Good Text Quality Despite Deficient Transcription Skills – An Analysis of Double Dissociation

Rebecca Maria Kreutz[1], Julie Wolf[1], Ann-Kathrin Hennes[2] & Alfred Schabmann[1]

[1]Department of Special Education and Rehabilitation, University of Cologne | Germany

[2]Department 02 Humanities and Cultural Sciences, University of Kassel | Germany

Abstract: Current writing models suppose that the production of a good-quality text is a complex interaction of low-level transcription skills (handwriting fluency and spelling) and higher-level cognitive and text-specific formulation skills (such as building cohesion and using diverse and appropriate vocabulary). Low-level transcription skills, albeit important for text quality, can only explain a small part of its variance. Therefore, in this paper we investigate whether there exist children who show discrepant achievements in text quality and low-level transcription skills and how they can be characterized in terms of higher-level text-specific formulation skills. A total of 186 native German-speaking 5th to 7th grade students had to complete transcription skills tests and write a narrative text. Text quality and text-specific formulation skills were measured. The following results were obtained: (1) Students with double dissociations (deficient transcription skills/ age-appropriate text quality or the reverse) could be identified; (2) these groups differed significantly in text-specific formulation skills, namely lexical diversity, correct sentence building, and local cohesion. Generally, children with the same low-level transcription skills (whether age-appropriate or deficient) who wrote better texts, achieved higher scores in these variables. The findings emphasize the importance of text-specific formulation skills and their consideration in writing instruction.

Keywords: spelling, handwriting fluency, text quality, text-specific formulation skills, double dissociation

journal of WRITING RESEARCH

Kreutz, R.M., Wolf, J., Hennes, A-K, & Schabmann, A. (2026). Good text quality despite deficient transcription skills – An analysis of double dissociation. *Journal of Writing Research, 18*(1), 49-73. DOI: https://doi.org/10.17239/jowr-2026.18.01.02

Contact: Rebecca Maria Kreutz, Department of Special Education and Rehabilitation, Faculty of Human Sciences, University of Cologne, Klosterstraße 79b, 50931 Cologne | Germany -rebecca.kreutz@uni-koeln.de

1. Introduction

There is a consensus that many children are not able to produce good-quality texts (Graham et al., 2023). However, composing a good-quality text is a key competence in social life, at school (Crossley & McNamara, 2016; Graham & Perin, 2007; National Commission on Writing, 2003) and at work (T. McNamara et al., 2019). Writing models have shown that the production of a good-quality text is a complex interaction of several subskills at different hierarchical levels (Berninger & Winn, 2006; Kim & Graham, 2022). At a lower level, transcription skills such as spelling and handwriting fluency build the basis; at a higher level, formulation skills at text level, such as lexical diversity or cohesion, and which are therefore referred to as (text-specific) formulation skills, are necessary for good text quality. The point here is that, while performance in low-level transcription skills alone might not be sufficient to produce good-quality texts, these skills are still needed (Philippek et al., 2024). Therefore, on the one hand it is hypothesized to find children with normal to high performance in transcription skills but poor text production (based on writing modeling assumptions (Philippek et al., 2024) due to poor performance in those higher-level text-specific formulation skills). On the other hand, the question arises whether there are children with poor performance in transcription skills but high competence in text production due to high performance in formulation skills. The aim of this study is therefore to examine whether these assumptions are correct and whether double dissociations exist in this form.

1.1 Writing competence

Producing a text is a complex process that requires a broad set of skills which vary depending on the specific demands of the writing task environment (Hayes, 2012). How well a text is written is operationalized by the text quality (Feenstra, 2021). High text quality is characterized by the writer's attention to the reader's needs, generation, selection and organization of ideas, familiarity with appropriate text patterns, and the ability to produce a coherent text that fulfils its intended purpose (Harsch et al., 2007; Hennes et al., 2018). To achieve a good-quality text, the writer must meet these demands in a goal-oriented manner, while also following the conventions of the specific genre (Beers & Nagy, 2009; Hayes, 2012). For decades, various writing models have sought to explain the subskills which are needed for producing a good-quality text (e.g., Berninger & Winn, 2006; Hayes, 2012; Kim & Graham, 2022). There is a consensus that a good-quality text should be orthographically correct, lexically diverse, coherent and adapted to the target group and text genre (Becker-Mrotzek et al., 2014; MacArthur et al., 2019; Olinghouse & Wilson, 2013).

Despite numerous models that already depict the writing process as well as the interaction of the various subskills very well, there is still a lack of sufficient information about the underlying construct of text quality, which makes operationalization difficult. A distinction is currently made between global and analytical ratings. The former involve judgements of the entire text: In most cases, a multi-level scale is used, the different levels of which are defined by various criteria of text quality (N. Lindauer & Sommer, 2018). The advantage of this variant

its time efficiency (N. Lindauer & Sommer, 2018; Neumann, 2017). However, global ratings vary significantly because each rater has their own criteria for what makes a text good (Myford & Wolfe, 2003). In analytical ratings, a text is assessed on the basis of individual dimensions in the form of items. The items within so-called analysis grids relate, for example, to the structure, content, and linguistic composition of a text. The advantages of this approach are that, on the one hand, raters can obtain a more differentiated picture of the text and, on the other hand, they can provide differentiated feedback (N. Lindauer & Sommer, 2018; Neumann, 2017). Here the question arises whether the chosen items are a valid representation of the underlaying construct of text quality (Sadler, 2009). Although, according to current studies, both methods represent an appropriate way of assessing text quality, there are also studies in which the results regarding reliability differ widely (Lesterhuis et al., 2017). For this reason, another form of text quality assessment, which also saves resources, has come to the foreground in research: comparative judgement. This is based on the premise that raters are more consistent in comparing performance than in giving grades for individual performance (Thurstone, 1927). According to this approach, holistic assessments are done by a group of raters whose task is to compare two texts globally in several rounds and decide which is of higher quality, resulting in a binary decision matrix (McMahon & Jones, 2015). Based on a logistic model, a ranking can then be created across all texts from best to worst text quality (Lesterhuis et al., 2017), which can be considered as the collective agreement of raters regarding the extent to which a particular text is of higher quality than the other texts (van Daal et al., 2019). Studies have shown that the comparative judgement method has a high reliability and consistency in the evaluation of text quality and thus offers the possibility of overcoming reliability problems in the evaluation of writing (Lesterhuis et al., 2017), provided that 12 or more comparisons per text are made (Verhavert et al., 2019). A major advantage of this method is that even inexperienced raters can achieve reliable results (Boonen et al., 2020). It therefore represents a solution for operationalization in research. However, it is not possible to derive indications for the promotion of writing competence from this method. It remains unclear which subskills should be given the most attention in writing instruction.

1.2 Transcription skills are important, but are they all that count?

The most basic skills in writing are transcription skills. These include spelling, which is the retrieval, recognition, and presentation of orthographic symbols, as well as handwriting fluency, which involves the fine motor movements needed to produce these orthographic symbols (Abbott & Berninger, 1993). Both spelling and handwriting fluency are key competences, especially for beginning writers (Graham & Santangelo, 2014; Santangelo & Graham, 2016).

