

## Schüler\*innen der Sekundarstufe I

INAUGURALDISSERTATION



Zur Erlangung des Doktorgrades  
der Humanwissenschaftlichen Fakultät  
der Universität zu Köln  
nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Rebecca Maria Kreutz

aus Jülich

Köln, im September 2025

Erstbetreuer der Arbeit: Prof. Dr. Alfred Schabmann (Universität zu Köln)

Zweitbetreuerin der Arbeit: Prof. Dr. Ann-Kathrin Hennes-Schuß (Universität Kassel)

Tag der mündlichen Prüfung: 09.02.2026

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln  
im Januar 2026 angenommen.

### Liste der Publikationen

Die vorliegende monographiebasierte Dissertation mit Teilpublikation (nach §10 Abs. 2b der Promotionsordnung vom 18.12.2018) basiert auf der folgenden Veröffentlichung, die ein *Double-blind Peer Review* Verfahren durchlaufen hat:

**Kreutz, R.M.**, Philippek, J., Hennes, A.-K. & Schabmann, A. (2025). Good text quality despite deficient transcription skills – An analysis of double dissociation. *Journal of Writing Research*, 300–321

Des Weiteren beinhaltet diese Dissertation folgende Studie:

Philippek, J.<sup>1</sup>, **Kreutz, R. M.**<sup>1</sup>, Hennes, A.-K., Schmidt, B. M. & Schabmann, A. (2025). The contributions of executive functions, transcription skills and text-specific skills to text quality in narratives. *Reading and Writing* (38), 651–670.  
<https://doi.org/10.1007/s11145-024-10528-5>

Im Rahmen meiner wissenschaftlichen Tätigkeit wurde außerdem folgende Studie veröffentlicht:

Wolf, J., **Kreutz, R.M.**, Hennes, A.-K. & Schabmann, A. (2025). Relevance of Executive Functions for Writing Competence in Skilled Writers. *Journal of Writing Research*, 300–321

---

<sup>1</sup> Julie Wolf (geb. Philippek) und Rebecca M. Kreutz haben in gleichem Maße zu dieser Arbeit beigetragen und teilen die Erstautorenschaft.

### **Erklärung zu den Publikationen**

Bei Studie Philippek/ Kreutz et al. (2025) handelt es sich um eine geteilte Erstautorenschaft mit Julie Wolf (geb. Philippek), die im Februar 2024 in der Fachzeitschrift *Reading and Writing* angenommen und im April 2024 online veröffentlicht wurde. Julie Wolf und ich waren zu gleichen Teilen an dieser Veröffentlichung beteiligt. Die Konzeption der Studie basiert auf den Überlegungen von Alfred Schabmann, Ann-Kathrin Hennes-Schuß und Barbara Maria Schmidt. Die Daten, die für die Berechnungen dieser Studie verwendet wurden, wurden von Studierenden des Lehrstuhls für Pädagogik und Didaktik des Förderschwerpunkts Lernen unter Supervision von Ann-Kathrin Hennes-Schuß erhoben. Julie Wolf und ich nahmen die Datenaufbereitung sowie die Weiterverarbeitung der Daten vor und generierten zusätzlich weitere Daten aus den Texten der Schüler\*innen unter beratender Mithilfe von Alfred Schabmann, Ann-Kathrin Hennes-Schuß und Barbara Maria Schmidt. Ebenfalls zu gleichen Teilen wurden von Julie Wolf und mir die Berechnungen unter Supervision von Alfred Schabmann vorgenommen. Wir verfassten beide zu gleichen Teilen das Manuskript inklusive Abbildungen und Tabellen. Frühere Versionen wurden mit Ann-Kathrin Hennes-Schuß, Barbara Maria Schmidt und Alfred Schabmann diskutiert. Das finale Manuskript wurde vor der Einreichung von allen Autor\*innen geprüft und bestätigt. Die Revisionen des Artikels fand in Zusammenarbeit mit Julie Wolf statt und in Absprache mit Alfred Schabmann und Ann-Kathrin Hennes-Schuß.

Die Studie Kreutz et al. (2025) wurde von mir als Erstautorin im Juli 2025 in der Fachzeitschrift *Journal of Writing Research* zur Veröffentlichung angenommen und wurde im Oktober 2025 online veröffentlicht. Für diese Studie wurde der gleiche Datensatz wie in der Studie Philippek/ Kreutz et al. (2025) verwendet. Für dieses Projekt wurde teilweise auf die von Julie Wolf und mir aufbereiteten Daten zurückgegriffen. Zusätzlich wurden die Daten von mir in anderer Form aufbereitet und daraus weitere Variablen generiert. Ich habe selbstständig unter Supervision von Alfred Schabmann die Berechnungen durchgeführt. Das

Manuskript wurde von mir verfasst und mit Julie Wolf, Ann-Kathrin Hennes-Schuß und Alfred Schabmann gemeinsam diskutiert. Auf Basis dessen wurde das finale Manuskript von mir angefertigt, welches von allen Co-Autor\*innen zur Einreichung in einem renommierten Journal bestätigt wurde. Die Revisionen fand in Absprache mit Alfred Schabmann statt.

### **Danksagung**

Danke an **Alfred**, meinen Doktorvater, dafür, dass ich dir beim Schreiben und Rechnen unzählige Male über die Schulter gucken durfte sowie für deine wertvolle fachliche Beratung. Durch deine Unterstützung konnte ich mich weiterentwickeln und habe immens viel gelernt.

Danke an **Annki**, meine Zweitbetreuerin, für dein kostbares fachliches Feedback sowie deine motivierenden und aufbauenden Worte, die das ein oder andere Mal nötig waren.

Danke an **Julie** für deine berufliche, seelische und „keksige“ Unterstützung. Ich bin dankbar, dass du meine gesamte Entwicklung von Beginn bis hin zur Angabe hautnah als Bürokollegin begleitet und dabei intensiv geprägt hast. Ich hätte mir mit keiner anderen Person lieber eine Erstautorenschaft geteilt.

Danke an **Barbara** für die zahlreichen Tipps und Tricks, die mich in meiner Rolle als wissenschaftliche Mitarbeiterin maßgeblich weitergebracht haben.

Danke an **Anja** und **Lisa**, die mir stets mit Rat und Tat zur Seite standen – sei es bei Fragen zu Lehrstuhlaufgaben oder zum Promotionsprozess.

Danke an all meine **Kolleg\*innen** für ein stets wertschätzendes und unterstützendes Arbeitsklima sowie die vielen inspirierenden Gespräche bei Kaffee und Keksen.

Danke an meine **Freund\*innen** und **Familie**, die für das „Life“ in der „Work-Life-Balance“ gesorgt haben und damit dazu beigetragen haben, dass ich stets die nötige Kraft für die Arbeit aufbringen konnte.

Danke an meine **Eltern**, die stets mitgefiebert haben. Mit dem Satz „*Der Tag hat 24 Stunden*“ habt ihr mir wohl ungeahnt die wichtigste Coping-Strategie für meine Doktorandenzeit mit auf den Weg gegeben.

Danke an **Alex** für deine unermüdliche Unterstützung in allen Lebenslagen – vom Bewerbungsgespräch bis hin zur Abgabe meiner Dissertation. Ohne dich wäre dieser Weg um einiges steiniger gewesen.



