



The contributions of executive functions, transcription skills and text-specific skills to text quality in narratives

Julie Philippek¹ · Rebecca Maria Kreutz¹ · Ann-Kathrin Hennes¹ ·
Barbara Maria Schmidt¹ · Alfred Schabmann¹

Accepted: 26 February 2024
© The Author(s) 2024

Abstract

To ensure adequate writing support for children, a profound understanding of the subskills of text quality is essential. Writing theories have already helped to better understand the contribution of different subskills to text quality, but empirical work is often limited to more general low-level transcription skills like handwriting fluency and spelling. Skills that are particularly important for composing a functional text, while theoretically seen as important, are only studied in isolation. This study combines subskills at different hierarchical levels of composition. Executive functions, handwriting fluency and spelling were modeled together with text-specific skills (lexically diverse and appropriate word usage and cohesion), text length and text quality in secondary school students' narratives. The results showed that executive functions, spelling and handwriting fluency had indirect effects on text quality, mediated by text-specific skills. Furthermore, the text-specific skills accounted for most of the explained variance in text quality over and above text length. Thus, it is clear from this study that, in addition to the frequently reported influence of transcription skills, it is text-specific skills that are most relevant for text quality.

Keywords Not-so-simple view of writing · Transcription skills · Cohesion · Lexical diversity · Text quality

Julie Philippek and Rebecca Maria Kreutz have contributed equally to this work and share first authorship.

✉ Rebecca Maria Kreutz
rebecca.kreutz@uni-koeln.de

¹ Department of Special Education and Rehabilitation, Faculty of Human Sciences, University of Cologne, Klosterstraße 79b, 50931 Cologne, Germany

Introduction

Composing competence (cf. Baer et al., 1995; Hennes et al., 2022) is an important predictor of educational and professional success and represents a key competence for social participation (Crossley & McNamara, 2016; Feenstra, 2021). It is a complex construct that entails a broad set of skills to be used depending on the demands of the writing task (Hayes, 2012). In both education and research, composing competence is usually operationalized via text quality (Feenstra, 2021). High text quality is achieved when the writer pays attention to the addressee's needs, generates, selects and organizes ideas, is familiar with appropriate text patterns and creates a coherent text that fulfils its intended function (Harsch et al., 2007; Hennes et al., 2018). To produce a functional text, the writer must meet these different demands in a way that is as goal-oriented as possible and meets the demands of the genre (Beers & Nagy, 2009; Hayes, 2012). This is a challenging cognitive process that is captured in theoretical writing process models (e.g. Hayes, 2012; Hayes & Flower, 1980). However, the particular subskills necessary for creating a functional text are not further defined in these models (Hennes, 2020). Research on subskills focuses on cognitive skills and transcription skills like handwriting fluency and spelling (Abbott et al., 2010; Graham et al., 1997; Kent & Wanzek, 2016; Salas & Silvente, 2020), but mastering these low-level transcription skills is not enough to write a functional text. Therefore, higher-level subskills are also needed, with text-specific vocabulary and cohesion discussed in the literature (Crossley & McNamara, 2016; Gómez Vera et al., 2016; Mathiebe, 2019; McNamara et al., 2010). Despite their confirmed relevance for text quality, the common influence of low- and higher-level subskills has so far been disregarded. Based on this, we conducted the present study to examine the joint influence and interactions of subskills on different hierarchical levels relevant to text quality.

The Not-So-Simple View of Writing model

Models specifying the particular subskills needed to produce a functional text and how they work together are rare. A prototypical model that considers the interaction of different subskills is the Not-So-Simple View of Writing model (Berninger & Winn, 2006). According to this model, text generation depends on the functioning of two elements, namely transcription and executive functions (EF). Transcription is composed of handwriting fluency (i.e., automatization of the motoric writing component) and spelling. EF enable writers to stay on task and switch between mental states (Drijbooms et al., 2015). Text generation itself contains “both idea generation and translation of those ideas into language representations” (Berninger et al., 2002, p. 292). Text generation, transcription and EF are supported by memory functions, which include working memory as well as long-term memory. Memory (temporarily) stores text representations, compares them to the previously produced text, and updates the text (Olive, 2012). The model further assumes that in novice writers, many of the limited cognitive resources of working memory and EF are taken up by low-level transcription skills. At this stage in development, text quality is hampered by non-automatized transcription skills. When these skills are automatized, writers

can make use of the released cognitive resources for higher-level text generation skills (Berninger & Winn, 2006).

Empirical data on the Not-So-Simple View of Writing model

Different aspects of the Not-So-Simple View of Writing model and their influence on text quality have been empirically tested in numerous studies, where there is no consensus on the operationalisation of these subskills (for a detailed overview of the measurement of subskills, see the [appendix](#); text quality is measured via holistic ratings in all considered studies).

Transcription skills and text quality: In their meta-analysis, Kent and Wanzek (2016) found positive correlations of handwriting fluency and spelling with text quality. Using structural equation modeling, Graham et al. (1997) found that handwriting fluency could explain variance in text quality of narratives from 1st to 6th grade. In their study, spelling had an indirect effect via handwriting fluency. A consistent direct effect of spelling on written expression (measured by word fluency, sentence combining and paragraph writing of narratives and expository) from 1st to 7th grade was shown by Abbott et al. (2010) in their longitudinal structural equation model. Handwriting fluency also contributed to written expression from 3rd to 4th grade.

Cognitive components and text quality: Berninger et al. (2010) showed that working memory accounted for variance in handwriting fluency, spelling and written expression (cf. Abbott et al., 2010) in 2nd and 4th grade and in spelling in 6th grade. Moreover, Connelly et al. (2012) found that working memory could explain 12%, spelling 18% and handwriting fluency 3% of the variance in text quality in narratives in 11-year-old students. These results were confirmed for 2nd graders by Cordeiro et al. (2020), who found that working memory, EF and transcription skills made significant contributions to explaining the variance in text quality in narratives.

Interactions: In some studies, interactions between the components of the Not-So-Simple View of Writing model were also considered. Salas and Silvente (2020) found a direct effect of inhibition (in 2nd, 4th and 8th grade) and of working memory (in 2nd grade) on text generation (number of words, words per clause) of opinion essays and narratives. Furthermore, they found indirect effects of working memory and inhibition via transcription skills on text generation (in 2nd, 4th and 8th grade), while working memory also affected text generation indirectly via spelling (in 8th grade). Moreover, they found indirect effects of inhibition on text generation via handwriting fluency (in 4th and 8th grades) and via spelling (in 8th grade). Drijbooms et al. (2015) included transcription, EF and written narratives (text length, syntactic complexity and story content) in their model. They found that EF indirectly affected all written narrative measures and were mainly mediated by handwriting fluency. Handwriting fluency had a direct effect on all written narrative measures, while spelling influenced syntactic complexity only.