Research findings suggest that the writing process can be affected by poor transcription skills. With poor spelling, for example, writers use a lot of cognitive resources to solve spelling problems, which may have a negative impact on text quality (Breetvelt et al., 1994). Similarly, poor handwriting fluency might impact idea generation because slow writers may find it

difficult to hold their ideas in mind while writing. In support of this claim, there is evidence that novice and slow writers tend to produce better texts when speaking than when writing (Bereiter & Scardamalia, 1987; Graham, 1990; Hayes & Berninger, 2010). Empirical studies support these assumptions by confirming that transcription skills have a significant influence on text quality from 1st to 9th grade, both directly and indirectly via other cognitive skills or higher-level writing skills. For example, Graham et al. (1997) found that handwriting fluency and spelling accounted for 25% of the variance in text quality in primary grades (1-3) and for 42% of the variance in text quality in intermediate grades (4-6). This is consistent with a longitudinal study by Abbott et al. (2010), which showed that spelling skills were a stable predictor of text quality from 1st to 7th grade, and that handwriting fluency had a direct effect on text quality from 3rd to 4th grade. Cordeiro et al. (2020) were also able to prove the relevance of transcription skills for text quality in 2nd graders, while Limpo and Alves (2013) found a direct effect of transcription skills on text quality for 4th to 6th graders, though this effect was indirect – via planning and self-efficacy – in secondary school students (grades 7 to 9). Limpo et al. (2017) found similar indirect effects of handwriting fluency via higher-level planning skills and indirect effects of spelling via higher-level translating skills (such as syntactic correctness and ability to combine sentences) in secondary school. The authors suggest that transcription skills continue to impact text quality in secondary school by constraining the higher-level skills.

In order not to hamper the writing process, it is necessary that transcription skills are automatized. This frees up cognitive capacity for higher-level writing skills. Transcription automatization is reached when students write quickly without misspellings (Graham & Santangelo, 2014; Limpo & Alves, 2013). The relationship between transcription skills and text quality becomes increasingly less important in higher grades, which is a desirable outcome of instruction (Sturm et al., 2017). Although it is not exactly clear at what age full automatization of transcription skills can be expected, it is evident that they can only be achieved through extensive and, above all, explicit practice (Graham & Santangelo, 2014; Santangelo & Graham, 2016). Transcription skills are attained through systematic training that includes the fast and correct transcription of letters, words, and even entire texts (Graham & Santangelo, 2014; Limpo & Alves, 2013). Improvement of transcription skills appears in the production of longer texts (Sturm, 2017), which is known to be associated with text quality (Fleckenstein et al., 2020; MacArthur et al., 2019). Studies have shown that this systematic training of transcription skills is effective for both primary school students (Alves & Limpo, 2015; Hurschler Lichtsteiner et al., 2018; Limpo & Graham, 2020) and weaker intermediate school students (Limpo et al., 2017). Meta-analyses on writing instruction from 1st to 12th grade also confirm the effectiveness of these trainings (Graham & Harris, 2018; Graham & Santangelo, 2014; Santangelo & Graham, 2016) and the relevance of daily practice to produce good-quality texts (Berninger et al., 2002; Limpo & Alves, 2018; Santangelo & Graham, 2016). Rocha et al. (2022) postulate that this explicit and extensive training should continue even beyond the primary school years until a sufficient level of automatization has been achieved. As a consequence of the importance of transcription skills for text quality, the promotion of

transcription skills in schools has become the focus of writing instruction. However, this means that writing instruction often lacks a balanced relationship between writing strategies and low- and higher-level subskills (Cordeiro et al., 2020; Graham & Harris, 1997; Moats et al., 2006). Although transcription skills are an important part of text composing, they are probably not the only or even the most important part of composing competence (T. Lindauer & Schmellentin, 2017).

1.3 The role of text-specific formulation skills

The importance of higher-level text-specific formulation skills is assumed in writing models, which attempt to understand the complex construct of writing competence by describing the subskills involved and how they interact to produce a good-quality text. A well-known writing model that includes various subskills of writing competence is the Not-So-Simple View of Writing Model (Berninger & Winn, 2006). In addition to transcription skills, this model includes executive functions and text generation. Here, transcription skills and executive functions support text generation in a system that coordinates working memory functions (Berninger & Winn, 2006). However, the Not-So-Simple View of Writing Model does not provide a detailed description of how the individual writing processes are coordinated. To address this, Olive (2014) developed a parallel and cascading model of writing processes. He postulated that, depending on the availability of cognitive resources, the different writing processes can occur simultaneously in parallel processing. Skilled writers can plan upcoming content while writing the current segment. Additionally, he conceptualized writing as a cascading process, in which information flows continuously from higher-level processes of text conceptualization (e.g., planning) to low-level processes (e.g., handwriting). Unlike strict serial models, the cascading process allows a low-level process to start before the higher-level one is fully completed (Olive, 2014). While Olive's parallel and cascading model of writing focuses on the writing process, Philippek et al. (2024) introduced a model called the Cascaded Model of Writing (CASMOW), which examines the interaction of different subskills and their influence on text quality. The model includes various subskills that are arranged in cascades: Executive functions, low-level transcription skills and text-specific formulation skills (termed higher-level text-specific skills) including the usage of a diverse and appropriate text-specific vocabulary and the ability to establish cohesion. Empirical findings of CASMOW show that, in 5th to 7th grade students, text-specific formulation skills mainly influence the text quality of narratives, while transcription skills and executive functions only indirectly contribute to text quality (Philippek et al., 2024).

The significant effects of these text-specific formulation skills on text quality have already been confirmed in numerous studies in which these subskills were considered in isolation (Cox et al., 1990; Gómez Vera et al., 2016; MacArthur et al., 2019; Olinghouse & Wilson, 2013). In the context of *text-specific vocabulary*, a prominent measurement is lexical diversity, which assesses the range of vocabulary used in a text (McCarthy & Jarvis, 2010). There is evidence that this measure is a significant predictor of text quality in late primary school and early secondary school (Cameron et al., 1995; Gómez Vera et al., 2016; Olinghouse & Wilson, 2013).

In the context of text-specific vocabulary, the use of appropriate words is also an important measure. Significant correlations with text quality have confirmed this (Mathiebe, 2019). *Coherence* is also important because a well-structured text makes it easier for the reader to understand the goal of the writer and to create a picture in their mind (Averintseva-Klisch, 2018). *Cohesion*, as the measurable part of coherence on the surface level of a text, can be captured by connectors and other references between sentences (Schwarz, 2001; Zifonun, 2000). Numerous studies verify the significant effect of cohesion measures on text quality from late primary and early secondary school (Cameron et al., 1995; Cox et al., 1990) to university students (Crossley & McNamara , 2010; MacArthur et al., 2019).