## Inhalt

Liste der Publikationen.....	I
Erklärung zu den Publikationen .....	II
Danksagung .....	IV
Inhalt.....	VI
Abbildungsverzeichnis .....	VIII
Tabellenverzeichnis.....	IX
1    Einleitung .....	10
2    Textproduktionskompetenz .....	12
2.1    Relevanz der Textproduktionskompetenz .....	14
2.2    Bewertungsmethoden der Textproduktionskompetenz .....	15
3    Modellannahmen des Schreibens .....	20
3.1    Das Schreibprozessmodell von Hayes and Flower (1980).....	20
3.2    Das kognitive Schreibmodell von Hayes (1996).....	22
3.3    Das Kognitive Schreibmodell von Hayes (2012).....	24
3.4    Simple View of Writing (1980-2006) .....	25
3.5    Direct and Indirect Effects Model of Writing (2017-2022) .....	28
3.6    Modellannahmen über den Entwicklungsverlauf des Schreibens .....	31
3.7    Möglichkeiten und Grenzen bisheriger Modelle.....	35
4    Teilfähigkeiten der Textproduktionskompetenz .....	38
4.1    Hierarchieniedrigere Transkriptionsfähigkeiten.....	39
4.2    Hierarchiehohe textspezifische Fähigkeiten.....	42

4.2.1	Textspezifisches aktives Vokabelwissen .....	43
4.2.2	Kohärenz .....	45
4.3	Kognitive Funktionen.....	47
4.3.1	Exekutivfunktionen .....	47
5	Philippek/ Kreutz et al. (2025): Forschungsfragen und Untersuchungsaufbau .....	51
6	The Contributions of Executive Functions, Transcription Skills and Text-specific Skills to Text Quality in Narratives.....	54
7	Philippek/ Kreutz et al. (2025): Zwischenfazit .....	79
8	Doppelte Dissoziationen .....	81
9	Kreutz et al. (2025): Forschungsfragen und Untersuchungsaufbau .....	83
10	Good text quality despite poor transcription skills – An analysis of double dissociation	87
11	Kreutz et al. (2025): Zwischenfazit .....	114
12	Diskussion.....	115
12.1	Limitationen .....	116
13	Fazit und Ausblick.....	118
13.1	Implikationen für die Praxis .....	118
13.2	Weitere Forschungsvorhaben .....	120
	Anhang A .....	127
	Anhang B.....	128
	Literaturverzeichnis.....	129

## Abbildungsverzeichnis

Abbildung 1: <i>Das Schreibprozessmodell (eigene Abbildung nach Hayes und Flower, 1980)</i> .....	22
Abbildung 2: <i>Das kognitive Schreibmodell (eigene Abbildung nach Hayes, 1996)</i> .....	23
Abbildung 3: <i>Das kognitive Schreibmodell (eigene Abbildung nach Hayes, 2012)</i> .....	25
Abbildung 4: <i>(Not-so-)Simple View of Writing (eigene Abbildung nach Berninger &amp; Winn, 2006)</i> .....	27
Abbildung 5: <i>Direct and Indirect Effects Model of Writing (eigene Abbildung nach Kim &amp; Park, 2019)</i> .....	30
Abbildung 6: <i>Expanded Direct and Indirect Effects Model of Writing (eigene Abbildung nach Kim &amp; Graham, 2022)</i> .....	31
Figure 1 <sup>2</sup> : <i>The Cascaded Model of Writing</i> .....	61
Figure 2 <sup>2</sup> : <i>Structural equation model of the Cascaded Model of Writing (1B)</i> .....	71
Figure 3 <sup>2</sup> : <i>Structural equation model of the Cascaded Model of Writing (2C)</i> .....	72
Figure 1 <sup>3</sup> : <i>Spelling/ Handwriting Fluency vs. Text Quality Double Dissociation Groups</i> .....	101
Figure 2 <sup>3</sup> : <i>ANOVA Statistics of Performance Profiles of Spelling and Text-Specific Formulation Skills</i> .....	104
Figure 3 <sup>3</sup> : <i>ANOVA Statistics of Performance Profiles of Handwriting Fluency and Text-Specific Formulation Skills</i> .....	106
Abbildung 7: <i>Cascaded Model of Writing Expanded</i> .....	120
Abbildung 8: <i>Das Kovergenzmodell (Hennes et al., 2018)</i> .....	125

---

<sup>2</sup> Abbildung aus Studie Philippek/ Kreutz et al. (2025)

<sup>3</sup> Abbildung aus Studie Kreutz et al. (2025)

**Tabellenverzeichnis**

Tabelle 1: <i>Schreibentwicklungsmodell nach Bereiter (1980)</i> .....	33
Table 1 <sup>4</sup> : <i>Descriptive Statistics</i> .....	70
Table 2 <sup>4</sup> : <i>Correlations of Variables</i> .....	70
Table 1 <sup>5</sup> : <i>Means and Standard Deviations for all Variables of the Study</i> .....	99
Table 2 <sup>5</sup> : <i>Correlations of Variables</i> .....	100
Table 3 <sup>5</sup> : <i>CFA: Overlap of Students with Poor/ Good Performance in Text Quality, Handwriting Fluency, and Spelling</i> .....	102
Table 4 <sup>5</sup> : <i>Performance on Text-Specific Formulation Skills. Discrepant and Nondiscrepant Groups Based on Spelling</i> .....	103
Table 5 <sup>5</sup> : <i>Performance on Text-Specific Formulation Skills. Discrepant and Nondiscrepant Groups Based on Handwriting Fluency</i> .....	105
Table 3 <sup>4</sup> : <i>Overview of Measurement in Cited Studies</i> .....	124

---

<sup>4</sup> Tabelle bzw. Anhang aus Studie Philippek/ Kreutz et al. (2025)

<sup>5</sup> Tabelle aus Studie Kreutz et al. (2025)

## 1 Einleitung

Die Produktion von Texten stellt eine Schlüsselkompetenz in Schule, Beruf und Alltag dar (Becker-Mrotzek et al., 2014; Crossley & McNamara, 2016; Feenstra, 2014). Denn erst durch das Schreiben von Texten entwickeln Schüler\*innen die Fähigkeit, ihre Gedanken zu ordnen und ihre Argumentation logisch aufzubauen. Es trägt dazu bei, das Verständnis von Konzepten zu vertiefen und Wissen dauerhaft im Gedächtnis zu speichern und spielt somit im fachlichen Lernprozess und für den Bildungserfolg eine zentrale Rolle (Forkarth & Manzel, 2024). Darüber hinaus stellt die Fähigkeit, Texte zu schreiben, eine wesentliche Grundlage für die Teilhabe an der Gesellschaft dar, vor allem seitdem E-Mail und Messenger zu den vorherrschenden Kommunikationsmitteln zählen (Becker-Mrotzek & Böttcher, 2012). Trotz dieser eindeutigen Relevanz der Schreibkompetenz, sind viele Schüler\*innen und Erwachsene nicht in der Lage, funktionale Texte zu produzieren (Graham et al., 2023; Neumann & Lehmann, 2008). Ein funktionaler Text zeichnet sich dabei durch seine orthografische Korrektheit, seine lexikalische Vielfalt, Kohärenz, Adressatenorientierung sowie der Verwendung eines angemessenen Textmusters in Abhängigkeit der Schreibaufgabe aus (Becker-Mrotzek et al., 2014; MacArthur et al., 2019; Olinghouse & Wilson, 2013). Angesichts dieser Merkmale funktionaler Texte haben empirische Studien verstärkt die Rolle verschiedener Teilfähigkeiten analysiert, die als zentral für die Produktion qualitativ guter Texte gelten. Hierzu zählen neben kognitiven Prozessen wie den Exekutivfunktionen auch hierarchieniedrige Transkriptionsfähigkeiten wie die Schreibflüssigkeit und die Rechtschreibung (Sturm, 2017). Darüber hinaus tragen zahlreiche hierarchiehohe Fähigkeiten – wie die Fähigkeit zur Herstellung von Kohärenz oder die Anwendung eines vielfältigen Wortschatzes – zur Produktion funktionaler Texte bei (z.B. Crossley & McNamara, 2016; Gómez Vera et al., 2016; Olive & Kellogg, 2002). Es konnte durch die Erstellung von Schreibmodellen gezeigt werden, dass die Textproduktion eine Interaktion dieser Fähigkeiten beinhaltet (Berninger & Winn, 2006; Kim & Graham, 2022; Kim & Park, 2019). Dabei lag