Text-specific skills

In the Not-So-Simple View of Writing model, text-specific skills might be necessary to accomplish tasks that lead to text generation. However, the term text generation is

only vaguely defined within the model itself, and it remains unclear which subskills fall under this concept. Accordingly, measurement of the construct appears somewhat ambiguous (Kim & Graham, 2022). While Salas and Silvente (2020) measured text generation with number of words and words per clause, Limpo and Alves (2013) used text quality to measure text generation. Other studies used different oral language skills (Kim & Graham, 2022; Oddsdóttir et al., 2021), although oral language skills and text generation skills cannot be considered equivalent (Arfé & Pizzocaro, 2016). For this reason, it is necessary to define the text-specific skills that are required to accomplish the text generation process in the Not-So-Simple View of Writing model. Two main subskills are described in the literature, namely the usage of a diverse and appropriate text-specific vocabulary and the ability to establish cohesion.

Text-specific vocabulary: Using a diverse and appropriate vocabulary is required to formulate good texts of different genres (Steinhoff, 2009) and can be considered differently: There is a distinction between written and oral vocabulary as well as between context-independent vocabulary, which is determined by specific tests, and context-dependent vocabulary, which can be determined in a text (e.g. Gómez Vera et al., 2016; Kim & Graham, 2022). When writing functional texts, writers draw especially on text-specific vocabulary, which is often captured through the measure of lexical diversity in the text (for a detailed overview of the measurement of lexical diversity in considered studies, see the [appendix](#)). Lexical diversity, which is defined as the active vocabulary used in a given piece of writing (Koizumi & In'nami, 2012), is a reliable predictor of text quality especially in narratives (Hiebert & Cervetti, 2011). For example, Gómez Vera and colleagues (2016) found that lexical diversity was a significant predictor of text quality in narratives written by 4th graders. Cameron et al. (1995) verified the role of lexical diversity for text quality in narratives written by nine-year old students. Together with complexity of utterances and text length, lexical diversity explained 16% of the variance in text quality. Moreover, Olinghouse and Wilson (2013) revealed that 8.4% of the variance in text quality in narratives could be explained by lexical diversity in grade 5. McNamara and colleagues (2010) found that text quality is also higher in argumentative essays written by university students when they are more lexically diverse.

However, a quantitative measure of lexical diversity alone cannot indicate whether these words are used correctly (Mathiebe, 2019). Thus, in addition to the use of a broad text-specific vocabulary, the selection of appropriate words in the textual context might also contribute to readers' comprehension of texts. Therefore, an additional measure is needed that assesses the appropriateness of word usage in the textual context. The predictive role of this measure has not yet been sufficiently investigated (Becker-Mrotzek et al., 2014). Initial results from German-speaking studies show that students in higher grades use more appropriate words in the textual context (Mathiebe, 2019).

Cohesion: Coherence structures and connects thoughts and ideas within a text, making it comprehensible to the reader (Becker-Mrotzek et al., 2014). Cohesion is the measurable part of coherence on the surface level of the text. By connections through connectors, repetitions or references, internal grammatical and semantic consistency is given to the text (Schwarz, 2001; Zifonun, 2008). The relevance of cohesion to text quality has been confirmed in just a few studies so far and moreover,

measuring cohesion in the text is difficult, which is why there are different ways to measure cohesion (for a detailed overview of the measurement of cohesion in considered studies, see the [appendix](#)).

Cox et al. (1990) found significant correlations between text quality and the frequency of appropriate cohesive ties in narratives by 3rd and 5th graders. Cameron et al. (1995) also examined a factor consisting of several cohesion measures in narratives. They found that the cohesion factor explained the largest proportion of variance in text quality for nine-year-old students, with lexical diversity, complexity of utterances and text length also considered in the model. MacArthur et al. (2019) modeled the influence of text length, writing-specific skills (lexical and syntactic complexity), referential cohesion and connectives on the text quality of argumentative essays by university students. They found that text length, measures of writing-specific skills and referential cohesion accounted for 48.7% of the variance in text quality, with referential cohesion alone explaining 8% of the variance. These results support findings by Crossley and McNamara (2010), that cohesive elements (as rated by experts) in persuasive essays correlated positively with text quality. In addition, Crossley and McNamara (2016) found that persuasive essays by students in which experts revised the essays in order to increase cohesion were rated better than those where cohesion was not increased.

Current study

The studies cited above show that both writers’ broad and appropriate usage of text-specific vocabulary and their ability to establish cohesion can explain variance in text quality. However, these subskills have received little attention in current theoretical writing models and are rarely considered together with transcription skills (Arfé & Pizzocaro, 2016; Oddsdóttir et al., 2021). Therefore, in this study we considered a working model – which we termed the Cascaded Model of Writing (CASMOW; Fig. 1) – including EF, low-level transcription skills, higher-level text-specific skills and text quality.

In arranging these subskills, we referred on the hierarchical relations hypothesis from the expanded Direct and Indirect Effects Model of Writing (DIEW) by Kim and Graham (2022), which states that not all subskills are directly related to text quality and that low-level transcription skills are related to text quality via higher-level subskills. Specifically, DIEW assumes that transcription skills have an influence on writ-

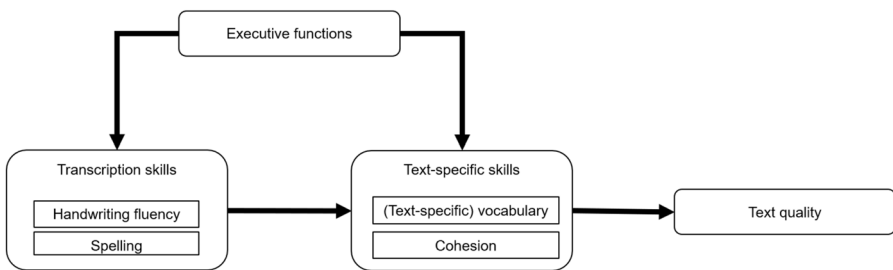


Fig. 1 The Cascaded Model of Writing

ten composition among others via discourse oral language. Therefore, in CASMOW, we hypothesized that the primary cascade goes from low-level transcription skills via higher-level text-specific skills to text quality, because the usage of appropriate and diverse text-specific vocabulary and the ability to produce local semantic and grammatical cohesion predict text quality. If a writer fails to use words appropriately or a text contains corresponding errors (e.g. false connections or the wrong tense), it is hard for the reader to establish a correct representation of the text in their mind (Erberich, 2022; Goblirsch, 2017). Of course, using appropriate words and cohesive ties requires the ability to write down words fluently and orthographically correctly. When spelling skills are insufficient, word omissions or incorrect words might be used because children avoid using words they do not know how to spell (Berninger et al., 2008).

In contrast to DIEW, however, CASMOW did not include oral vocabulary, but rather context-dependent text-specific vocabulary. This adaptation was necessary because although oral vocabulary clearly is a predictor of text quality, a broad and appropriate text-specific vocabulary requires different knowledge, is more demanding than an active vocabulary test and is therefore more suitable for 5th – 7th graders (Mathiebe, 2019).

