced by the tool. Interestingly, this contrasts with findings from keylogging studies indicating that children with reading and writing difficulties engage very little in reading the text they have written so far (Johansson et al., 2008). Given that the

detection and subsequent correction of errors in a text can trigger other types of revisions (Conijn et al., 2021), one important future question to investigate is whether composition using STT offers an opportunity to engage children in other, including more high-level, revisions than the mere correction of transcription errors made by the tool. However, the error-correction aspect of composing by means of STT may in and of itself cause a heavy cognitive load, meaning that this mode of transcription could simply redirect the problem-solving focus of children with spelling difficulties from spelling to correcting errors made by the tool.

5.2 Association between individual abilities and strategies for composing, on the one hand, and error correction, on the other

Neither burst length, nor burst accuracy, nor the choice of error-correction strategy was predicted by spelling ability, decoding ability or working-memory function. Our failure to find an effect of any of these factors indicates that participants with and without difficulties behaved similarly during composition. This strengthens the feasibility of STT as a useful writing tool for children with difficulties. However, accuracy when composing was predicted by how well the STT tool 'understood' the participants under optimal conditions.

5.3 Fluency during composition

Burst length and burst accuracy, but not correction-success rate, predicted the production rate. That is, how the children behaved and interacted with the STT system during initial composing affected their production rate. Further, working memory—but no other individual ability—also predicted the production rate. This suggests that even when the demands of spelling are removed, composing text is a cognitively complex process placing a heavy strain on working memory.

Because this study is the first to investigate production rate when children with difficulties compose using STT, comparison with previous studies is difficult. However, if we compare the difference in production rate between the children with and without difficulties in our study with the corresponding difference found in a Swedish study investigating production rate in slightly older children (15 years) with reading and writing difficulties composing by means of a keyboard (Wengelin et al., 2014), we can see that the difference between the groups in our study is smaller: 6.46 words/minute for *Spell* and 9.02 words/minute for *Ref*, compared with 11.06 and 20.3 words/minute, respectively, for keyboarding teenagers with and without difficulties (Wengelin et al., 2014). This suggests that STT may have potential as a facilitatory tool, especially for children with difficulties.

However, as mentioned before, production rate is a large-grained measure. What is more, a slow production rate is not necessarily negative. Events other than actual text production during the composing process, such as pauses for planning or rewording content (reflecting higher-level reviewing processes), could bias this; hence

this result should be interpreted with caution (cf. McCutchen, 1996). For this reason, there is a need to further explore the revisions performed during composition with STT in terms of their level, but that is unfortunately outside the scope of this paper.

5.4 Limitations

Some limitations of the present study should be reported. First, the small number of participants makes it necessary to interpret our results with caution. More studies on the topic are needed to draw strong conclusions. Second, the study did not analyse the participants' voice quality or speech rate, these factors could potentially influence how accurately the STT tool understands the children. Third, we did not investigate the existence of words that are phonologically similar to the non-accurately transcribed words, which could affect burst accuracy.

Further, the writing session most likely differed from the participants everyday writing, since they were not allowed to use spell checker or other compensatory tools during the session, which could have affected their behaviour during composing.

While manual annotation was chosen over automatic for the identification of pauses because of its greater reliability (Fors, 2015), there is also a risk of errors due to the human factor. However, the inter- and intra-rater agreement rates were acceptable.

Finally, while the results from this study have implications for how a speech-recognition tool can be implemented to enable a high production rate, it must be pointed out that the present study did not relate production rate to text quality. Hence a future issue to investigate is how transcription fluency and production rate are related to text quality.

6. FUTURE RESEARCH AND PRACTICAL IMPLICATIONS

Our results support the view that composing text using STT is a cognitively complex process placing heavy demands on working memory. Further, since an STT user has to monitor the accuracy of the tool's output while composing, interaction with an STT system may change how the planning of linguistic content is managed and how that planning draws upon working memory during composition. Whilst it is safe to assume that users engage in such monitoring of accuracy during the pauses they make, the possibility that they are also simultaneously holding linguistic content for the next burst in working memory, such that planning extends to two or more bursts, cannot be ruled out. It could be valuable to further investigate the pause distribution of composing processes to gain insight into users' planning processes and to analyse whether, as it might be hypothesised, longer pauses are related to larger syntactic units and shorter pauses to the word, phrase or sentence level. However, we cannot exclude the possibility that burst length does not tell us anything about the writer's planning processes *per se* but rather reflects a strategy to handle the limitations of

working-memory capacity. Since any more extensive, higher-level plan will be impossible to transform into words in one go, writers may use bursts to keep their plan active in memory, perhaps by adjusting it to take into account that part of it has been put into words.