der Fokus oftmals auf der Untersuchung des Einflusses hierarchieniedriger Transkriptionsfähigkeiten sowie kognitiver Fähigkeiten auf die Textqualität. Trotz bestätigten isolierten Einflusses der hierarchiehöheren schreibspezifischen Fähigkeiten auf die Textqualität, blieben diese bislang in Schreibmodellen unterrepräsentiert. Ziel dieser Arbeit ist es daher, ein tiefergehendes Verständnis über die zugrundeliegenden Interaktionen der hierarchieniedrigen und hierarchiehohe Teilfähigkeiten der Textproduktion sowie über ihren Einfluss auf die Produktion funktionaler Texte gemessen anhand der Textqualität zu erlangen. Hierzu wurden zwei empirische Studien durchgeführt. Die Studie Philippek/ Kreutz et al. (2025) modelliert die Schreibkompetenz unter Berücksichtigung von Teilfähigkeiten verschiedener hierarchischer Ebenen, die sich in Isolation bereits als wichtige Einflussfaktoren für die Produktion guter Texte erwiesen haben (Kapitel 6). Der Fokus der Studie Kreutz et al. (2025) liegt anschließend auf einer differenzierten Analyse des Zusammenhangs von Transkriptionsfähigkeiten und Textqualität bei Schüler\*innen mit unterschiedlichen Leistungsausprägungen in diesen beiden Bereichen (Kapitel 10). Bisherige Studien zur Modellierung des Schreibens widmeten sich hauptsächlich Grundschüler\*innen oder erwachsenen Schreibenden (Kapitel 3). Die beiden im *Peer-Review* Verfahren veröffentlichten Studien befassen sich mit Kinder der 5.-7. Klasse, die sich in einer zentralen Entwicklungsphase des Schreibens befinden, in der Transkriptionsfähigkeiten bereits nahezu automatisiert sind und hierarchiehohe Fähigkeiten zunehmend an Bedeutung gewinnen.

Da die Studien vornehmlich im schulischen Kontext messbare Teilfähigkeiten der Textproduktion untersuchen, lassen sich auf Basis der Resultate wichtige Implikationen für den Schreibunterricht ableiten. Die Ergebnisse können folglich genutzt werden, um eine zielgerichtete Förderung des Schreibens anbieten zu können (Kapitel 13.1). Ebenso zeigen die beiden Studien auf, welche Aspekte in der Forschung noch unbeantwortet bleiben und geben somit ebenfalls wichtige Implikationen für die Wissenschaft (Kapitel 13.2).

## 2 Textproduktionskompetenz

„Kompetenz“ ist aufgrund seines weitverbreiteten Gebrauchs innerhalb verschiedener Fachrichtungen ein umfangreicher Begriff. Allgemein lässt sich Kompetenz definieren als

*die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können. (Weinert, 2001, S. 27)*

Die empirische Bildungsforschung befasste sich seit der empirischen Wende, in der eine fundamentale Veränderung in der politischen und didaktischen Reflexion von Schule und Unterricht in Deutschland stattfand, im Zuge der PISA-Studien intensiv damit, den Begriff sowohl theoretisch als auch empirisch für das eigene Forschungsgebiet näher zu definieren (Gniewosz, 2011). Viele Wissenschaftler\*innen versuchten sich seither an einer konkreten Definition der Schreibkompetenz. Da es sich dabei jedoch um ein vielschichtiges Konstrukt handelt, dessen vielfältige Einflussfaktoren bisher noch nicht vollständig erforscht sind, existiert bis heute keine einheitliche Definition. Einigkeit besteht jedoch darüber, dass sich das Schreiben im Sinne der Textproduktion deutlich vom Sprechen unterscheidet und mehr als pures Aufschreiben von einzelnen Wörtern, Buchstaben oder Sätzen beinhaltet. Schreiben wird hier im Kontext der Kompetenz immer als komplexe Fähigkeit verstanden und meint die Produktion eines funktionalen Textes.

In der Forschung lassen sich verschiedene Ansätze finden, die die zur Textproduktion notwendigen Teilfähigkeiten unterschiedlich definieren und gewichten. Sobald ein Schreibziel besteht, wird Schreiben zu einem zielgerichteten Prozess, dem sowohl hierarchieniedrige und hierarchiehohe Fähigkeiten unterliegen, welche in einem schulischen Vermittlungsprozess mühsam angeeignet werden müssen (Merz-Grötsch, 2022). Die Schreibkompetenz hebt sich somit von den hierarchieniedrigen Fähigkeiten der Rechtschreibung und Schreibflüssigkeit ab, die zwar ein integraler Bestandteil der Textproduktion und somit der Schreibkompetenz sind, jedoch nicht deren entscheidende Komponenten darstellen. So unterscheidet sich die

Textproduktion eines Grundschuldkindes, das sich vorrangig noch auf die Schreibflüssigkeit und die Rechtschreibung konzentrieren muss, deutlich von der eines geübten Schreibenden, bei dem die Transkription vollständig automatisiert verläuft (Merz-Grötsch, 2022). Die Schreibkompetenz bei geübten Schreibenden umfasst die Fähigkeit, einen Text, in welchem alle notwendigen Informationen gegeben werden, in einem angemessenen Schreibstil und zusammenhängend zu produzieren (National Assessment of Educational Progress [NEAP], 2011). Doch da Schreiben immer eine kommunikative Funktion übernimmt und in einen sozialen Handlungskontext eingebunden ist, zählt zur Schreibkompetenz ebenso die pragmatische Angemessenheit (Becker-Mrotzek et al., 2014). Je nach Textgenre und Schreibziel werden an die Schreibenden verschiedene Herausforderungen gestellt wie beispielsweise Sachverhalte erklären oder Unterhaltung bieten. So gehören ebenfalls die Fähigkeiten, die Adressat\*innen präzise zu informieren, überzeugend zu argumentieren oder Wörter kreativ auszuwählen, zur Schreibkompetenz (Becker-Mrotzek & Böttcher, 2012; Harsch et al., 2007; Kultusministerkonferenz [KMK], 2022). Darüber hinaus muss ein Schreibender auch inhaltliches Wissen, Textstrukturwissen und Sprachwissen miteinander verknüpfen (Fix, 2010). Zusammenfassend lässt sich aus den verschiedenen sogenannten Teilfähigkeiten ableiten, dass die Schreibkompetenz nicht einfach durch ein vordefiniertes Set von Fähigkeiten beschrieben werden kann. Vielmehr handelt es sich um ein Kompetenzbündel verschiedener Teilfähigkeiten, die während des Schreibprozesses und abhängig vom Schreibziel simultan agieren (Becker-Mrotzek et al., 2014; Becker-Mrotzek & Schindler, 2007; Crossley et al., 2014; Kim et al., 2013). Welche Fähigkeiten zur Schreibkompetenz beitragen, kann je nach Textgenre deutlich variieren (Beers & Nagy, 2009; Kim & Graham, 2022). Im Folgenden wird in dieser Arbeit der Fokus auf die Produktion von Narrativen gelegt und die dafür relevanten Fähigkeiten näher betrachtet, da dieses Genre in der Primar- sowie Sekundarstufe I einen Schwerpunkt in den Lehrplänen einnimmt (Dockrell et al., 2015; MSB NRW, 2019). Es handelt sich außerdem um das erste Genre, das in der



Schule thematisiert wird, da es aufgrund seines chronologischen Aufbaus und der Handlungsorientierung für Schüler\*innen leichter zu produzieren ist als beispielsweise expositorische Texte, die diskursive Fähigkeiten erfordern. Hinzu kommt, dass die Themen in Narrativen häufiger an die Lebenswelt der Schüler\*innen anknüpfen und somit eine gewisse inhaltliche Vertrautheit besteht (Scott & Windsor, 2000).