Like DIEW, we also incorporated EF in CASMOW as fundamental for all kinds of skills necessary to write a functional text. Since the term EF covers a wide range of cognitive skills, in this study, we focused on the three most commonly postulated EF: shifting, inhibition and updating, because these three functions are well circumscribed and can therefore be operationalised quite precisely (Miyake et al., 2000). These EF were also used in DIEW, but supplemented by attentional control. Inhibition affects handwriting fluency and spelling by, for example, inhibiting other letters, motor movements or incorrect spelling patterns (Salas & Silvente, 2020); updating influences spelling by keeping a phonological form active until an orthographic rule is applied (Berninger & Richards, 2010) and shifting enables the writer to switch between different writing tasks (Olive, 2012). Moreover, these lower-level EF form the basis of higher-level EF such as planning, revising and monitoring (Berninger & Richards, 2010).

In the literature, it is assumed that writing functional texts is a general competence that manifests itself independently of genre. However, to write texts of different genres, the writer must have knowledge of concrete features of specific text types (Kim & Graham, 2022). Therefore, the relative contributions of each subskill to text quality may vary depending on the genre. In children's everyday school life, writing narratives is a common task and one of the most prominent genres in writing curricula for early secondary education (Dockrell et al., 2015; MSB NRW, 2019). In this study, the model is therefore tested with narratives.

Research questions

Based on these considerations, two research questions emerged:

1. To what extent can text quality in narratives be predicted by the Cascaded Model of Writing?

2. How do the subskills of the Cascaded Model of Writing interact with each other?

To answer the research questions, in a first step, we tested the model and all its direct and indirect effects using structural equation modeling. In a second step, we added text length to our model. Text length, although theoretically not an element of text quality per se, is associated with text quality ratings, i.e. longer texts tend to be rated better (Fleckenstein et al., 2020; MacArthur et al., 2019). Therefore, we wanted to measure the effect of all predictors of CASMOW on text length and the effect of text length on text quality.

Method

Participants

186 native German speakers between 9 and 14 years old ($M(\text{age})=11.75$; $SD=1.05$; 89 girls) participated, with 57 fifth graders, 63 sixth graders and 66 seventh graders from three classes each. The study was conducted at two randomly chosen secondary schools at which the general qualification for university entrance could be completed (academic-track school: 34.4%; comprehensive school¹: 65.6%). These schools in middle-class urban regions in North Rhine-Westphalia were as homogeneous as possible in terms of socio-economic conditions. Controlling for students' socio-economic status was not possible for data protection reasons.

The data was collected completely anonymously and could not be linked back to individual students.

Measures

A good writing task should be free of subject-related knowledge, culturally neutral and stimulating enough for students to write longer texts on it. It should also be appropriate to the students' level of writing development (Jost, 2022). These criteria resulted in the writing task for secondary school students: "What if you could fly? Think of a story about that and write it down". The students had 15 min to write a narrative based on the given phrase "If I could fly, ...".

The children's written compositions were used to determine text quality, text length and the text-specific skills. In comparison to related models such as DIEW, in which all subskills were measured with specific tests, in CASMOW text-specific skills were measured in the text itself, which is in line with Cameron's et al. (1995) statement that it is relevant to consider the retrieval of these skills in the textual context. Furthermore, it is also closest to what is required and happens in the classroom (Mathiebe, 2019). On the contrary, transcription skills are more general writing skills (Sturm, 2018) and thus should be measured independent of context with standardized tests. The main advantage of this measure is that it is an empirically validated

¹ Comprehensive secondary school encompassing all possible tracks.

assessment and ensures better comparability, since all children had to write down the same items.

Executive functions

EF were assessed using the Star Counting Test (SCT; de Jong & Das-Smaal, 1990). The test consists of six items with a pattern of stars with plus and minus signs between them. The children were asked to count the stars, starting from a given number. Depending on the sign between the stars, they had to count forwards or backwards so that one ongoing process has to be inhibited and another activated. This task imposes demands on the three basic EF of inhibition, updating and shifting. One point was given for each correct answer with a maximum of six points possible. Higher scores indicate better performance. The test's reliability (Cronbach's $\alpha=0.83-0.88$ depending on the test version) and convergent validity are satisfactory (de Jong & Das-Smaal, 1990).

Handwriting fluency

The alphabet task from the Detailed Assessment of Speed of Handwriting (DASH; Barnett et al., 2007) was conducted to assess handwriting fluency. Children were asked to write the alphabet in lowercase letters as often and legibly as possible for 60 s. The number of correct letters corresponded to the final score, so that higher scores indicate better handwriting fluency. A letter was scored as correct if it was legible and in the correct alphabetical order. Substitutions, transpositions, additions and omissions were considered errors. The interrater reliability (ICC=0.99) and the convergent validity of this subtest of the DASH are satisfactory (Barnett et al., 2009).

Spelling

To assess children's spelling ability, the standardised spelling test Hamburger Schreibprobe [Hamburger Spelling Test] (HSP; May et al., 2018) was used. Words or sentences requiring knowledge of alphabetic, orthographic and morphemic rules were dictated to the children. The number of correctly spelled words was used as the spelling score, so that higher scores indicate better spelling. The HSP has a reliability of Cronbach's $\alpha=0.94$ at the word level and a satisfactory convergent validity (May et al., 2018).

Cohesion

In this study, similar to McNamara et al. (2010), ratings were carried out by two raters with expertise in linguistics. In this rating, various grammatical and semantic aspects within the children's texts that interrupted the reader's process of understanding were assessed. Three cohesion measures were identified and included in the model: Local cohesion errors (lack of references or connection of elements that do not belong together), sentence errors (incorrect syntax caused by missing words or incorrect word order) and tense errors (inappropriate change between tenses). The interrater

reliability between the raters was calculated with Cohen's Kappa. This was $\kappa=0.87$ ($p<.001$) for local cohesion errors, $\kappa=0.89$ ($p<.001$) for sentence errors and $\kappa=0.92$ ($p<.001$) for tense errors. In case of disagreement between the raters, the case was discussed with two further experts until a consensus was reached. Since text length is related to text quality, many linguistic indices also correlate with it. Therefore, the measure of cohesion needs to control for this problem, which is why quotients were formed (MacArthur et al., 2019): Local cohesion errors were divided by the number of propositions, sentence errors by the number of sentences, and tense errors by the number of verbs. For simpler interpretation, scores were inverted so that higher scores correspond to more cohesive texts.

Text-specific vocabulary

There are several indicators related to the characteristics and quantity of text-specific vocabulary. Of particular relevance here is lexical diversity, which can be calculated by the ratio of types to tokens in a text. It is an objective indicator of the amount of vocabulary available to the author. There are various measures of lexical diversity, such as the type-token ratio (TTR), which divides all the types in the text by all the tokens, and the Guiraud index, which uses the square root of the number of tokens as the denominator. However, text length always has an influence on these measures (Koizumi & In'nami, 2012).