1.4 The current study

Based on the assumptions of the writing models outlined above and the reviewed literature, we hypothesised that there are children with age-appropriate performance in transcription skills but poor text production, which could be explained by the construct of double dissociations. Double dissociations of this kind are generally interpreted as an indication that the respective cognitive functions – even if they are correlated – might involve different cognitive mechanisms (Gegenfurtner et al., 2022). In this case, writing competence (measured by text quality) might imply different mechanisms than basic transcription skills. In this study, we investigated whether there are students who show this double dissociation of transcription skills and text quality. We therefore examined spelling and handwriting fluency separately with regard to double dissociation to determine whether this applies to both components. Where groups of double dissociation existed, we determined whether they can be characterized by their performance in text-specific formulation skills.

The following research questions will be addressed:

1. Is there clearly definable double dissociation in the performance of low-level transcription skills and text quality in 5th to 7th grade students (age-appropriate text quality/ deficient low-level transcription skills or deficient text quality/ age-appropriate low-level transcription skills)?
2. If double dissociation exists, does it differ with regard to text-specific formulation skills, so that these skills explain students' performance in text production?

2. Method

2.1 Participants

The sample comprised 186 native speaking students from two randomly selected schools in North Rhine-Westphalia (Germany), where students can obtain a higher education entrance

qualification.[1] Students were between nine and 14 years old (M = 11.75; SD = 1.05; 48% female). Fifty-seven (30.6%) of these were 5th graders, 63 (33.9%) 6th graders and 66 (35.5%) 7th graders from three classes each.

2.2 Measures

Handwriting fluency

To test handwriting fluency, the alphabet task of the Detailed Assessment of Speed of Handwriting (DASH; Barnett et al., 2007) was conducted. For this, students were asked to write down the alphabet in lowercase letters in correct order as often as possible in 60 seconds. The number of correct letters corresponded to the final score: The higher the score, the faster the handwriting fluency. The alphabet task has a satisfactory convergent validity and interrater reliability (ICC = .99) (Barnett et al., 2009).

Spelling

The Hamburger Schreibprobe (Hamburger Spelling Test [HSP]; May et al., 2018) was used to assess spelling abilities. This test consists of a dictation of words and sentences in which alphabetical, orthographic, and morphemic strategies must be applied. The number of correctly spelt words was used for calculation. The higher the score, the better the spelling ability. The HSP has a satisfactory convergent validity and a reliability of Cronbach's α = .94 at the word level (May et al., 2018).

Text Quality

Students were asked to write a narrative on the topic "What if I could fly…". Their texts were typed and corrected for spelling. To assess text quality, the comparative judgement method was used as this is a robust method for comparing complex constructs like text quality (Pollitt, 2012b; van Daal et al., 2022). This procedure involves a pairwise comparison of texts by a large number of raters. One rater was randomly shown two texts at the same time. The rater was then asked to decide which of the two texts was better by clicking on it. Raters were not trained or given specific criteria to focus on: Each rater could decide individually what constitutes a good-quality text. Therefore, it is possible that some raters had a broad view and focused on multiple elements, while others had a narrow view and focused only on single elements (van Daal et al., 2019). Furthermore, it remains unclear which aspects of the text surface were most important when evaluating the texts. However, comparative judgement assumes that the multiple ratings of each text and the resulting final ranking reflects a common consensus of all ratings and thus also the construct of text quality, which contributes to its validity and reliability (I. Jones et al., 2015; Pollitt, 2012a). Therefore, the students' texts

[1] Corresponds to the German school type "Gymnasium" and "Gesamtschule".

were scored by 65 naïve raters with no prior text scoring experience through 15 pairwise comparisons per rater, resulting in each text being scored 79 times. The absolute number of comparisons made was 961, resulting in a satisfactory scale separation reliability of r = .73 (analogous to Cronbach's alpha). Using a logistic model, a logit score was then calculated from the ratings of all texts indicating the probability for a particular text to "win" the pairwise comparison (Pollitt, 2012b). The higher the logit score, the better the text (L. Jones & Karadeniz, 2016). In this way, the ranking reflects the different views on text quality and also reduces the discrepancies between the raters (van Daal et al., 2022). In this study, the entire rating was carried out using the online tool Comproved (www.comproved.com).

Text-specific vocabulary

In addition to text quality, children's texts were used to determine text-specific vocabulary, as this ability is one of the text surface characteristics which should be assessed in context (Cameron et al., 1995). There are various measures for operationalizing text-specific vocabulary. The best-known is lexical diversity, which describes the range of the vocabulary of an entire text. In this study, the Measure of Textual Lexical Diversity (MTLD) was calculated, which is largely unaffected by the length of the text and is therefore reliable (McCarthy, 2005). Satisfactory validity is demonstrated by MTLD's correlation with other measures of lexical diversity (McCarthy & Jarvis, 2010). The MTLD was calculated by counting how often a type-token ratio (TTR) of .72 (determined as the stabilization point) was reached in the text. This was calculated from the beginning to the end of the text. The text was then analyzed again from the reverse to calculate a second MTLD score. The average of the two scores formed the final MTLD score. The higher the score, the better the lexical diversity. The MTLD was calculated in this study using R (R Core Team, 2022).

Appropriate Word Usage

Equally relevant for the reader's understanding and thus for a good-quality text is the selection of appropriate words (Mathiebe, 2019). In this study, word appropriateness was measured by counting inappropriate word-usages (children wrote for example: "*Fire would **burst** from my wings" or "[The world] **expands** with the tears of people")*. These were determined and counted by two linguistic experts. As the number of inappropriate words is influenced by the length of the text, the count was divided by the number of words in the text. The interrater reliability was $\kappa = .88$ ($p < .001$). Disagreements were discussed until a consensus was reached. The higher the score, the higher the appropriate word-usage.

Cohesion

To determine cohesion in the text, the same procedure was used as for appropriate word-usage. For this purpose, various cohesion measures were determined for assessment: The measure of local cohesion (lacks of references or connection of elements that do not belong together) as well as grammatical aspects such as sentence errors (incorrect syntax caused by

missing words or incorrect word order) and tense errors (inappropriate change between tenses), which can also create incoherence due to internal grammatical inconsistency (Klotz, 2022; Schwarz, 2001), were evaluated. The interrater reliability between the raters was κ = .87 ($p < .001$) for local cohesion errors, κ = .89 ($p < .001$) for sentence errors, and κ = .92 ($p < .001$) for tense errors. Again, in case of disagreement the raters discussed until reaching consensus. Because these measures were also influenced by the text length, quotients were formed by dividing local cohesion errors by the number of propositions, sentence errors by the number of sentences, and tense errors by the number of verbs. For easier interpretation, scores were inverted: The higher the score, the more cohesive the text.

2.3 Procedure

Data collection was carried out for one school lesson of 45 minutes on two different days. All tests were performed in the same order in each class: In the first lesson the writing task was conducted and in the second lesson the alphabet task and HSP were carried out. For the assessment the students used paper and pencil.