Während Schreibkompetenz (Englisch: writing competence) zwar in diesem Forschungsbereich die geläufigere Bezeichnung ist, wird im Folgenden der synonyme Begriff der Textproduktionskompetenz (Englisch: composing competence) verwendet, da hieraus deutlicher hervor geht, dass sich auf die Produktion von Texten bezogen wird und beispielsweise nicht auf die Produktion einzelner Sätze. Diesen Unterschied hervorzuheben ist wichtig, da die Produktion von Texten den Schreibenden nochmal vor größere Herausforderungen stellt, da hierbei deutlich mehr Teilfähigkeiten Verwendung finden, die notwendig sind, um beispielsweise die einzelnen Sätze zu verbinden und einen stimmigen und kohärenten Text produzieren zu können.

### **2.1 Relevanz der Textproduktionskompetenz**

Die Textproduktionskompetenz stellt einen wichtigen Prädiktor für den schulischen und beruflichen Erfolg dar. Darüber hinaus ist sie auch im sozialen Miteinander von äußerster Relevanz, da das Verfassen von Texten in vielen Lebensbereichen erforderlich ist (Becker-Mrotzek et al., 2014; Crossley & McNamara, 2016; Feenstra, 2014; MacArthur et al., 2016; T. McNamara et al., 2019). Bereits zu Beginn der Schulzeit spielt die Fähigkeit zur Textproduktion eine zentrale Rolle – sowohl im Lernprozess selbst als auch bei der Darstellung des Erlernenen. Zum einen ist die Textproduktionskompetenz an sich Gegenstand schulischer Bildung (Becker-Mrotzek, 2022) und deren Vermittlung sollte somit die Hauptaufgabe des Schreibunterrichts sein (Graham et al., 2023). Zum anderen ist sie aber auch wichtig für die Aneignung von Wissen in sämtlichen (schulischen) Bereichen, da Schreiben es ermöglicht, Wissen zu entwickeln, zu organisieren und mit anderen zu teilen

(Schmölzer-Eibinger & Thürmann, 2015). Beherrschen die Schüler\*innen die Produktion von Texten nicht, sinkt auch die Chance, nach der Schule ein Studium zu absolvieren, da dort die Leistungsüberprüfung hauptsächlich schriftlich erfolgt (Graham & Perin, 2007a). Außerdem erfordern heutzutage die meisten Berufe, die ein existenzsicherndes Einkommen gewährleisten, eine sichere Textproduktion, da in einer Vielzahl von Berufen von den Mitarbeitenden die Anfertigung schriftlicher Unterlagen, Präsentationen und Berichte erwartet wird. Zudem ist eine gesellschaftliche Teilhabe von Erwachsenen, die nicht schriftlich kommunizieren können, kaum möglich, seitdem E-Mails das Telefon als bevorzugtes Kommunikationsmittel ersetzen (Graham & Perin, 2007a). Die Förderung der Textproduktionskompetenz ist daher essentiell, wenn es um soziale Gerechtigkeit und Chancengleichheit geht (Harris, 2024).

### **2.2 Bewertungsmethoden der Textproduktionskompetenz**

Trotz der hohen Relevanz der Textproduktionskompetenz stellt deren Bewertung keine triviale und einheitlich gehandhabte Vorgehensweise dar (N. Lindauer, 2024). Sowohl in der Forschung (Kim et al., 2017; Schoonen, 2012) als auch im schulischen Kontext (Birkel & Birkel, 2002; N. Lindauer, 2024) stellt die Messung der Textproduktionskompetenz eine große Herausforderung dar. In beiden Kontexten sollen die Endprodukte der Schreibenden gemäß der psychometrischen Gütekriterien bewertet werden, um eine gerechte und nachvollziehbare Urteilsfindung zu gewährleisten. Die Bewertung dient dabei verschiedenen Zwecken, darunter die Evaluation von Lehr- und Unterrichtserfolgen (nicht nur im Deutschunterricht), das Feedback für Schreibende zur Optimierung ihrer Textproduktion sowie die Beantwortung wissenschaftlicher Forschungsfragen (Peschel, 2024). Die Schwierigkeit einer angemessenen Bewertung der Textproduktionskompetenz sowie mögliche Lösungsansätze werden bereits seit Jahren intensiv diskutiert (Neumann, 2017). Auch wenn ein gewisses intuitives Verständnis darüber besteht, was einen qualitativ guten Text ausmacht, lassen sich zugrundeliegende Bewertungskriterien nur schwer definieren. Sowohl in der

Bildungsforschung als auch im schulischen Kontext besteht jedoch Einigkeit darüber, dass die Textqualität als ein geeigneter Indikator für die zugrundeliegende Textproduktionskompetenz herangezogen werden sollte (Grabowski et al., 2014). Dies löst jedoch nicht das Problem der Bewertung, da hierbei einige Herausforderungen bestehen: Im Gegensatz zum Lesen gibt es bei der Produktion von Texten keinen vordefinierten Zielzustand, anhand dessen das Endergebnis abgeglichen werden kann. Denn beim Schreiben von Texten handelt es sich um eine produktive Aufgabe (Neumann, 2007), die mehrere individuelle Lösungen zulässt (Feenstra, 2014). Eine weitere Herausforderung besteht darin, dass die Textproduktionskompetenz, wie in Kapitel 2 beschrieben, ein latentes Konstrukt ist, das durch eine Kombination mehrerer Merkmale zustande kommt, welche bei der Messung mitunter gleichzeitig bewertet werden müssen. Dabei lässt die bewertende Person häufig ihre persönlichen Vorstellungen eines guten Textes in die Bewertung einfließen, was zu sehr unterschiedlichen Bewertungen verschiedener Personen bei dem gleichen Text führen kann (Bouwer et al., 2023). Ebenfalls ist denkbar, dass die Bewertenden bestehende Kriterien unterschiedlich interpretieren bzw. sie nicht separat voneinander betrachten können (Cooksey et al., 2007; Eckes, 2008). So ist die Beurteilung der orthografischen Korrektheit zwar ein hoch reliables Maß, das jedoch andere Aspekte bei der Gesamtbeurteilung maskieren könnte (Mathiebe et al., 2024). Ein weiterer potenzieller Einflussfaktor bei der Bewertung mehrerer Texte ist die Reihenfolge, in der diese beurteilt werden. Der sogenannte „Reihenfolgeeffekt“ beschreibt das Phänomen, dass die Bewertung eines Textes durch den direkten Vergleich mit zuvor gelesenen Texten beeinflusst wird und die Reihenfolge somit zu einer systematischen Verzerrung der Beurteilung führen kann (Hales & Tokar, 1975; Hughes et al., 1980). Diese Probleme treten nicht nur bei der Textbewertung durch Lehrkräfte auf (z. B. Birkel & Birkel, 2002; Skar & Jølle, 2017), sondern auch innerhalb des Forschungskontextes (Hennes, 2020).