The finding that bursts of different lengths have different accuracy rates raises new questions about how writers package linguistic content in bursts. The production of, say, three-word bursts (one of the burst lengths with the highest accuracy) requires the ability to (at least occasionally) package output with a non-fixed syntactic structure, and also to split a sentence, or even a phrase, into smaller units. Future research should therefore include syntactic analysis of the bursts. In line with Olive and Cislaru (2015) a preliminary "test drilling" indicated that the participants in our study seemed to prefer nesting the entire noun phrase within one burst, but apart from this, we found little by way of obvious syntactic patterns. For example, both verb phrases and prepositional phrases could be split across bursts, and similar observations were made for subordinate clauses, where the conjunction or pronoun starting the clause sometimes was not in the same burst as the rest of the clause. It would therefore clearly be of interest for future studies to investigate more thoroughly to what extent different syntactical structures tend to be divided across bursts, and whether flexibility in this respect is dependent on language ability or working-memory function. It could for example be possible that a person with language constraints and less linguistic flexibility will be more rigid and dependent on syntactic structure and so will have difficulties adjusting burst length in violation of syntactic structure. However, to our knowledge there is no existing research on syntactic structures in bursts in relation to language ability. Previous research has shown that children with a developmental language disorder compose the same number of bursts as their age-matched peers on a group level, but that their bursts are shorter (Connelly et al, 2012). What information that is lost in those shorter bursts is yet unexplored. Further investigation of the distribution of grammatical structures across bursts in different conditions and its relationship to working memory and general language ability would give us valuable insights about those whose writing could be facilitated by STT, and about what can be done to help writers benefit from this modality.

One interesting aspect of the potential usefulness of STT is that it could be more motivational to correct transcription errors made by a tool than to correct spelling errors made by oneself. In fact, one participant in our study reflected on this, stating that error correcting when composing with STT was more motivating than correcting spelling errors, because the errors that needed correcting had been "made by someone else". At a general level, motivation is an important predictor of writing performance (Camacho et al., 2021), meaning that it is important for future research to investigate motivation (including its development over time) in composition using STT.

Our results showed that the characteristics of the participants' interaction with the STT tool during composition affected their production rate. Both burst length

and burst accuracy predicted the production rate. Together with working-memory capacity, they explained 53.43% of the production rate. However, none of spelling ability, decoding ability and error-correction accuracy explained either burst length or burst accuracy, indicating that the participants' spelling and decoding ability did not predict their behaviour during composition. By contrast, working-memory capacity significantly predicted the production rate. This indicates that composing with STT is a demanding, complex activity—as is indeed writing in general (Hayes, 2012). On the surface, STT may seem to reduce the cognitive requirements of spelling, but it presumably places a heavy load on executive processes involved in writing, such as self-regulation and error detection and correction, quite possibly to a similar extent as the processes involved in the management of spelling difficulties. On the positive side, strategies for self-regulation can be taught (for a meta-analysis, see Graham et al., 2012), and there also seems to be an obvious potential for optimising the choice between dictating and keyboarding for error correction. Hence the possibility of combining instruction on STT transcription and instruction on self-regulation should be explored. Furthermore, it should be noted that the participants in our study had received only brief instructions about how to use the STT tool when they composed their texts, and they had no previous experience of composing text with STT. Our results also showed that the interaction between the participants and the tool in the STT test under optimal conditions predicted transcription accuracy during composing, and it is highly likely that, with training and experience, user and tool will come to understand each other better. For this reason, it is important to consider individual STT accuracy rates under optimal conditions before STT is implemented as a writing tool.