2.4 Statistical Analysis

Data analysis was conducted using SPSS version 29 (IBM Corp., 2023) as well as R (R Core Team, 2022). In order to determine whether there are students with poor transcription skills but age-appropriate performance in text quality or the reverse, students were categorized (separately) based on their spelling, handwriting fluency, and text quality skills. A percentile score at or below 20 was categorized as deficient achievement, while a percentile score at or above 25 was defined as age-appropriate performance. Percentile scores between 20 and 25 were not considered (Moll & Landerl, 2009). Since in this study the standard values of the HSP are related to categorization as deficient and age-appropriate, performance below 20% corresponded to a percentile rank of the HSP of between 1% and 31% for all school types. Of these, 63% of the weaker sample have a percentile rank below 15%, which can be regarded as an indication of the need for support, showing that the categorization made in this study corresponds to the weaker students.

Based on this categorization, in a next step, spelling or handwriting fluency skills were related to performance in text quality. A configural frequency analysis (CFA) was conducted to examine the patterns of symptom occurrence across variables (text quality, spelling, handwriting fluency). The Pearson chi-square test was used to compare observed and expected frequencies, and configurations with significant deviations were classified as types (significantly more frequent) or antitypes (significantly less frequent). The group differences in text-specific formulation skills were then analyzed using analyses of variance (ANOVAs).

Nonnormality and outliers: After eliminating outliers based on the 1.5 x IQR criteria, handwriting fluency and text quality were found to be normally distributed. The other variables were not normally distributed (Table 1). Therefore, the parametric analysis was repeated based on 1000 bootstrap samples, which revealed essentially the same results for all analyses. Descriptive statistics for all variables are shown in Table 1.

Table 1. Means and Standard Deviations for all Variables of the Study

Measures	Mean	Standard deviation	Range	Skewness	Kurtosis
Handwriting fluency	63.25	23.00	8 – 138	-0.0	0.09
Spelling	36.94	8.24	13 – 49 (49)a	-0.69	-0.38
Lexical diversity	58.17	23.30	15.85 –153.00	1.01	1.78
Appropriate words	0.99	0.02	0.91 –1.00	2.01	5.0
Local cohesion	0.89	0.12	0.38 –1.00	1.34	2.22
Correct tense	0.89	0.13	0.46 – 1.00	1.18	0.63
Correct sentences	0.90	0.16	0.00 –1.00	2.61	9.38
Text quality	0.06	1.34	-3.84 – 2.99	-0.38	0.12

[a] theoretical maximum

3. Results

3.1 Correlations

Zero order correlations revealed that both transcription skills and formulation skills (with the exception of correct tense) had an effect on text quality with lexical diversity and (local) cohesion being the best predictors (Table 2). What is also shown is that – as theoretically assumed – transcription skills had an influence on some text-specific formulation skills, namely lexical diversity, cohesion, and correct sentences (the latter only for spelling, not for handwriting fluency).

Table 2. Correlations of Variables

	Handwriting fluency	Spelling	Lexical diversity	Appropriate words	Local cohesion	Correct tense	Correct sentences
Spelling	.41**						
Lexical diversity	.25**	.34**					
Appropriate words	.01	.07	.18*				
Local cohesion	.01	.17*	.19**	.06			
Correct tense	.03	.05	.07	.01	.02		
Correct sentences	.04	.33**	.05	.01	.03	.06	
Text quality	.30**	.44**	.53**	.20**	.28**	.09	.36**

Note. ** p< .01, * p< .05

Double dissociation groups of low-level transcription skills and text quality

According to our categorization, four groups were identified for spelling and handwriting fluency as transcription skills: 1) students with age-appropriate text quality (TQ) and age-appropriate spelling (SP; 67.0%; N = 114) or handwriting fluency (HF; 68.5%; N = 115) skills (TQ+/ SP+ or HF+); 2) students with age-appropriate text quality and deficient spelling (11.8%; N = 20) or handwriting fluency (11.9%; N = 20) skills (TQ+/ SP- or HF-); 3) students with deficient text quality and age-appropriate spelling (11.8 %; N = 20) or handwriting fluency (11.3 %; N = 19) skills (TQ-/ SP+ or HF+); and 4) students with deficient text quality and deficient spelling (9.4 %; N = 16) or handwriting fluency (8.3 %; N =14) skills (TQ-/SP- or HF-) (Figure 1).

16 students did not fall into any spelling group, and 18 students did not fall into any handwriting fluency group. Ten of these students were identical.

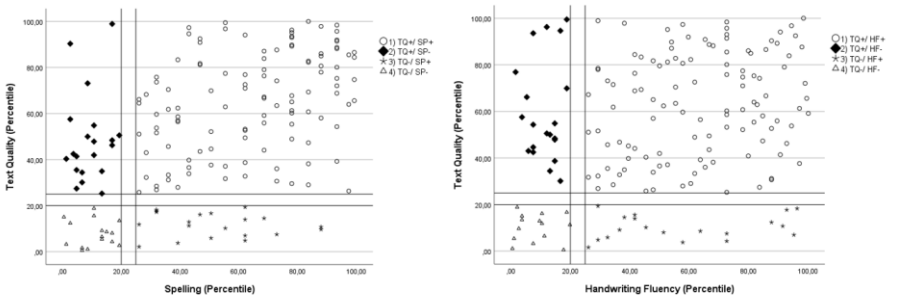

Figure 1: Spelling/ Handwriting Fluency vs. Text Quality Double Dissociation Groups

Note. X-axis: fractional rank percent of scores of HSP; y-axis: fractional rank percent of text quality; SP = spelling; TQ = text quality; + = age-appropriate performance; - = deficient performance.

As the profiles for spelling and handwriting fluency were created separately, it is also interesting to look at whether there is an overlap of students in the respective groups. The CFA revealed several configurations where an overlap of students was found across different combinations of variables. If the overlaps are linked to the group sizes of the individual profiles, it can be seen that the students overlap to some extent between handwriting fluency and spelling. The overlap of students who had developed age-appropriately in all three variables was highest and was also significantly more frequent than expected (observed = 98; expected = 79.91; p = .022). The lowest overlap of students was found for poor text quality and handwriting fluency, but age-appropriate spelling. This provided an observed frequency of 6 students, which was expected (p = .403). The CFA also provides information on the extent of overlap between the four groups determined by spelling and handwriting fluency performance. A total of 98 identical students were assigned to Group 1 for both spelling and handwriting fluency (TQ+/ SP+/ HF+). Group 2 (TQ+/ SP-/ HF-) comprised nine, Group 3 (TQ-/ SP+/ HF+) comprised 12 and Group 4 (TQ-/ SP-/ HF-) comprised 8 identical students (Table 3).

Table 3. CFA: Overlap of Students with Poor/ Good Performance in Text Quality, Handwriting Fluency, and Spelling

Text quality	Spelling	Handwriting fluency	Observed frequency	Expected frequency	Result	z	p
< 20	< 20	< 20	8	1.50	type	5.32	> .001
< 20	< 20	> 25	7	5.63	neutral	0.58	.283
< 20	> 25	< 20	6	5.43	neutral	0.25	.403
< 20	> 25	> 25	12	20.44	antitype	-1.87	.031
> 25	< 20	< 20	9	5.85	neutral	1.30	.096
> 25	> 25	< 20	11	21.23	antitype	-2.22	.013
> 25	< 20	> 25	11	22.02	antitype	-2.35	.009
> 25	> 25	> 25	98	79.91	type	2.02	.022

Note. Global test $\chi^2(4) = 48.38$*. Types and antitypes based on Bonferroni corrected alphas (5 %).