Es gibt verschiedene Ansätze, die versuchen, sich einer validen, reliablen und objektiven Messung anzunähern. Dabei wird grundsätzlich zwischen analytischen und

holistischen Ratingverfahren unterschieden, wobei der Übergang zwischen den Methoden teilweise fließend ist (Schipolowski & Böhme, 2016). Innerhalb des analytischen Ratings wird ein Text anhand einzelner Dimensionen in Form von verschiedenen Items bewertet. Die Items innerhalb sogenannter Analyseraster beziehen sich dabei beispielsweise auf die Struktur, den Inhalt und auf die sprachliche Gestaltung eines Textes und werden meist dichotom oder auf einer Skala eingeschätzt (Böhme et al., 2009). Für diese Bewertungsmethode ist ein item- statt textgeleitetes Vorgehen empfehlenswert, da so der Gesamteindruck des Textes die Bewertung nicht beeinflusst (N. Lindauer & Sommer, 2018). Vorteile dieses Ansatzes sind, dass zum einen ein differenzierteres Bild des Textes erstellt werden kann und zum anderen somit die Möglichkeit besteht, differenziertes Feedback zu geben (Neumann, 2017). Dennoch bleibt die Beurteilung mittels dieser Raster vage, weil sie häufig keine differenzierte Beschreibung der Items beinhalten (Sturm et al., 2018). Zudem ist die Beurteilung aufwändig, da die Texte wiederholt zur Bewertung herangezogen werden müssen (N. Lindauer, 2024).

Eine alternative Methode zur Evaluation der Textqualität basiert auf holistischen Bewertungen. Dabei handelt es sich um Globalurteile eines gesamten Textes. Meist wird zur Bewertung eine mehrstufige Skala herangezogen (Böhme et al., 2009). Anders als bei analytischen Ratings lassen diese Skalen deutlich mehr Beurteilungsspielraum zu, da die Items weniger aufgabenspezifisch sind. Dies hat zum Nachteil, dass es mit höherer Wahrscheinlichkeit zu unterschiedlicheren Urteilen kommt, bringt aber den Vorteil mit sich, dass Aspekte, die eventuell nicht im Kriterienkatalog explizit aufgeführt werden, dennoch in die Bewertung miteinbezogen werden können (Weigle, 2002). Dies könnte die Textproduktionskompetenz angemessener abbilden als aufsummierte einzelne Items (N. Lindauer, 2024). Um die Vorteile der holistischen Bewertung zu nutzen und gleichzeitig das Problem der unterschiedlichen Urteile zu beheben, kann diese Form der Bewertung durch sogenannte Benchmarktexte ergänzt werden. Dabei handelt es sich um ein globales Rating

unter Hinzunahme von Beispieltexten, die verschiedene Qualitätsstufen abbilden und mit den zu bewertenden Texten in ihrer Qualität verglichen werden können (Bouwer et al., 2024; N. Lindauer & Sommer, 2018). Dadurch soll intra- und interpersonellen Unterschieden in der Bewertung der Texte vorgebeugt werden (N. Lindauer, 2024). Vorteil dieser Variante ist die zeitliche Ökonomie durch die schnelle Handhabbarkeit (N. Lindauer & Sommer, 2018; Neumann, 2017). Ein Nachteil dieser Methode ist der relativ hohe Aufwand der Konstruktion der Ratingskala mit Benchmarktexten, die mit einem Pre-Ratingverfahren mit mehreren Bewertenden verbunden ist (N. Lindauer, 2024).

Zahlreiche Studien stellten Vergleiche zwischen den Verfahren bezüglich der Gütekriterien an. Trotz vermindertem Interpretationsspielraum beim analytischen Rating, zeigen vergleichende Studien, dass bezüglich der Beurteiler-Reliabilität keins der drei Ratingverfahren dem anderen überlegen ist. Dabei war es irrelevant, ob die Bewertenden trainiert oder untrainiert waren (Grabowski et al., 2014). Hinsichtlich der Validität zeigt sich, dass bei allen Verfahren zwingend mehrere Texte pro Schüler\*in von verschiedenen Bewertenden beurteilt werden müssen, um aussagekräftige Ergebnisse zur Textproduktionskompetenz zu erhalten. Die derzeitige Studienlage deutet aber auf einen leichten Vorteil der holistischen Ratings mit Benchmarktexten hin (Feenstra, 2014; N. Lindauer & Sommer, 2018; van den Bergh et al., 2012). Trotz zufriedenstellender Ergebnisse hinsichtlich der Gütekriterien zeigt die Bewertungsforschung jedoch, dass Beurteilende dazu neigen, überwiegend den mittleren Bereich von Ratingskalen zu nutzen, während Extremwerte selten vergeben werden. Dies führt bei allen drei Verfahren zu einer Verteilung mit zentraler Tendenz anstelle einer Normalverteilung (N. Lindauer, 2024).

Daher rückte in der Forschung eine andere ebenfalls ressourcenschonende Form der Textqualitätsbewertung in den Vordergrund: Das Comparative Judgement. Übersetzt bedeutet es „vergleichende Urteile“ und basiert auf ganzheitlichen Beurteilungen durch eine Gruppe von Bewertenden, die als Einzelpersonen voneinander unabhängige Urteile fällen. Aufgabe

der Bewertenden ist es, in mehreren Runden jeweils zwei Texte global miteinander zu vergleichen und zu entscheiden, welcher der beiden Texte der bessere ist. Für jede Paarung wird somit ein Text als „Verlierer“ und der andere Text als „Gewinner“ gekennzeichnet, was in einer binären Entscheidungsmatrix resultiert (McMahon & Jones, 2015). Auf Basis eines logistischen Modells kann infolgedessen eine Rangfolge über alle Texte hinweg erstellt werden von der besten bis zur schlechtesten Textqualität (Lesterhuis et al., 2017).

Die Grundidee hinter dieser Aufgabenart stammt von Thurstone (1927), der argumentierte, dass Stimuli durch Vergleiche beurteilt werden können, wenn sie ein Merkmal teilen, das sich einordnen lässt (Boonen et al., 2020). Studien haben gezeigt, dass die Methode des Comparative Judgements bei der Bewertung der Textqualität eine hohe Zuverlässigkeit und Konsistenz aufweist und somit die Möglichkeit bietet, vorherrschende Probleme bei der Bewertung der Textqualität zu überwinden (Lesterhuis et al., 2017). Die Anzahl der Beurteilungen eines Textes ist dabei ein wichtiger Faktor. Es konnte gezeigt werden, dass nach 20 Vergleichen pro Text, die Zuverlässigkeit der Rangliste einen Höchststand erreichte (Verhavert et al., 2019). Ein großer Vorteil für die Forschung ist, dass auch unerfahrene Bewertende mit dieser Methode zuverlässige Ergebnisse erzielen können (Boonen et al., 2020). Daher wurde in den dargestellten Studien der vorliegenden Arbeit das Comparative Judgement Verfahren zur Erfassung der Textqualität genutzt.

Es sei jedoch darauf hingewiesen, dass sich dieses Verfahren weniger für den schulischen Bereich eignet, da für ein reliables Ergebnis eine gewisse Anzahl an Bewertenden benötigt wird. Des Weiteren lässt sich hieraus lediglich ein Ranking der eingeschlossenen Texte von gut nach schlecht aufstellen, was jedoch nicht zu einer Förderindikation führt.