Therefore, in this study, the Measure of Textual Lexical Diversity (MTLD), which is found to be least affected by text length (Koizumi & In'nami, 2012), was calculated. The MTLD represents "the mean length of sequential word strings in a text that maintain a given TTR value" (McCarthy & Jarvis, 2010, p. 384). McCarthy and Jarvis (2010) showed that the TTR curves trend to reach a stabilization point at around 0.72. It is counted how many times the text reaches this TTR value starting at the beginning of the text and continuing to the end of the text. Then, the mean word count is calculated. To do so, the number of tokens is counted and divided by the number of times the text reaches the specified TTR value. Once this first cycle is complete and an initial MTLD score has been calculated, the entire text is analysed again in reverse order. This re-analysis results in another MTLD score. The final MTLD score is the value obtained by taking the mean of the forward and the reverse MTLD scores. This measure correlates highly with other measures of lexical diversity and therefore has satisfactory convergent validity. Moreover, MTLD is reliable, in that shorter sections of a text have similar MTLD scores to those of the whole text and the MTLD scores of these text sections do not correlate with text length (McCarthy, 2005; McCarthy & Jarvis, 2010).

In this study, MTLD scores were calculated in RStudio (RStudioTeam, 2020). Higher scores indicate greater lexical diversity (McCarthy & Jarvis, 2010).

Also relevant for the reader's comprehension process is the writer's selection of appropriate words (Becker-Mrotzek et al., 2014). To measure appropriate word usage, words in the text that did not fit the context were counted as word errors, applying the same ratings used to determine cohesion. In this case, the interrater reliability was $\kappa=0.88$ ($p<.001$), disagreements were discussed again and a consensus was reached. Since this measure is influenced by text length, the number of word

errors was divided by the total number of words, resulting in the variable “appropriate words”. Higher scores correspond to a more appropriate usage of words.

Text length

Text length was determined by the number of words written.

Text quality

To evaluate text quality, the texts were typed up and spelling mistakes were corrected. The most frequently used method of measuring text quality are holistic ratings. However, even trained raters, such as teachers, seem to have difficulties giving holistic ratings to texts (Hennes et al., 2022). Higher reliability and consistency can be achieved by directly comparing texts, which is why the Comparative Judgement method was chosen for the present study. In this method, texts are randomly paired off and raters must decide which of the two texts is the better. The texts are then ranked on a scale from worst to best (Lesterhuis et al., 2017). To achieve this ranking, a logit score per text is determined using a logistic model. This score indicates the probability of winning a comparison with a reference text and can be used as a text quality parameter (Pollitt, 2012). Higher scores indicate better text quality.

The reliability of the estimated logit scores can be determined using the scale separation reliability (analogous to Cronbach’s alpha) (Jones & Karadeniz, 2016). In order to achieve a satisfactory reliability of 0.7, each text must be compared 10 to 14 times on average, while good values for convergent validity are achieved with 15 or more comparisons per text. With this number of comparisons, both experts and naïve raters can make reliable and valid assessments (Bouwer et al., 2023; Verhavert et al., 2019).

In this study, comparative judgements were conducted using the online comparing tool Comproved (www.comproved.com). 65 independent naïve raters were asked to complete 15 holistic, pairwise comparisons of the children’s texts, in which they had to choose the better text in each case by mouse click. They received no specific training and performed the comparisons at their own pace. Thus, each text was judged at least 79 times, resulting in a total of 961 comparisons. The scale separation reliability yielded in a mean score of 0.73 and due to the high number of comparisons, the conditions for satisfactory convergent validity were also given.

Procedure

Data collection took place in class on two days for 45 min each day. The first session included the SCT (de Jong & Das-Smaal, 1990) and the writing task. In the second session, the alphabet task (Barnett et al., 2007) and the HSP (May et al., 2018) were conducted. The order of administration remained the same across all grades. All tasks were carried out using paper and pencil, as this is the most common modality in the German school system.

Statistics

All effects described in CASMOW were transferred into a structural equation model (model 1A). This included direct paths from all variables to text quality, direct paths from EF to transcription and text-specific skills and direct paths from transcription to text-specific skills. As we had no theoretical assumption about causal effects of spelling and handwriting fluency, we did not specify a path from one to the other (but estimated correlations). All corresponding indirect paths were also examined. The model fit was evaluated by Chi²-test, the Root-Mean-Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI) and the Bentler-Bonett Normed Fit Index (NFI). After fitting model 1A, we omitted non-significant paths and reran the analysis (model 1B). We then added text length to the model (model 2A), omitted non-significant paths (model 2B) and variables with no significant paths (model 2C).

Results

Descriptive statistics

Children's mean performance on the different tasks were evaluated and the variables were tested for normal distribution using the Kolmogorov-Smirnov Test. The variables handwriting fluency and text quality were normally distributed ($p > .05$). The non-normal distribution of the other variables was counteracted by using bootstrapping in the structural equation model. Descriptive statistics for all variables are shown in Table 1. The correlations presented in Table 2 show that all of the variables except for correct tense were moderately correlated with text quality and text length and partially correlated with each other.

Modeling the influence of subskills on text quality

When specifying all theoretically sensible effects, model 1A demonstrated good fit to the data (Chi² (10)=12.62, $p = .246$, RMSEA=0.038, CFI=0.989, NFI=0.952).

Table 1 Descriptive statistics

Measures	Mean	Standard deviation	Range
Executive functions	2.64	1.48	0–6 (6) ^a
Handwriting fluency	63.25	23.00	8–138
Spelling	36.94	8.24	13–49 (49) ^a
Lexical diversity	58.17	23.30	15.85–153.00
Appropriate words	0.99	0.02	0.91–1.00
Local cohesion	0.89	0.12	0.38–1.00
Correct tense	0.89	0.13	0.46–1.00
Correct sentences	0.90	0.16	0.00–1.00
Text length	103.4	57.2	18–369
Text quality	0.06	1.34	–3.84–2.99

^aTheoretical maximum

Table 2 Correlations of variables

	1	2	3	4	5	6	7	8	9
1. Executive functions									
2. Handwriting fluency	0.27**								
3. Spelling	0.34**	0.41**							
4. Lexical diversity	0.08	0.25**	0.34**						
5. Appropriate words	0.15*	0.01	0.07	0.18*					
6. Local cohesion	0.09	0.00	0.17*	0.19**	0.06				
7. Correct tense	0.01	-0.03	-0.05	-0.07	-0.01	0.02			
8. Correct sentences	0.18*	0.04	0.33**	0.05	-0.01	0.03	-0.06		
9. Text length	0.29**	0.47**	0.44**	0.48**	0.15*	0.22**	-0.01	0.21**	
10. Text quality	0.20**	0.30**	0.44**	0.53**	0.20**	0.28**	0.09	0.36**	0.60**