Group differences in text-specific formulation skills

ANOVAs with double dissociation groups as independent variables and formulation skills as dependent variables indicated that the use of text-specific formulation skills differed significantly depending on the transcription skills/TQ group.

For spelling, profiles differed significantly in the use of lexical diversity ($F (3, 166) = 10.37$; $p < .001$; $\eta p^2 = .16$), correct sentences ($F (3, 166) = 7.97$; $p < .001$; $\eta p^2 = .13$), and local cohesion ($F (3, 166) = 4.12$; $p = .008$; $\eta p^2 = .07$), but not in the use of appropriate words ($F (3, 166) = 2.12$; $p = .1$) and correct tense ($F (3, 166) = 0.71$; $p = .55$). Within the text-specific formulation skills, various homogeneous subgroups crystallized, which showed the same use of the particular skills. Means and post hoc comparisons revealed that, in spite of deficient spelling, Group 2 (those with age-appropriate text quality) had fewer problems in building correct sentences and establishing local cohesion than Group 4 (deficient spellers/ deficient writers). For lexical diversity, Group 2 formed a homogeneous group with Group 4. The deficient text quality/ good spelling group (Group 3) had more problems than the good text quality/ good spelling group (Group 1) in using a lexical diversity of words. The nondiscrepant groups (Groups 1 and 4) differed in lexical diversity, correct sentences, and local cohesion (Table 4, Figure 2).

Figure 2 illustrates the difference between the double dissociation groups (Groups 2 and 3) and the two groups with equal performance in spelling and text quality (Groups 1 and 4). It shows that the students who were able to produce good texts despite deficient spelling (Group 2) used a lexically more diverse vocabulary, produced more correct sentences and linked these together significantly more often using cohesive devices than the students with poor spelling and poor text quality (Group 4). A comparison of the two groups who showed age-appropriate spelling skills revealed that those students who were able to produce good texts (Group 1) showed greater lexical diversity and more local cohesion than their peers who showed similarly good spelling skills but deficient text quality (Group 3).

Table 4. Performance on Text-Specific Formulation Skills. Discrepant and Nondiscrepant Groups Based on Spelling

	(1) TQ+/ SP+	(2) TQ+/ SP-	(3) TQ-/SP+	(4) TQ-/ SP-	F (3, 166)
Lexical diversity	.27a	-.23ab	-.62b	-.83b	10.37**
Appropriate words	.10	-.028	-.45	.03	2.12
Correct sentences	.24a	-.07a	-.29a	-.89b	7.97**
Local cohesion	.16a	.05a	-.35ab	-.63b	4.12**
Correct tense	-.01	.25	-.2	-.05	0.71

Note. Text-specific formulation skills are z-standardized; TQ = text quality; SP = spelling; *p < .05; **p < .01. Superscript letters represent homogeneous subgroups.

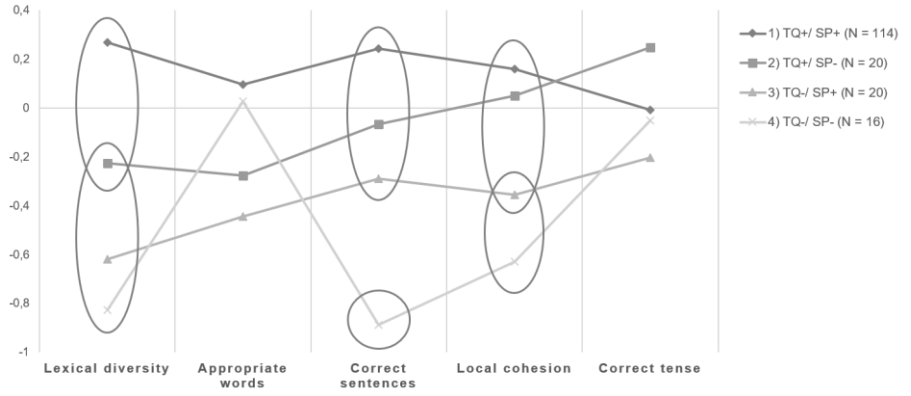

Figure 2: Anova Statistics of Performance Profiles of Spelling and Text-Specific Formulation *Note.* The y-axis shows the z-standardised measures of the text-specific formulation skills. The circles indicate the homogeneous subgroups.

For handwriting fluency, groups differed significantly in lexical diversity ($F (3, 164) = 10.43$; $p < .001$; $\eta p^2 = .16$), correct sentences ($F (3, 164) = 8.03$; $p < .001$; $\eta p^2 = .13$), and local cohesion ($F (3, 164) = 3.51$; $p = .017$; $\eta p^2 = .06$), but not in the use of appropriate words ($F (3, 164) = 1.65$; $p = .179$) and correct tense ($F (3, 164) = 0.89$; $p = .448$). Means and post hoc comparisons revealed that the group of good writers – in spite of deficient handwriting fluency (Group 2) – had fewer evident problems in building correct sentences and using a lexically-diverse vocabulary than the deficient handwriting fluency/ deficient text quality group (4). For local cohesion students from Group 4 were somewhat weaker than the other groups (except deficient text quality/ good handwriting fluency (Group 3)), but this result did not reach statistical significance. The deficient text quality/ good handwriting fluency group (3) had more problems than the good text quality/ good handwriting fluency group (1) in using

lexically-diverse words, but remarkably fewer problems than the nondiscrepant deficient text quality/ handwriting fluency group (4). Group 3 also had fewer problems in building correct sentences than the deficient nondiscrepant group (4), and than the good text quality/ deficient handwriting fluency group (2). Building local cohesion was more difficult for Group 3 than for the good text quality groups (Groups 1 and 2). The nondiscrepant groups differed in lexical diversity and correct sentences (Table 5, Figure 3).

Figure 3 illustrates the difference between the groups with double dissociation (Groups 2 and 3) and the two groups with equal performance in handwriting fluency and text quality (Groups 1 and 4). It shows that the students who were able to produce good texts despite poor handwriting fluency (Group 2) used a more lexically diverse vocabulary and produced more correct sentences than the students with poor handwriting fluency and text quality (Group 4). A comparison of the two groups with age-appropriate handwriting fluency showed that the students who were able to produce good texts (Group 1) had greater lexical diversity and more local cohesion than their peers who showed similarly good handwriting fluency but poor text quality (Group 3).