Further, our results highlight certain behaviours and strategies that may be useful during composing with STT, and those behaviours and strategies can be taught to users as part of implementing STT as a writing tool. For example, it was found that one strategy for enhancing STT accuracy was to change the burst length—but not to a length of one word. In this study, the group with spelling difficulties had a lower production rate than the group without them and also produced more one-word bursts (even though these differences were not statistically significant). Teaching children to avoid one-word bursts while composing could potentially reduce some of the variance in production rate, and investigating the implementation of this strategy in an intervention study would be an important issue for future research.

A further suggestion is to instruct children to combine keyboard use with STT: short, highly frequent words are well suited for typing (especially for correction purposes, but perhaps also for original composition), while longer words are more appropriate for dictation. Explicitly instructing users that STT is useful for the composition of long words (that are hard to spell) might influence how they use the tool. Kraft et al. (2019), who did not notice any difference in word length when comparing texts written on a keyboard with texts written using STT by children aged 9–12 years, suggested that this could be because the children had not yet realised that STT can enable the production of long words that are hard to spell and so could let them use

words they would steer clear of when typing. By contrast, Higgins and Raskind (1995) reported differences in word length between these modalities in adults with difficulties, and the participants in their study did show an awareness that the tool was useful for producing long words that were hard to spell. The teaching of writing-process strategies has previously been shown to have positive effects on the writing process (Hayes & Berninger, 2014), and future research should address an intervention targeting STT-tool instruction (transcription) in combination with writing instruction (composition).

Our regression model for production rate did not explain all of the variance, suggesting that aspects of writing other than mere transcription—which go beyond this study—could be influential. For example, it is presumably the case that the participants' genre knowledge and their reading and writing habits are important. Previous research has shown that reading and writing difficulties have far-reaching consequences and that people with reading and writing difficulties read and write less than their peers without such difficulties, causing them to have fewer text experiences and hence less genre knowledge (Stanovich, 1986). Thus, it is highly unlikely that STT as a composition tool for children with spelling difficulties will be enough in and of itself for helping those children improve their writing skills. This further strengthens the case for combining instruction about how to use the tool with instruction pertaining to higher-level writing processes such as planning or revising content.

To conclude, it would be naïve to implement STT as a writing tool for children with reading and writing difficulties and expect this to immediately (after 15 minutes instruction) solve all of the problems that are involved in writing (for an overview of this complexity, see Skar et al., 2022). However, the results from this study do show that children use a variety of strategies when they dictate, and some of those strategies contribute to raising their fluency. Our next step will therefore be to carry out an intervention study targeting such strategies alongside writing instruction and to compare text quality before and after the intervention.

ACKNOWLEDGEMENTS

We gratefully acknowledge the insightful comments and questions from two anonymous reviewers and the editor. We also express our gratitude to Marcus and Amalia Wallenberg foundation for enabling us to carry out the research reported here. We also would like to acknowledge Johan Segerbäck for language editing. Any remaining mistakes and errors are the sole responsibility of the authors.

REFERENCES

Alves, R. A., Branco, M., Castro, S. L., & Olive, T. (2011). Effects of Handwriting Skill, Output Modes, and Gender on Fourth Graders' Pauses, Language Bursts, Fluency, and Quality. In V. Berninger (Ed.). *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 415-428). Hove: Psychology Press.

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading*, *19*(5), 374–391. https://doi.org/10.1080/10888438.2015.1059838

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beers, S. F., Mickail, T., Abbott, R., & Berninger, V. (2017). Effects of transcription ability and transcription mode on translation: Evidence from written compositions, language bursts and pauses when students in grades 4 to 9, with and without persisting dyslexia or dysgraphia, compose by pen or by keyboard. *Journal of Writing Research*, *9*(1), 1. https://doi.org/10.17239/jowr-2017.09.01.01

Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy*, *5*(1), 1–43. https://doi.org/10.1075/wll.5.1.02ber

Berninger, V. W. (2006). A developmental approach to learning disabilities. In R. K. Ann & I. E. Sigel (Eds.), *Handbook of child psychology* (Vol. IV Child Psychology in Practice. No. 2 Research Advances and Implications for Clinical Applications). Wiley Online Library. https://doi.org/10.1002/9780470147658.chpsy0411

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., . . . Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of educational psychology*, *94*(2), 291. https://doi.org/10.1037/0022-0663.94.2.291

Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp—the corpus infrastructure of språkbanken. In *Proceedings of lrec 2012. istanbul: Elra* (Vol. Accepted, pp. 474–478). retreived from: https://spraakbanken.gu.se/resurser/bloggmix2012

Camacho, A., Alves, R. A., & Boscolo, P. (2021). Writing motivation in school: a systematic review of empirical research in the early twenty-first century. *Educational Psychology Review*, *33*(1), 213–247. https://doi.org/10.1007/s10648-020-09530-4

Conijn, R., Speltz, E. D., Zaanen, M. van, Waes, L. V., & Chukharev-Hudilainen, E. (2022). A Product- and Process-Oriented Tagset for Revisions in Writing. *Written Communication*, *39*(1), 97–128. https://doi.org/10.1177/07410883211052104

Connelly, V., Dockrell, J. E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology*, *25*(1), 99–107. https://doi.org/10.1080/0144341042000294912

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment. *Written Communication*, *29*(3), 278–302. https://doi.org/10.1177/0741088312451109

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

ELAN (Version 5.6) [Computer software]. (2019). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Elwér, Å., Fridolfsson, I., Samuelsson, S., & Wiklund, C. (2011). *LäSt: Test i läsning och stavning [LäSt: Test in reading and spelling].* Hogrefe, Psykologiförlaget.

Fors, K. L. (2015). *Production and perception of pauses in speech* (Doctoral dissertation, Department of Philosophy, Linguistics, and Theory of Science, University of Gothenburg).

Gaulin, C. A., & Campbell, T. F. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and motor skills*, *79*(1), 55–64. https://doi.org/10.2466/pms.1994.79.1.55

Grabowski, J. (2008). The internal structure of university students' keyboard skills. *Journal of writing research*, *1*(1). https://doi.org/10.17239/jowr-2008.01.01.2

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of educational psychology*, *104*(4), 879. https://doi.org/10.1037/a0029185

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? a meta-analytic review. *Reading and Writing*, *27*(9), 1703–1743. https://doi.org/10.1007/s11145-014-9517-0

Gustafson-Capková, S., & Hartmann, B. (2008). Manual of the Stockholm Umeå corpus version 2.0. Stockholm University. *Unpublished Work*.

Harley, T. A. (2014). Recognizing visual words. In T. Harley (Ed.) *The psychology of language: From data to theory (4th ed.).* (pp. 167–208). Psychology Press.

Haug, K. N., & Klein, P. D. (2018). The effect of speech-to-text technology on learning a writing strategy. *Reading & Writing Quarterly*, *34*(1), 47–62. https://doi.org/10.1080/10573569.2017.1326014

Hayes, J. R. (2009). From idea to text. In D. Myhill (Ed.) *The Sage handbook of writing development* (pp. 64–79). London: Sage.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hayes, J. R., & Berninger, V. W. (2014). Cognitive processes in writing: A framework. In J. Dockrell, B. Arfé & V. Berninger (Eds.) *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction*, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199827282.003.0001

Higgins, E. L., & Raskind, M. H. (1995). Compensatory effectiveness of speech recognition on the written composition performance of postsecondary students with learning disabilities. *Learning Disability Quarterly*, *18*(2), 159–174. https://doi.org/10.2307/1511202

Järpsten, B., & Taube, K. (2010). *DLS: för skolår 4–6. Handledning [DLS: for school years 4–6. Manual].* Hogrefe, Psykologiförlaget.

Johansson, R., Johansson, V., Wengelin, Å., & Holmqvist, K. (2008). Reading during writing: Four different groups of writers. *Working papers/Lund University, Department of Linguistics and Phonetics, 53,* 43-59.

Kaufer, D. S., Hayes, J. R., & Flower, L. S. (1986). Composing written sentences. *Research in the Teaching of English, 20*(2), 121-140. http://www.jstor.org/stable/40171073

Kent, S. C., & 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

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

Kraft, S., Thurfjell, F., Rack, J., & Wengelin, Å. (2019). Lexikala analyser av muntlig, tangentbordsskriven och dikterad text producerad av barn med stavningssvårigheter. *Nordic Journal of Literacy Research*, *5*(3), 102-122. https://doi.org/10.23865/njlr.v5.1511

Kraft, S. (2023). Revisions in written composition: Introducing speech-to-text to children with reading and writing difficulties. *Frontiers in Education, 8*(1133930), 1-17. https://doi.org/10.3389/feduc.2023.1133930

Leijten, M. (2007). *Writing and speech recognition: Observing error correction strategies of professional writers*. Netherlands Graduate School of Linguistics.