* $p < .05$; ** $p < .01$

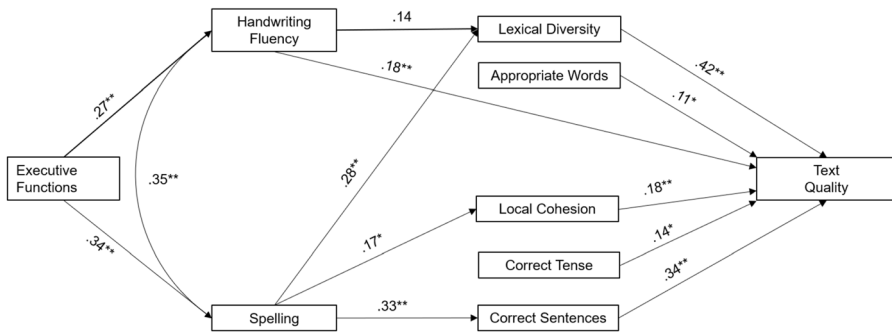


Fig. 2 Structural equation model of the Cascaded Model of Writing (1B). Standardized path coefficients of the interactions between executive functions, transcription skills, text-specific skills and text quality. * $p < .05$; ** $p < .01$

When omitting non-significant paths, the more parsimonious model (1B; Fig. 2) appeared to fit equally well ($\chi^2(23) = 27.17, p = .249, RMSEA = 0.031, CFI = 0.982, NFI = 0.897$; model comparison: $\Delta \chi^2(13) = 14.559, p = .336$).

In this model, significant direct effects could be found from all text-specific skills and from handwriting fluency to text quality. Moreover, there was an indirect effect from handwriting fluency to text quality via lexical diversity ($\beta = 0.06; p = .048$); thus, handwriting fluency had a total effect of $\beta = 0.24 (p = .002)$ on text quality. Spelling influenced text quality only indirectly via lexical diversity, local cohesion and correct sentences (total effect: $\beta = 0.26; p = .002$), and EF influenced text quality indirectly as well, resulting in a total effect of $\beta = 0.15 (p = .002)$. Additionally, there was a significant correlation between handwriting fluency and spelling ($r = .35; p < .001$). Overall, model 1B could explain 49.4% of the variance in text quality.

Comparing model 1B and model 2B (with variable text length and removal of non-significant paths) revealed no significant difference ($\Delta \chi^2(7) = 13.24, p = .066$), and the more parsimonious model (2B) fit the data well ($\chi^2(30) = 38.72, p = .132, RMSEA = 0.040, CFI = 0.975, NFI = 0.900$). Because appropriate word usage had

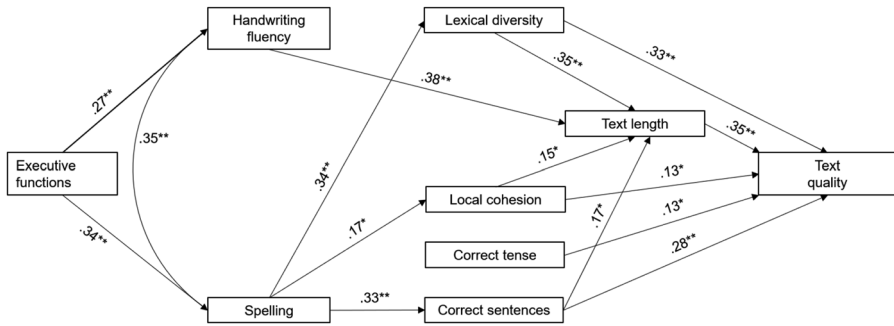


Fig. 3 Structural equation model of the Cascaded Model of Writing (2C). Standardized path coefficients of the interactions between executive functions, transcription skills, text-specific skills, text length and text quality. * $p < .05$; ** $p < .01$

no significant effect of any kind in this model, the model was rerun omitting this variable, which resulted in model 2C, depicted in Fig. 3 ($\chi^2(21) = 23.58, p = .314, RMSEA = 0.026, CFI = 0.992, NFI = 0.937$).

In this model, significant direct effects to text quality were again found from all text-specific skills and additionally from text length. The direct path from handwriting fluency to text quality from model 1B was now mediated by text length. Furthermore, lexical diversity, local cohesion and correct sentences showed both direct effects on text quality and indirect effects via text length. In sum, model 2C could explain 54% of the variance in text quality.

Discussion

In the present study, we examined the influence of subskills at different hierarchical levels on text quality of narratives in 5th – 7th graders. We theorised causal pathways for EF, low-level transcription skills and text-specific skills. Although CASMOW is based on the Not-So-Simple View of Writing model, it goes beyond it by specifying – at least for narratives – what subskills are needed for text generation (which is only discussed as a more general entity within the Not-So-Simple View of Writing model).

In general, the two final models presented (1B and 2C) confirm our assumptions.

In model 1B, 49.4% of the variance and in model 2C 54% of the variance in secondary school students’ text quality could be explained by the components of CASMOW. The models show that all considered subskills are relevant for text quality in narratives and interact with each other.

When considering both hierarchical levels of writing skills, the findings are in line with the assumption of a cascade from transcription skills via text-specific skills to text quality. In particular, in model 1B, all text-specific skills have a direct influence on text quality, which is consistent with previous studies that have confirmed the relevance of these subskills for text quality (Cameron et al., 1995; Cox et al., 1990; Gómez Vera et al., 2016; MacArthur et al., 2019; McNamara et al., 2010; Olinghouse & Wilson, 2013), whereas the effects of spelling on text quality are mediated by text-specific skills. These results are in line with the hierarchical relations hypothesis by

Kim and Graham (2022), who showed that not all subskills are directly related to text quality and that low-level subskills are related to text quality via higher-level subskills, even though text generation is operationalized in terms of text-specific skills in this study.

In model 1B, there is one exception that does not fit our assumption or the hierarchical relations hypothesis and is difficult to explain, namely the direct effect of handwriting fluency on text quality. One explanation of this result is that poor handwriting fluency influences text quality over and above the text-specific skills considered in this study. However, model 2C clearly shows that the text quality rating might be influenced by text length (Fleckenstein et al., 2020; MacArthur et al., 2019), which in turn at least partially a function of handwriting fluency, was correct.

The change from a direct to an indirect effect of handwriting fluency on text quality does not show up in the text-specific skills when adding text length (model 2C). Rather, the text-specific skills continue to have a direct effect on text quality independently of each other. This means that text-specific skills – in contrast to handwriting fluency – contribute to text quality beyond the effects of text length in this age group, which is in line with other research examining students with automatized handwriting fluency, showing that higher-level skills could be more relevant at this stage of development (MacArthur et al., 2019).