Table 5. Performance on Text-Specific Formulation Skills. Discrepant and Nondiscrepant Groups Based on Handwriting Fluency

	(1) TQ+/ HF+	(2) TQ+/ HF-	(3) TQ-/ HF+	(4) TQ-/ HF-	F (3, 164)
Lexical diversity	.26^{a}	-.28^{ab}	-.50^{bc}	-.97^{c}	10.43**
Appropriate words	.10	-.16	-.40	-.10	1.65
Correct sentences	.19^{ab}	.35^{b}	-.32^{a}	-.95^{c}	8.03**
Local cohesion	.13^{a}	.29^{a}	-.55^{b}	-.15^{ab}	3.51*
Correct tense	-.01	.23	-.05	-.34	0.89

Note. Text-specific formulation skills are z-standardized; TQ = text quality; SP = spelling; *p < .05; **p < .01. Superscript letters represent homogeneous subgroups.

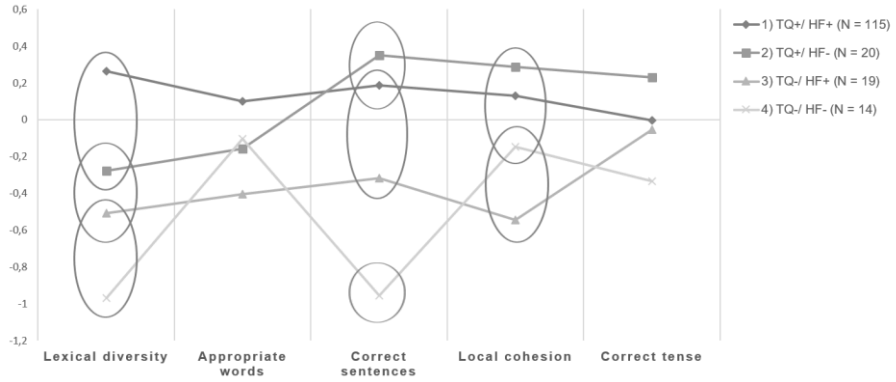

Figure 3: ANOVA Statistics of Performance Profiles of Handwriting Fluency and Text-Specific Formulation Skills

Note. The y-axis shows the z-standardised measures of the text-specific formulation skills. The circles indicate the homogeneous subgroups.

4. Discussion

In this study, we examined double dissociation between low-level transcription skills and text quality in a sample of German-speaking secondary school children and investigated the role of text-specific formulation skills in students with discrepant achievements.

The initial research question was whether double dissociation groups can be identified. As hypothesized, despite the influence of transcription skills on text quality (Abbott et al., 2010; Cordeiro et al., 2020; Graham et al., 1997; Limpo & Alves, 2013), we found double dissociation groups in which the performance of transcription skills (spelling and handwriting fluency) and text quality did not match. These results applied to both handwriting and spelling. Here, CFA showed that most students performed the same in handwriting fluency and spelling. This in turn shows that the two low-level skills develop similarly, if not interdependently, which is also assumed in existing writing models (e.g., Berninger & Winn, 2006). Discrepant groups and the nondiscrepant group with poor achievement in both basic transcription skills and text quality were approximately of the same size (about 10% of the sample). As mentioned, double dissociations might be interpreted as an indication that related cognitive functions involve different cognitive mechanisms (Gegenfurtner et al., 2022). Our hypothesis was that composing competence (measured by text quality) implies different mechanisms than basic transcription skills. According to some writing models, e.g. the Not-So-Simple View of Writing or the Cascaded Model of Writing (Berninger & Winn, 2006; Philippek et al., 2024), this has to be assumed. However, as a consequence of past (positive) findings concerning the effect of transcription skills on students' writing skills, the focus of instruction is often on transcription

skills, but less on formulation skills (Cordeiro et al., 2020; Graham & Harris, 1997; Moats et al., 2006).

Since a double dissociation was found, the second research question sought to ascertain whether there were differences in the use of formulation skills among the various groups. Results showed that children who were good writers/ poor spellers had fewer problems in creating correct sentences and local cohesion in comparison to children who showed deficient performance in both text quality and spelling. Likewise, poor writers/ good spellers had more problems with the implementation of a lexically-diverse vocabulary in their texts and were weaker in building correct sentences and local cohesion in comparison to children who were good in both text quality and spelling, although the latter did not reach statistical significance. Comparing the two non-dissociation groups, the deficient group showed significantly more problems in the use of lexical diversity, correct sentences, and local cohesion.

As with handwriting fluency, good writers/ poor handwriters had fewer problems with lexical diversity and correct sentence building and were somewhat better, but not significantly, in the use of local cohesion than children showing deficits in both handwriting fluency and text quality. Children who were good writers/ good handwriters were better in using lexically-diverse words and local cohesion than children who were poor writers/ good handwriters. They were also somewhat, but not significantly, better in forming correct sentences. Comparing the deficient group with poor writers/ poor handwriters with the higher-performance group with good writers/ good handwriters, the latter were clearly better at using lexically-diverse vocabulary and building correct sentences and tended to be better in local cohesion. In general, all comparisons showed that lexical diversity, correct sentence building, and local cohesion were responsible for the production of good texts, regardless of whether transcription skills were developed to an age-appropriate level. This is in line with previous research that confirmed the influence of these subskills on text quality in narratives (Cox et al., 1990; Gómez Vera et al., 2016; MacArthur et al., 2019; Olinghouse & Wilson, 2013), particularly for children who have already completed basic writing instruction and for whom the relationship of transcription skills and text quality thus decreases with increasing schooling (Sturm et al., 2017), even though transcription skills are still important (Abbott et al., 2010; Cordeiro et al., 2020; Graham et al., 1997). In later schooling, there are indirect effects of transcription skills on text quality via higher-level skills (Limpo & Alves, 2013), which can be explained by the ongoing automatization of transcription skills, freeing up more cognitive capacity to use higher-level skills (Berninger & Winn, 2006). In light of the aforementioned considerations, it was unexpected that there existed a group with poor transcription skills but good text-specific formulation skills. This result is not consistent with existing theories of writing that require full automation of transcription skills for good text production (Berninger & Winn, 2006; Kim & Graham, 2022). It seems that the study may have identified students whose writing activity cannot be fully explained by existing theories of writing as there are students who are able to use higher-level skills to produce good-quality texts, despite not having fully automated transcription skills.

The results of this study therefore show that, regardless of the automation of transcription skills, at least for some children, higher-level text-specific formulation skills are crucial for text quality.

To complete the results, appropriate word usage and correct tense did not differ significantly between the different groups, in spelling or in handwriting fluency. It seems that these subskills have less impact on differences in text quality. In narrative texts, a change in tense and a wrong choice of words could be compensated by the context of a story and thus are not as important for the quality of a text as, for example, lexical diversity (Olinghouse & Wilson, 2013). These skills might be more relevant for other text genres (e.g., factual texts), as different skills are required of the writer depending on the genre (Beers & Nagy, 2009; Kim & Graham, 2022). Although inappropriate words can influence the reader's comprehension of the text (Mathiebe, 2019), we found little variance in the data in this study, which may have led to the fact that we found no significant differences between the groups.