Lindgren, E., Westum, A., Outakoski, H., & Sullivan, K. P. (2019). Revising at the leading edge: Shaping ideas or clearing up noise. In *Observing writing* (pp. 346–365). Brill. https://doi.org/10.1163/9789004392526_017

Lu, X., Li, S., & Fujimoto, M. (2020). Automatic speech recognition. In Y. Kidawara, E. Sumita, & H. Kawai (Eds.) *Speech-to-speech translation* (pp. 21–38). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-15-0595-9_2

MacArthur, C. A. (2009). Reflections on research on writing and technology for struggling writers. *Learning Disabilities Research & Practice*, *24*(2), 93–103. https://doi.org/10.1111/j.1540-5826.2009.00283.x

MacArthur, C. A., & Cavalier, A. R. (2004). Dictation and speech recognition technology as test accommodations. *Exceptional Children*, *71*(1), 43-58. https://doi.org/10.1177/001440290407100103

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

Mortimore, T., & Crozier, W. R. (2006). Dyslexia and difficulties with study skills in higher education. *Studies in higher education*, *31*(2), 235–251. https://doi.org/10.1080/03075070600572173

Olive, T., & Cislaru, G. (2015). Linguistic forms at the process-product interface: Analysing the linguistic content of bursts of production. In G. Cislaru (Ed.) *Writing(s) at the Crossroads: The process-product interface.* (pp. 99–124). John Benjamins Publishing Company. https://doi.org/10.1075/z.194.06oli

Quinlan, T. (2004). Speech recognition technology and students with writing difficulties: Improving fluency. *Journal of Educational Psychology*, *96*(2), 337. https://doi.org/10.1037/0022-0663.96.2.337

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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*, 1–17. https://doi.org/10.1007/s00426-021-01625-z

Skar, G. B., Lei, P.-W., Graham, S., Aasen, A. J., Johansen, M. B., & Kvistad, A. H. (2022). Handwriting fluency and the quality of primary grade students' writing. *Read Writ*, *35*, 509–538. https://doi.org/10.1007/s11145-021-10185-y

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407. https://doi.org/10.1177/0022057409189001-204

Sumner, E., Connelly, V., & Barnett, A. L. (2013). Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution. *Reading and Writing*, *26*(6), 991–1008. https://doi.org/10.1007/s11145-012-9403-6

Sumner, E. and Connelly, V. (2020). Writing and revision strategies of students with and without dyslexia. *Journal of Learning Disabilities 53*, 189–198. https://doi.org/10.1177/0022219419899090

Svensson, I., Nordström, T., Lindeblad, E., Gustafson, S., Björn, M. et al. (2021) Effects of assistive technology for students with reading and writing disabilities. *Disability and Rehabilitation: Assistive Technology, 16*(2): 196–208. https://doi.org/10.1080/17483107.2019.1646821

TechSmith Corporation (2018). *Camtasia* (Version 3.1.6) [Computer software]. Copyright 2006–2018 TechSmith Corporation.

Torrance, M., Rønneberg, V., Johansson, C., & Uppstad, P. H. (2016). Adolescent Weak Decoders Writing in a Shallow Orthography: Process and Product. *Scientific Studies of Reading, 20*(5), 375–388. https://doi.org/10.1080/10888438.2016.1205071

Wengelin, Å. (2007). The word-level focus in text production by adults with reading and writing difficulties. In Rijlaarsdam, G. (Series Ed.) and M. Torrance, L. van Waes & D. Galbraith (Volume Eds.). *Writing and Cognition: Research and Applications* (Studies in writing, Vol. 20, pp. 67–82). Amsterdam: Elsevier.

Wengelin, Å., Johansson, R., & Johansson, V. (2014). Expressive Writing in Swedish 15-Year-Olds with Reading and Writing Difficulties. In Arfé, B., Dockrell, J. & Berninger, V. (Eds.) *Writing development in children with hearing loss, dyslexia or oral language problems: Implications for assessment and instruction* (pp. 244–256). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199827282.003.0018