The largest direct effect and the highest correlation of text-specific skills can be seen between lexical diversity and text quality. This confirms other recent studies (e.g. Gómez Vera et al., 2016; Olinghouse & Wilson, 2013) and furthermore demonstrates that lexical diversity is the strongest predictor of text quality compared to other subskills when considered together.

Nevertheless, there are also indirect effects of some text-specific skills on text quality via text length. Thus, it seems that children who write more lexically diverse and cohesive texts – and thus are more proficient with language – are able to write longer texts, which are then often rated better (Crossley et al., 2014).

In accordance with our hypothesis, following the cascade, spelling has an indirect effect on text quality via lexical diversity, correct sentences and local cohesion. This could be due to students' avoidance of words they cannot spell, which could lead to less lexically diverse texts, syntax errors (due to word omissions) and cohesion errors (Berninger et al., 2008). Poor spelling thus seems to mainly inhibit text-specific skills. In contrast to spelling, the assumed cascade from handwriting fluency via text-specific skills to text quality was not observed. This could be due to the fact that in the present sample of secondary school students, handwriting fluency is automatized and thus (no longer) has an inhibitory influence on text-specific skills.

In accordance with our assumptions, we found direct effects of EF on transcription skills, which is in line with previous studies depicting the relevance of inhibition and updating for handwriting fluency (Salas & Silvente, 2020) and spelling (Berninger & Richards, 2010). Moreover, we found an indirect effect on text quality, which is also in line with CASMOW and prior studies in which EF were assumed to be relevant for text quality but to influence it only via transcription skills (Salas & Silvente, 2020). Contrary to our hypothesis, we did not find a direct effect from EF to text-specific skills. A possible explanation for this is that lower- and higher-level EF contribute differentially to the various levels of subskills of CASMOW. Thus it might be that

text-specific skills are more closely related to higher-level EF like planning, revising and monitoring (Goldstein & McGoldrick, 2021; Kim & Graham, 2022), which were not considered in the present assessment.

Limitations and future research

The interpretation of the current results is constrained by some limitations that may point to possible directions for future research.

First of all, text quality and text-specific skills were measured in this study using a single writing sample. However, a single writing sample might not provide a reliable estimate of students' writing abilities (Graham et al., 2016). Therefore, it would be more informative to assess multiple writing samples from all students in the same genre.

Furthermore, the results relate only to narratives and might neither be easily transferred to other text genres or other prompts (Beers & Nagy, 2009), as different writing tasks require different skills (Kim & Graham, 2022), nor to other age groups, as composing competence develops until adolescence and subskills are relevant differently at various stages of development. In future studies, CASMOW should also be tested with respect to other genres and other age groups.

However, the most important aspect concerning future directions might be the following: The paths described in the model explain a large proportion of the variance, but since a significant amount of variance in text quality remains unexplained (46%), further subskills need to be included in the future CASMOW. We assume that these subskills are on an even higher-level than the subskills we included and are located in the model between the text-specific skills and text quality. According to Kim and Graham (2022) and Hennes (2020), on this higher level the use of suitable text patterns (text structure knowledge) and information management are necessary to produce a functional text. Furthermore, the ability to adopt the reader's perspective and to create global coherence represent further subskills on the highest level. Moreover, as lower-level EF scaffold higher-level EF like the ability to plan, revise or monitor the text while writing, these higher-level EF should be included in the future CASMOW. These skills could be measured both by standardised EF tests or writing-specific measures in the text (e.g. by measuring bursts) (Kim & Graham, 2022; Limpo & Alves, 2013).

Educational implications

The results of the current study are important for understanding the relationships between the different subskills of text quality and can be used to draw implications for writing instruction. In primary school, if addressed at all, composing instruction for beginning writers mainly contains the automatization of transcription skills (Cordeiro et al., 2020; Graham et al., 2013; Salas & Silvente, 2020). As the study results show that text-specific skills are most important for text quality, these should also be trained as early as possible in combination with low-level transcription skills. Within these interventions, explicit use of text-specific skills in the text should be practiced. This should be initiated even before the transition to secondary school. With regard

to EF, there is evidence that isolated training of individual EF does not transfer to academic skills (Melby-Lervåg & Hulme, 2013). Instead, studies show that it makes more sense to provide students with strategies, such as guiding students to focus on certain domains (e.g. spelling) at certain stages of revision, thus reducing the cognitive load for the rest of the writing process (McNamara et al., 2010; Salas & Silvente, 2020). These strategies should be combined with the comprehensive intervention programmes just described to scaffold the writing of functional texts.

Appendix

Overview of measurement in cited studies

Subskill	Measurement
Handwriting fluency	<ul style="list-style-type: none"> Standardized alphabet task (Abbott et al., 2010; Berninger et al., 2010; Connelly et al., 2012; Cordeiro et al., 2020; Graham et al., 1997; Salas & Silvente, 2020) Standardized sentence copying task (Berninger et al., 2010; Cordeiro et al., 2020; Graham et al., 1997) Standardized paragraph copying task (Berninger et al., 2010; Drijbooms et al., 2015)
Spelling	<ul style="list-style-type: none"> Standardized dictation test (Abbott et al., 2010; Berninger et al., 2010; Connelly et al., 2012; Drijbooms et al., 2015; Graham et al., 1997) Dictation task of (isolated) words (Cordeiro et al., 2020; Salas & Silvente, 2020)
Lexical diversity	<ul style="list-style-type: none"> Corrected Type-Token Ratio in the text (Gómez Vera et al., 2016) Type-Token Ratio in the text (Cameron et al., 1995) Measure of Textual Lexical Diversity in the text (McNamara et al., 2010; Olinghouse & Wilson, 2013)
Appropriate words	<ul style="list-style-type: none"> Rating (Mathiebe, 2019)
Cohesion	<ul style="list-style-type: none"> Number of appropriate cohesive ties/ number of t-units in the text (Cameron et al., 1995; Cox et al., 1990) Analysis of cohesion features in the text by a computer program (Coh-Metrix) (MacArthur et al., 2019) Tool for automatic analysis of cohesion (Crossley & McNamara, 2016) Expert rating (Crossley & McNamara, 2010)

Acknowledgements We would like to thank all the children and students who participated in the study.

Author contributions RMK and JP contributed equally to the concept of the article and performed the data analysis. The first draft of the manuscript was written by RMK and JP and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding We received no financial support for the research. Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest We have no conflict of interest to disclose.