4.1 Limitations and future research

This study contains some limitations that should be considered when interpreting the results. First, one single writing sample was used to determine composing competence. Probably, more than one text sample would obtain a more reliable measure of text quality (Graham et al., 2016). Second, the results relate only to narratives, the text genre chosen for this study. In particular, results for text-specific formulation skills might change when assessing other genres (Beers & Nagy, 2009), as different writing tasks require various writing skills (Kim & Graham, 2022). For example, lexical diversity is very important in narratives, but not in informative texts. Depending on the text genre, it would be interesting to consider other text-specific formulation skills such as information management or audience-orientation, which are known to be relevant in informative texts (Berman & Nir-Sagiy, 2007; Weston et al., 2011). Third, results from this study only apply to students from the 5th to 7th grades. The distribution of students in the performance groups could change as the grade level increases or decreases, as text-specific formulation skills develop in secondary school. In subsequent studies, it would be interesting to analyze the possible occurrence of dissociation groups in younger and older students and the relationship to text-specific formulation skills. Fourth, this study included students from schools where the university entrance qualification can be obtained. Thus, mainly higher-performing students were likely included. Weaker students from other types of school were not included in the calculations, which is why the range of writing skills was probably limited. In future studies, students from all school types should be included in order to represent the entire population of students.

4.2 Educational implications

As mentioned, writing instruction (at best) focuses primarily on the automatization of transcription skills in order to create cognitive capacities for the development of text-specific formulation skills and, thereby, the production of good-quality texts (Cordeiro et al., 2020; Graham et al., 2013). However, this study has shown that the automatization of basic

transcription skills alone is not sufficient to produce good-quality texts. Knowing that there are different groups, the focus in writing lessons should be on the individual needs of the students.

On the one hand, some children fail to write good-quality texts despite good spelling/ handwriting fluency skills, which in the current context is probably the most interesting group in terms of writing instruction. From our data, it is clear that these children need training in text-specific formulation skills, particularly as this becomes increasingly more important over years of instruction (Sturm et al., 2017). Writing instruction for these students should not neglect the importance of text-specific formulation skills and perhaps even focus on them from an earlier stage of development. As soon as students show problems in the production of texts, not only transcription skills, but also the use of text-specific formulation skills should be trained in common, since there is considerable evidence that young children can use writing in an academic setting before they are able to consistently apply the conventions of handwriting fluency and spelling (Mason & Allen, 1986; Palincsar & Klenk, 1993). Our study also suggests that it is not necessary to fully automatize transcription skills before promoting text-specific formulation skills.

On the other hand, our data has shown that there are also students in secondary school who produce good-quality texts but show deficits in the use of transcription skills. Spelling and handwriting fluency exercises are no longer the main focus of writing lessons in grades 5 to 7 (Kultusministerkonferenz [KMK], 2022). However, these students may profit from explicit instruction in transcription skills beyond primary school to improve their basic writing abilities.

Author note

Data availability: The data associated with this study can be downloaded from the Open Science Framework at the following link:
https://osf.io/v2p6f/?view_only=0927380a9a71455a8e254e4701faec6c

Funding: No funding was received for this research.

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

References

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

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

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

Averintseva-Klisch, M. (2018). Textkohärenz (2nd ed.). Universitaetsverlag Winter GmbH. https://api.pageplace.de/preview/dt0400.9783825379780_a40046660/preview-9783825379780_a40046660.pdf

Barnett, A. L., Henderson, S. E., Scheib, B., & Schulz, J. (2007). The Detailed Assessment of Speed of Handwriting (DASH). 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/10.1348/000709909x421937

Becker-Mrotzek, M., Grabowski, J [Joachim], 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

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

Berman, R., & Nir-Sagiy, B. (2007). Comparing Narrative and Expository Text Construction Across Adolescence: A Developmental Paradox. *Discourse Processes, 2*(43), 79–120. https://doi.org/10.1080/01638530709336894

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., & 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.

Boonen, N., Kloots, H., & Gillis, S. (2020). Rating the overall speech quality of hearing-impaired children by means of comparative judgements. *Journal of Communication Disorders, 83*, 105969. https://doi.org/10.1016/j.jcomdis.2019.105969

Breetvelt, I., van den Bergh, H., & Rijlaarsdam, G. (1994). Relations Between Writing Processes and Text Quality: When and How? *Cognition and Instruction, 12*(2), 103–123. https://doi.org/10.1207/s1532690xci1202_2

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

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. https://doi.org/10.2307/747987

Crossley, S. A., & McNamara, D. (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., & McNamara D. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32), 351–370. https://escholarship.org/content/qt6n5908qx/qt6n5908qx.pdf

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

Gegenfurtner, K. R., Goldstein, E. B., & Cacciamani, L. (Eds.). (2022). Lehrbuch. Wahrnehmungspsychologie: Der Grundkurs (10. vollständig überarbeitete und aktualisierte Auflage). Springer. https://doi.org/10.1007/978-3-662-65146-9

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. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology, 82*(4), 781–791. https://doi.org/10.1037//0022-0663.82.4.781

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., & Harris, K. R. (1997). It Can Be Taught, But It Does Not Develop Naturally: Myths and Realities In Writing Instruction. *School Psychology Review, 26*(3), 414–424. https://doi.org/10.1080/02796015.1997.12085875

Graham, S., & Harris, K. R. (2018). Evidence-Based Writing Practices: A Meta-Analysis of Existing Meta-Analyses. In R. Fidalgo, K. R. Harris, & M. Braaksma (Eds.), Studies in writing: volume 34. Design principles for teaching effective writing: Theoretical and empirical grounded principles (pp. 13–37). BRILL. https://doi.org/10.1163/9789004270480_003

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

Graham, S., Kim, Y.-S., Cao, Y., Lee, J. w., Tate, T., Collins, P., Cho, M., Moon, Y., Chung, H. Q., & Olson, C. B. (2023). A meta-analysis of writing treatments for students in grades 6–12. *Journal of Educational Psychology, 115*(7), 1004–1027. https://doi.org/10.1037/edu0000819

Graham, S., & Perin, D. (2007). Writing next: Effective strategies to improve writing of adolescents in middle and high schools: A report to Carnegie Corporation of New York. Alliance for Excellent Education. https://education.illinoisstate.edu/downloads/casei/5-15-writingnext.pdf

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

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., & Berninger, V. W. (2010). Relationships between idea generation and transcription: How the act of writing shapes what children write. In C. Bazerman, R. Krut, K. Lunsford, S. McLeod, S. Null, P. Rogers, & A. Stansell (Eds.), Traditions of Writing Research (pp. 166–180). Taylor & Francis/Routledge.

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.

Hurschler Lichtsteiner, S., Wicki, W., & Falmann, P. (2018). Impact of handwriting training on fluency, spelling and text quality among third graders. *Reading and Writing, 31*(6), 1295–1318. https://doi.org/10.1007/s11145-018-9825-x

IBM Corp. (2023). IBM SPSS Statistics for Windows (Version 29.0.2.0) [Computer software]. IBM Corp.