### 3 Modellannahmen des Schreibens

Ogleich die Messung der Textqualität ein großer Forschungsaspekt ist, fokussiert sich die Schreibforschung gleichermaßen auf die Ergründung der Faktoren, die die Textqualität beeinflussen. Zu diesem Zweck werden in der Forschung Modellannahmen genutzt. Sie dienen dazu, komplexe Phänomene zu verstehen, Hypothesen zu testen, Vorhersagen zu treffen und wissenschaftliche Erkenntnisse zu kommunizieren. Wie in den vorausgehenden Kapiteln beschrieben, stellt die Textproduktionskompetenz ein komplexes Konstrukt dar, dessen Teilfähigkeiten noch nicht gänzlich erfasst werden konnten. Gerade in der Schreibforschung wird deshalb versucht, das Konstrukt und dessen zugrundeliegende Lernprozesse anhand von Modellen zu erklären, um weiterführend Diagnostik- oder Fördermöglichkeiten zu entwickeln. Die Schreibforschung mit ihren Modellen ist noch nicht alt. Erst in den letzten vier Jahrzehnten haben sich Wissenschaftler\*innen mit den Einflussfaktoren und deren Interaktion bei der Textproduktion sowie mit dem Schreibprozess beschäftigt und dabei verschiedenste Modellansätze veröffentlicht. Die ersten Modelle wurden in den 1980er Jahren entwickelt, während die aktuellsten Modelle aus diesem Jahrzehnt stammen. Dies verdeutlicht, dass die Modellierung der Textproduktionskompetenz noch längst nicht abgeschlossen ist. Im Folgenden werden die bislang bekanntesten Modelle chronologisch dargestellt und inhaltlich näher erläutert.

#### 3.1 Das Schreibprozessmodell von Hayes and Flower (1980)

In den 1970er Jahren fand innerhalb der Schreibforschung eine Verlagerung des Fokus vom Schreibprodukt auf den Schreibprozess statt. Diese Verlagerung basierte auf Erkenntnissen der Kognitionsforschung, die postulierte, den Schreibprozess mit allen Überarbeitungsschritten und Teilfähigkeiten sowie deren Interaktion zu betrachten (Hayes & Flower, 1980, 1986). Das daraus resultierende wohl bekannteste und gleichzeitig eines der ersten Schreibmodelle ist das von Hayes und Flower aus dem Jahre 1980 (Abbildung 1). Das Modell versteht die Textproduktion als eine Art Problemlöseprozess, bei dem unter

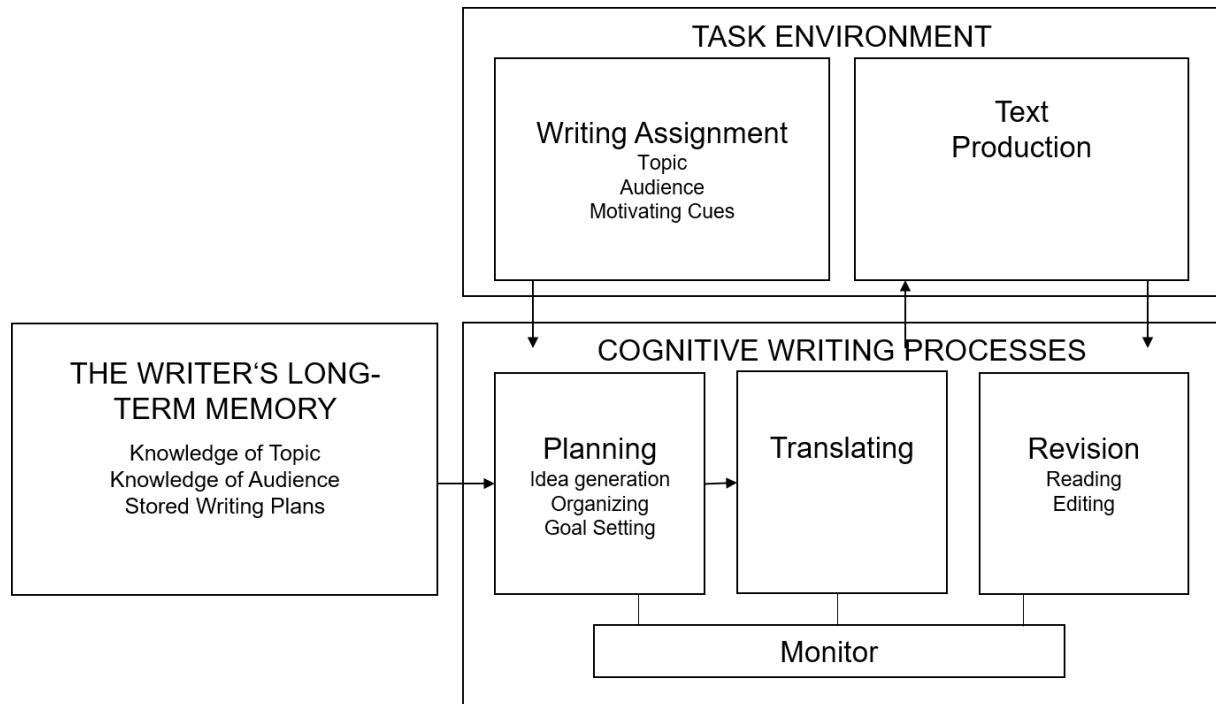
Verwendung verschiedener kognitiver und sprachlicher Ressourcen ein Text sukzessiv produziert wird. Erstellt wurde das Modell daher mit sogenannten „Lautes-Denken-Protokollen“, bei denen erwachsene Schreibende aufgefordert wurden, alles zu äußern, was ihnen beim Schreiben in den Sinn kam. Resultierend daraus teilten Hayes und Flower den Schreibprozess in drei Elemente ein: der *Schreibprozess* an sich (cognitive writing processes), das *Aufgabenumfeld* (task environment) und das *Langzeitgedächtnis* (long-term memory) des Schreibenden. Das *Aufgabenumfeld* beinhaltet alle externen Informationen über die Schreibaufgabe, wie z.B. Thema, Adressaten und Ressourcen des Schreibenden, wie etwa Notizen sowie den Schreibanlass. Dieser Teil bleibt im Verlauf des gesamten Prozesses stabil, während sich das Schreibprodukt kontinuierlich ändert. Um die Schreibaufgabe zu bewältigen, greift das *Langzeitgedächtnis* auf Vorwissen bezüglich des Schreibthemas (content knowledge), auf Textmusterwissen (discourse process knowledge), auf Faktenwissen (declarative knowledge) sowie auf Informationen über Orthografie, Syntax und Textgestaltungsmittel wie Kohärenz (procedural knowledge) zurück. Das *Aufgabenumfeld* und das *Langzeitgedächtnis* stellen die Rahmenbedingungen für den *Schreibprozess* dar, welcher wiederum aus drei verschiedenen Einzelprozessen – dem *Planen* (planning), dem *Übersetzen* (translating) und dem *Überarbeiten* (revision) – besteht. *Planen* involviert das Generieren von Ideen und deren Organisation sowie die Zielsetzung. Diese Phase ist noch nicht mit Schrift assoziiert. Erst in der darauffolgenden Phase des *Übersetzens*, in welcher die generierten Ideen schriftsprachlich verfasst werden, findet eine Assoziation mit Schrift statt. In dieser Phase erfolgt vor allem der Rückgriff auf das *Langzeitgedächtnis*. Innerhalb des *Überarbeitungsprozesses* wird der Text immer wieder gelesen und angepasst. Dieser Prozess verläuft automatisch, was zur Folge hat, dass andere Prozesse bei der Entdeckung von Fehlern kontinuierlich unterbrochen werden. Der gesamte *Schreibprozess* verläuft rekursiv, d.h., dass die drei zentralen Prozesse Planen, Übersetzen und Überarbeiten wiederholt im Wechsel stattfinden, bis das Ziel des Schreibenden erfüllt ist. Die Steuerung, wann zwischen den



einzelnen Schritten gewechselt werden muss und wann der Schreibprozess beendet werden kann, übernimmt dabei der sogenannte Monitor (Hayes & Flower, 1980) (Abbildung 1).