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. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*(2), 281–298. <https://doi.org/10.1037/a0019318>.
- Arfè, B., & Pizzocaro, E. (2016). Sentence Generation in Children with and Without Problems of Written Expression. In J. Perera, M. Aparici, E. Rosado, & N. Salas (Eds.), *Written and Spoken Language Development across the Lifespan* (pp. 327–344). Springer. https://doi.org/10.1007/978-3-319-21136-7_19.
- Baer, M., Fuchs, M., Reber-Wyss, M., Jurt, U., & Nussbaum, T. (1995). Das “Orchester-Modell” der Textproduktion. In J. Baumann, & R. Weingarten (Eds.), *Schreiben: Prozesse, Prozeduren Und Produkte* (pp. 173–200). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-97050-3_9.
- Barnett, A. L., Henderson, S., Scheib, B., & Schulz, J. (2007). *The detailed Assessment of Speed of Handwriting (DASH). Manual*. Pearson Education.
- Barnett, A. L., Henderson, S. E., Scheib, B., & Schulz, J. (2009). Development and standardization of a new handwriting speed test: The detailed Assessment of Speed of Handwriting. *British Journal of Educational Psychology, 2*(6), 137–157. <https://doi.org/10.1348/000709909X241937>.
- Becker-Mrotzek, M., Grabowski, J., Jost, J., Knopp, M., & Linnemann, M. (2014). Adressatenorientierung und Kohärenzherstellung im Text. Zum Zusammenhang kognitiver und sprachlicher realisierter Teilkompetenzen von Schreibkompetenz. *Didaktik Deutsch Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur, 19*(37), 21–43. <https://doi.org/10.25656/01:17151>.
- Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing, 22*(2), 185–200. <https://doi.org/10.1007/s11145-007-9107-5>.
- Berninger, V. W., & Richards, T. (2010). Inter-relationships among behavioral markers, genes, brain and treatment in dyslexia and dysgraphia. *Future Neurology, 5*(4), 597–617. <https://doi.org/10.2217/fnl.10.22>.
- Berninger, V. W., & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 96–114). Guilford Press.
- Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*(2), 291–304. <https://doi.org/10.1037/0022-0663.94.2.291>.
- Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology, 46*(1), 1–21. <https://doi.org/10.1016/j.jsp.2006.11.008>.
- Berninger, V. W., Abbott, R. D., Swanson, H. L., Lovitt, D., Trivedi, P., Lin, S. J., Gould, L., Youngstrom, M., Shimada, S., & Amtmann, D. (2010). Relationship of Word- and Sentence-Level Working Memory to Reading and Writing in Second, Fourth, and Sixth Grade. *Language Speech and Hearing Services in Schools, 41*(2), 179–193. [https://doi.org/10.1044/0161-1461\(2009/08-0002\)](https://doi.org/10.1044/0161-1461(2009/08-0002)).
- Bouwer, R., Lesterhuis, M., de Smedt, F., van Keer, H., & de Maeyer, S. (2023). Comparative approaches to the assessment of writing: Reliability and validity of benchmark rating and comparative judgement. *Journal of Writing Research, 300*-320. <https://jowr.org/pkp/ojs/index.php/jowr/article/view/867>.
- Cameron, C. A., Lee, K., Webster, S., Munro, K., Hunt, A. K., & Linton, M. J. (1995). Text cohesion in children's narrative writing. *Applied Psycholinguistics, 16*(3), 257–269. <https://doi.org/10.1017/S0142716400007293>.

- 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>.
- Cordeiro, C., Limpo, T., Olive, T., & Castro, S. L. (2020). Do executive functions contribute to writing quality in beginning writers? A longitudinal study with second graders. *Reading and Writing, 33*(4), 813–833. <https://doi.org/10.1007/s11145-019-09963-6>.
- Cox, B. E., Shanahan, T., & Sulzby, E. (1990). Good and Poor Elementary Readers' Use of Cohesion in Writing. *Reading Research Quarterly, 25*(1), 47–65. <https://doi.org/10.2307/747987>.
- Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. *Proceedings of the Annual Meeting of the Cognitive Science Society, 32*(32), 984–989.
- Crossley, S. A., & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research, 7*(3), 351–370. <https://doi.org/10.17239/jowr-2016.07.03.02>.
- Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication, 31*(2), 184–214. <https://doi.org/10.1177/0741088314526354>.
- de Jong, P. F., & Das-Smaal, E. A. (1990). The star counting test: An attention test for children. *Personality and Individual Differences, 11*(6), 597–604. [https://doi.org/10.1016/0191-8869\(90\)90043-Q](https://doi.org/10.1016/0191-8869(90)90043-Q).
- Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children's writing products: The role of curriculum based measures. *British Educational Research Journal, 41*(4), 575–595. <https://doi.org/10.1002/berj.3162>.
- Drijbooms, E., Groen, M. A., & Verhoeven, L. (2015). The contribution of executive functions to narrative writing in fourth grade children. *Reading and Writing, 28*(7), 989–1011. <https://doi.org/10.1007/s11145-015-9558-z>.
- Erberich, M. (2022). Die Wortwahl. In M. Erberich (Ed.), *Einfach und verständlich schreiben: Techniken von Profis für Beruf und Studium* (pp. 5–15). Springer. https://doi.org/10.1007/978-3-662-66276-2_2.
- Feenstra, H. (2021). *Assessing writing ability in primary education: On the evaluation of text quality and text complexity* [Dissertation, University of Twente]. Research Information University of Twente. <https://doi.org/10.3990/1.9789036537254>.
- Fleckenstein, J., Meyer, J., Jansen, T., Keller, S., & Köller, O. (2020). Is a Long Essay Always a Good Essay? The Effect of Text Length on Writing Assessment. *Frontiers in Psychology, 11*, 562462. <https://doi.org/10.3389/fpsyg.2020.562462>.
- Goblirsch, G. (2017). Was zum Textverständnis nötig ist. In G. Goblirsch (Ed.), *Gebrauchstexte schreiben: Systemische Textmodelle für Journalismus und PR* (pp. 3–17). Springer Fachmedien. https://doi.org/10.1007/978-3-658-17601-3_2.
- Goldstein, S., & McGoldrick, K. D. (2021). The Future Role of Executive Functions in Education. In T. Limpo & T. Olive (Eds.), *Executive Functions and Writing* (First edition, pp. 288–296). Oxford University Press. <https://doi.org/10.1093/oso/9780198863564.003.0013>.
- Gómez Vera, G., Sotomayor, C., Bedwell, P., Domínguez, A. M., & Jéldrez, E. (2016). Analysis of lexical quality and its relation to writing quality for 4th grade, primary school students in Chile. *Reading and Writing, 29*(7), 1317–1336. <https://doi.org/10.1007/s11145-016-9637-9>.
- Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*(1), 170–182. <https://doi.org/10.1037/0022-0663.89.1.170>.
- Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing, 26*(1), 1–15. <https://doi.org/10.1007/s11145-012-9395-2>.
- Graham, S., Hebert, M., Paige Sandbank, M., & Harris, K. R. (2016). Assessing the Writing Achievement of Young Struggling Writers. *Learning Disability Quarterly, 39*(2), 72–82. <https://doi.org/10.1177/0731948714555019>.
- Harsch, C., Neumann, A., Lehmann, R., & Schröder, K. (2007). Schreibfähigkeit. In E. Klieme & B. Beck (Eds.), *Sprachliche Kompetenzen. Konzepte und Messungen* (pp. 42–62). Beltz. <https://doi.org/10.25656/01:3232>.
- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369–388. <https://doi.org/10.1177/0741088312451260>.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the Organization of Writing Processes. In L. Gregg & Erwin R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 3–30). Routledge.