Jones, I., Swan, M., & Pollitt, A. (2015). Assessing Mathematical Problem Solving Using Comparative Judgment. *International Journal of Science and Mathematics Education, 13*(1), 151–177. https://doi.org/10.1007/s10763-013-9497-6

Jones, L., & Karadeniz, I. (08.2016). An alternative approach to assessing achievement. Proceedings of the 2016 4th Conference of the International Group for the Psychology of Mathematics Education, Szeged, Hungary. https://repository.lboro.ac.uk/articles/an_alternative_approach_to_assessing_achievement/9373622/files/16984337.pdf

Kim, G. Y.-S., & 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

Klotz, P. (2022). Grammatik und Textualität. In H. Gornik (Ed.), Deutschunterricht in Theorie und Praxis Series. Sprachreflexion und Grammatikunterricht (2. überarbeitete und erweiterte Auflage, pp. 138–154). W. Verlag Bertelsmann GmbH & Co. KG.

Kultusministerkonferenz (Ed.). (2022). Bildungsstandards für das Fach Deutsch Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA). https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Deutsch.pdf

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

Limpo, T., & Alves, R. A. (2018). Tailoring Multicomponent Writing Interventions: Effects of Coupling Self-Regulation and Transcription Training. *Journal of Learning Disabilities, 51*(4), 381–398. https://doi.org/10.1177/0022219417708170

Limpo, T., Alves, R. A., & Connelly, V. (2017). Examining the transcription-writing link: Effects of handwriting fluency and spelling accuracy on writing performance via planning and translating in middle grades. *Learning and Individual Differences, 53*, 26–36. https://doi.org/10.1016/j.lindif.2016.11.004

Limpo, T., & Graham, S. (2020). The role of handwriting instruction in writers' education. British *Journal of Educational Studies, 68*(3), 311–329. https://doi.org/10.1080/00071005.2019.1692127

Lindauer, N., & Sommer, T. (2018). Verfahren der Textbeurteilung. Merkmale und Vorzüge eines holistischen Benchmarkratings. Leseräume, 6(5), 1–14. https://doi.org/10.26041/fhnw-2043

Lindauer, T., & Schmellentin, C. (2017). Rechtschreibung: System, Kompetenzen und Erwerb. In M. Philipp (Ed.), Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben (pp. 20–35). Beltz Juventa.

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

Mason, J. M., & Allen, J. (1986). A Review of Emergent Literacy with Implications for Research and Practice in Reading. In E. Rothkranz (Ed.), Review of research in education (Vol. 13, pp. 27–43). American Educational Research Association. https://doi.org/10.3102/0091732X013001003

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?pq-origsite=gscholar&cbl=18750&diss=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

McMahon, S., & Jones, I. (2015). A comparative judgement approach to teacher assessment. *Assessment in Education: Principles, Policy & Practice, 22*(3), 368–389. https://doi.org/10.1080/0969594X.2014.978839

McNamara, T., Knoch, U., & Fan, J. (2019). Fairness, Justice and Language Assessment. Oxford University Press.

Moats, L., Foorman, B., & Taylor, P. (2006). How Quality of Writing Instruction Impacts High-risk Fourth Graders' Writing. *Reading and Writing, 19*(4), 363–391. https://doi.org/10.1007/s11145-005-4944-6

Moll, K., & Landerl, K. (2009). Double Dissociation Between Reading and Spelling Deficits. *Scientific Studies of Reading, 13*(5), 359–382. https://doi.org/10.1080/10888430903162878

Myford, C. M., & Wolfe, E. W. (2003). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I. *Journal of Applied Measurement, 4*(4), 286–422.

National Commission on Writing. (2003). The neglected "r": The need for a writing revolution. College Entrance Examination Board.

Neumann, A. (2017). Zugänge zur Bestimmung von Textqualität. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), Forschungshandbuch empirische Schreibdidaktik (pp. 203–219). Waxmann.

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. (2014). Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination. *Journal of Writing Research, 6*(2), 173–194. https://doi.org/10.17239/jowr-2014.06.02.4

Palincsar, A. S., & Klenk, L. (1993). Third Invited Response: Broader Visions Encompassing Literacy, Learners, and Contexts. *Remedial and Special Education, 14*(4), 19–25. https://doi.org/10.1177/074193259301400405

Philippek, J., Kreutz, R. M., Hennes, A.-K., Schmidt, B. M., & Schabmann, A. (2024). The contributions of executive functions, transcription skills and text-specific skills to text quality in narratives. *Reading and Writing*, 1–20. https://doi.org/10.1007/s11145-024-10528-5

Pollitt, A. (2012a). Comparative judgement for assessment. *International Journal of Technology and Design Education, 22*(2), 157–170. https://doi.org/10.1007/s10798-011-9189-x

Pollitt, A. (2012b). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice, 19*(3), 281–300. https://doi.org/10.1080/0969594X.2012.665354

R Core Team. (2022). R: A Language and Environment for Statistical Computing [Computer software]. https://www.R-project.org/.

Rocha, R. S., Castro, S. L., & Limpo, T. (2022). The role of transcription and executive functions in writing: A longitudinal study in the transition from primary to intermediate Grades. *Reading and Writing, 35*(8), 1911–1932. https://doi.org/10.1007/s11145-022-10256-8

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education, 34*(2), 159–179. https://doi.org/10.1080/02602930801956059

Santangelo, T., & Graham, S. (2016). A Comprehensive Meta-analysis of Handwriting Instruction. *Educational Psychology Review, 28*(2), 225–265. https://doi.org/10.1007/s10648-015-9335-1

Schwarz, M. (2001). Establishing Coherence in Text. Conceptual Continuity and Textworld Models. *Logos and Language, 2*(1), 15–24.

Sturm, A. (2017). Förderung hierarchieniedriger Schreibprozesse. In M. Philipp (Ed.), Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben (pp. 266–284). Beltz Juventa.

Sturm, A., Nänny, R., & Wyss, S. (2017). Entwicklung hierarchieniedriger Schreibprozesse. In M. Philipp (Ed.), Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben (pp. 84–104). Beltz Juventa.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34*(4), 273–286. https://doi.org/10.1037/h0070288

van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V., & Maeyer, S. de (2019). Validity of comparative judgement to assess academic writing: examining implications of its holistic character and building on a shared consensus. *Assessment in Education: Principles, Policy & Practice, 26*(1), 59–74. https://doi.org/10.1080/0969594X.2016.1253542

van Daal, T., Lesterhuis, M., Maeyer, S. de, & Bouwer, R. (2022). Editorial: Validity, reliability and efficiency of comparative judgement to assess student work. *Frontiers in Education, 7*, Article 1100095. https://doi.org/10.3389/feduc.2022.1100095

Verhavert, S., Bouwer, R., Donche, V., & Maeyer, S. de (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

Weston, J. L., Crossley, S. A., McCarthy, P. M., & McNamara, D. (March 2011). Number of words versus number ideas: Finding a better predictor of writing quality. Twenty-Fourth International FLAIRS Conference, Florida, USA. https://cdn.aaai.org/ocs/2618/2618-11356-1-pb.pdf

Zifonun, G. (2000). Textkonstitutive Funktionen von Tempus, Modus und Genus Verbi. In K. Brinker (Ed.), Text- & Gesprächslinguistik (pp. 315–330). Walter de Gruyter. https://doi.org/10.1515/9783110194067-035