### Abbildung 1

*Das Schreibprozessmodell (eigene Abbildung nach Hayes und Flower, 1980)*



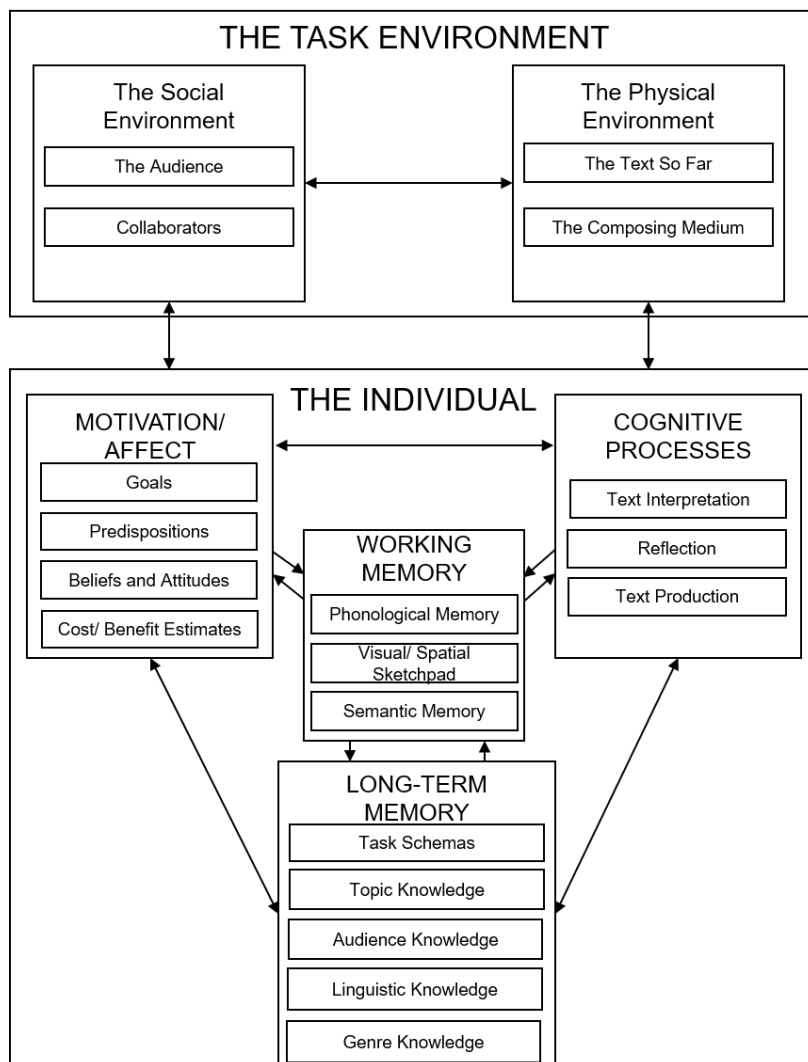
### 3.2 Das kognitive Schreibmodell von Hayes (1996)

Weitere auf dem Schreibprozessmodell von Hayes und Flower (1980) basierende Forschung führte zur Entwicklung des kognitiven Schreibmodells von Hayes (Abbildung 2), welches als individuell-umweltbezogenes Modell bezeichnet werden kann, da in diesem Modell viele Einflussfaktoren auf Ebene der Umwelt oder des Individuums von Relevanz sind (Hayes, 1996). Das Modell unterscheidet sich vom Schreibprozessmodell (Hayes & Flower, 1980) hauptsächlich in vier Aspekten: Neben dem *Langzeitgedächtnis* (long-term memory), das hierin spezifiziert wurde, spielt hier auch das *Arbeitsgedächtnis* (working memory), das für die Verarbeitung der einzelnen Dimensionen zuständig ist, eine wichtige Rolle im Schreibprozess. Außerdem wurde dem Modell die *Schreibmotivation* sowie *affektive Faktoren* (motivation/affect) hinzugefügt. Die *kognitiven Prozesse* (cognitive processes) wurden um weitere Forschungsergebnisse ergänzt und basierend darauf umbenannt in

generellere Termini: Planen wurde in Reflexion umbenannt und enthielt fortan kognitive Prozesse des Problemlösens und der Entscheidungsfindung. Übersetzen wurde durch Textproduktion ersetzt, bei der die interne Vorstellung in ein schriftliches Produkt umgewandelt wird. Überarbeiten wurde umbenannt zu Textinterpretation, wozu das Lesen, Zuhören und Interpretieren von visuell-räumlichen und schriftlichen Darstellungen gehört, um die interne Repräsentation der sprachlichen und grafischen Inhalte zu erstellen. Alle Teilprozesse bedingen sich gegenseitig und stehen in Relation zueinander (Hayes, 1996) (Abbildung 2).