- Hennes, A. K. (2020). *Schreibprodukte bewerten: die Rolle der Expertise bei der Bewertung der Textproduktionskompetenz* [Dissertation, Universität zu Köln]. Kölner UniversitätPublikationsServer. <http://kups.ub.uni-koeln.de/id/eprint/11414>.
- Hennes, A. K., Schmidt, B. M., Zepnik, S., Linnemann, M., Jost, J., Becker-Mrotzek, M., Rietz, C., & Schabmann, A. (2018). Schreibkompetenz diagnostizieren. Ein standardisiertes Testverfahren für die Klassenstufen 4–9 in der Entwicklung. *Empirische Sonderpädagogik*, *10*(3), 294–310.
- Hennes, A. K., Schmidt, B. M., Yanagida, T., Osipov, I., Rietz, C., & Schabmann, A. (2022). Meeting the Challenge of Assessing (Students’) Text Quality: Are There any Expert Teachers Can Learn from or Do We Face a More Fundamental Problem? *Psychological Test and Assessment Modeling*, *64*(3), 272–303.
- Hiebert, E. H., & Cervetti, G. N. (2011). What Differences in Narrative and Informational Texts Mean for the Learning and Instruction of Vocabulary. Reading Research Report #11.01. *Online Submission*. <https://eric.ed.gov/?id=ed518047>.
- Jones, I., & Karadeniz, I. (2016, August 3). *An alternative approach to assessing achievement*. Proceedings of the 2016 40th Conference of the International Group for the Psychology of Mathematics Education, Szeged, Hungary.
- Jost, J. (2022). Schreibaufgaben zur Indikation von Schreibkompetenz. In M. Becker-Mrotzek, & J. Grabowski (Eds.), *Schreibkompetenz in der Sekundarstufe: Theorie, Diagnose und Förderung* (pp. 117–132). Waxmann.
- Kent, S. C., & Wanzenk, J. (2016). The relationship between component skills and writing quality and production across developmental levels. *Review of Educational Research*, *86*(2), 570–601. <https://doi.org/10.3102/0034654315619491>.
- Kim, Y. S. G., & Graham, S. (2022). Expanding the Direct and Indirect Effects Model of Writing (DIEW): Reading-Writing Relations, and Dynamic Relations As a Function of Measurement/Dimensions of Written Composition. *Journal of Educational Psychology*, *114*(2), 215–238. <https://doi.org/10.1037/edu0000564>.
- Koizumi, R., & In’nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, *40*(4), 554–564. <https://doi.org/10.1016/j.system.2012.10.012>.
- Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & de Maeyer, S. (2017). Comparative Judgement as a Promising Alternative to Score Competences. In E. Cano & G. Ion (Eds.), *Innovative Practices for Higher Education Assessment and Measurement* (pp. 119–138). IGI Global. <https://doi.org/10.4018/978-1-5225-0531-0.ch007>.
- Limpo, T., & Alves, R. A. (2013). Modeling writing development: Contribution of transcription and self-regulation to Portuguese students’ text generation quality. *Journal of Educational Psychology*, *105*(2), 401–413. <https://doi.org/10.1037/a0031391>.
- MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing*, *32*(6), 1553–1574. <https://doi.org/10.1007/s11145-018-9853-6>.
- Mathiebe, M. (2019). Wortschatzfähigkeiten in der Sekundarstufe I: Plädoyer für eine textorientierte Perspektive. *Forschung Sprache*, *7*(3), 96–106.
- May, P., Vieluf, U., & Malitzky, V. (2018). *HSP+ : Hamburger Schreib-Probe 1–10*. Ernst Klett Verlag.
- McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (3199485) [Dissertation, The University of Memphis]. ProQuest Dissertations and Theses Global. <https://search.proquest.com/openview/860b2901fa90c6e68e46cd9111bd2d1c/1?pqorigsite=gscholarcbl=18750diss=y>.
- McCarthy, P. M., & Jarvis, S. (2010). MtlD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. <https://doi.org/10.3758/BRM.42.2.381>.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. <https://doi.org/10.1177/0741088309351547>.
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. <https://doi.org/10.1037/a0028228>.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>.
- MSB NRW (Ed.) (2019). *Schule in NRW. Kernlehrplan für Sekundarstufe I Gymnasium in Nordrhein-Westfalen*. <https://www.schulentwicklung.nrw.de/lehrplaene/>.

- Oddsdóttir, R., Ragnarsdóttir, H., & Skúlason, S. (2021). The effect of transcription skills, text generation, and self-regulation on Icelandic children's text writing. *Reading and Writing, 34*(2), 391–416. <https://doi.org/10.1007/s11145-020-10074-w>.
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26*(1), 45–65. <https://doi.org/10.1007/s11145-012-9392-5>.
- Olive, T. (2012). Writing and working memory: A summary of theories and of findings. In E. L. Grigorenko, E. Mambrino, & D. D. Preiss (Eds.), *Cognitive psychology. Writing: A mosaic of new perspectives* (pp. 125–140). Psychology Press.
- Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles Policy & Practice, 19*(3), 281–300. <https://doi.org/10.1080/0969594X.2012.665354>.
- RStudioTeam. (2020). *RStudio: Integrated development for R [Computer software]*. RStudio, PBC. <http://www.rstudio.com>.
- Salas, N., & Silvente, S. (2020). The role of executive functions and transcription skills in writing: A cross-sectional study across 7 years of schooling. *Reading and Writing, 33*(4), 877–905. <https://doi.org/10.1007/s11145-019-09979-y>.
- Schwarz, M. (2001). Establishing Coherence in Text. Conceptual Continuity and Textworld Models. *Logos and Language, 2*(1), 15–24.
- Steinhoff, T. (2009). Der Wortschatz als Schaltstelle des schulischen Spracherwerbs. *Didaktik Deutsch Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur, 14*(27), 33–52. <https://doi.org/10.25656/01:21338>.
- Sturm, A. (2018). *Empfehlungen zur Sprachförderung im Pilotprojekt ALLE*. Pädagogische Hochschule FHNW.
- Verhavert, S., Bouwer, R., Donche, V., & de Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles Policy & Practice, 26*(5), 541–562. <https://doi.org/10.1080/0969594X.2019.1602027>.
- Zifonun, G. (2008). Textkonstitutive Funktionen von Tempus, Modus und Genus Verbi. In K. Brinker (Ed.), *Text- & Gesprächslinguistik, 1. Halbband* (pp. 315–330). Walter de Gruyter. <https://doi.org/10.1515/9783110194067-035>.

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