### Abbildung 2

*Das kognitive Schreibmodell (eigene Abbildung nach Hayes, 1996)*



### 3.3 Das Kognitive Schreibmodell von Hayes (2012)

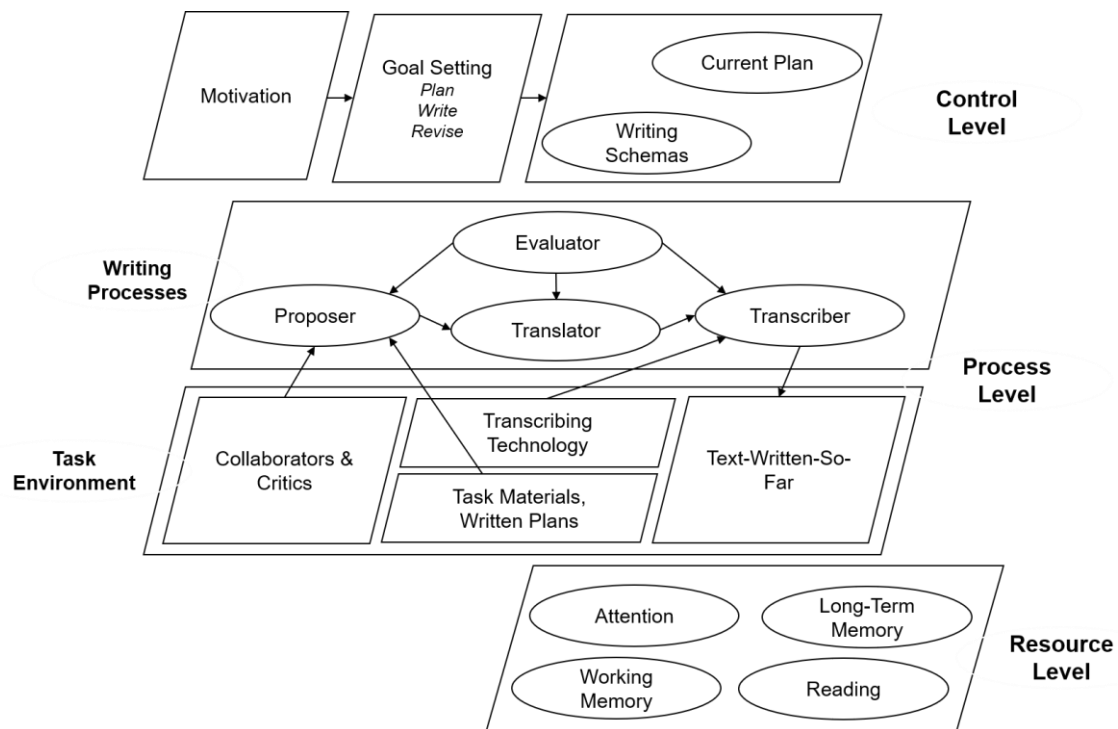
Mehr als ein Jahrzehnt später veröffentlichte Hayes eine erneute Überarbeitung seines kognitiven Schreibmodells, in dem die kognitiven Voraussetzungen des Schreibens noch detaillierter dargestellt wurden. Alle Teilfähigkeiten des vorherigen Modells wurden auf die drei Ebenen – *Kontrollebene*, *Prozessebene* und *Ressourcenebene* – verteilt und um weitere Faktoren ergänzt (Abbildung 3). Auf der *Kontrollebene* (control level) findet die Koordination und Kontrolle der kognitiven Prozesse (plan, write, revise: planen, schreiben, überarbeiten) statt. Darin ebenso verankert ist der Schreibplan mit dem Schreibziel. Zudem wird die Schreibmotivation aus dem vorherigen Modell auf dieser Ebene integriert. Die *Prozessebene* (process level) lässt sich in den *Schreibprozess* (writing process) und die *Schreibumgebung* (task environment) unterteilen. Innerhalb des *Schreibprozesses* erfolgt die simultane Abfolge sämtlicher in einer Person ablaufenden kognitiven Prozesse. Auf dieser Ebene interagieren Evaluator, Proposer, Translator und Transkriber miteinander. Der Proposer greift auf mögliche Inhalte im Langzeitgedächtnis zurück und generiert Ideen, welche der Translator in Schriftsprache übersetzt und vom Transkriber verschriftlich werden. Der Evaluator entscheidet als Kontrollinstanz, was dem Schreibziel dienlich ist und beeinflusst somit die drei anderen Prozesse während des gesamten Schreibprozesses stetig. Die *Schreibumgebung*, welche bereits aus dem vorherigen Modell bekannt ist, enthält alle Faktoren außerhalb der schreibenden Person, wie z.B. das Schreibthema oder die Adressaten. Sie interagiert ebenfalls mit dem *Schreibprozess* und beeinflusst den Proposer sowie den Transkriber. Die *Ressourcenebene* (resource level) beinhaltet die sprachlichen und kognitiven Voraussetzungen, die für das Schreiben relevant sind. Auf dieser Ebene sind Arbeits- und Langzeitgedächtnis, die bereits relevanter Bestandteil des Vorgängermodells waren, angesiedelt, aber auch die Lesefähigkeit sowie Aufmerksamkeitssteuerung gelten als wichtig für den Schreibprozess und sind auf dieser Ebene wiederzufinden. Im Modell werden die kognitiven Voraussetzungen des Schreibenden im Vergleich zu den beiden

Vorgängerversionen expliziter dargestellt. Dies erfolgt einerseits durch die Einteilung der einzelnen Teilfähigkeiten auf verschiedene Ebenen und andererseits durch das Verdeutlichen von Beziehungen zwischen den Teilfähigkeiten (Hayes, 2012) (Abbildung 3).

Die beschriebenen Schreibmodelle beziehen sich auf erwachsene, erfahrene Schreibende, die den Schriftspracherwerb bereits vollständig abgeschlossen haben. Es ist anzunehmen, dass sich der Schreibprozess bei sich in der Entwicklung befindenden Schreibenden von dem in diesem Modell dargestellten Prozess unterscheidet. Daher können die Modellannahmen von Schreibexpert\*innen nicht auf Schreibanfänger\*innen übertragen werden (Berninger, 1996).

### Abbildung 3

*Das Kognitive Schreibmodell (eigene Abbildung nach Hayes, 2012)*



### 3.4 Simple View of Writing (1980-2006)

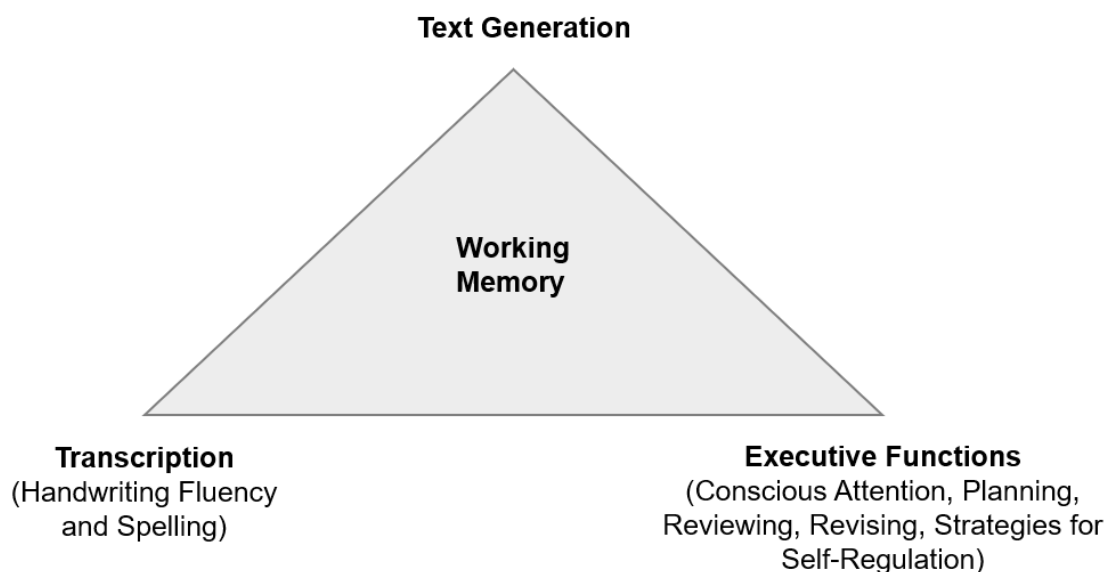
Um zu verstehen vor welchen Herausforderungen Schreibanfänger\*innen bei der Textproduktion stehen, deren Schriftspracherwerb noch nicht abgeschlossen ist, wurde 1986 von Juel et al. die *Simple View of Reading and Writing* entwickelt. Mit diesem Modell sollte die Entwicklung im Lesen und Schreiben bei Schulkindern dargestellt werden. Die *Simple View of Writing*, die im Folgenden fokussiert wird, bestand in dieser ersten Version aus den

Teilfähigkeiten *Rechtschreibung* und *Ideenfindung*. Eine longitudinale Studie mit Kindern der ersten bis vierten Klasse zeigte, dass Rechtschreibung in der ersten Klasse den größten Einfluss auf das Schreiben hatte. Bei älteren Schüler\*innen wurde die Ideenfindung als entscheidender Faktor identifiziert. Die Beziehung zwischen diesen beiden Fähigkeiten konnte erstmals verschiedene Entwicklungsstatus von Kindern erklären (Swanson & Berninger, 1996). 2002 überarbeiteten Berninger et al. mit Hilfe von Forschungsergebnissen aus den Bereichen Bildung, Kognition, Linguistik, kindliche Entwicklung und Neuropsychologie die *Simple View of Writing* Theorie. Die Ideenfindung wurde zur *Textgenerierung* (text generation), welche die Fähigkeit beschreibt, Ideen zu generieren und zu organisieren. Textgenerierung kann hierbei auf Wort-, Satz- und Diskursebene verortet werden (Berninger et al., 2002). Die Rechtschreibung wurde durch die Handschrift ergänzt und findet sich in der Komponente *Transkription* wieder, welche per Definition als die Fähigkeit, Buchstaben und Wörter orthografisch korrekt und flüssig zu schreiben, bezeichnet wird. Der *Simple View of Writing* hinzugefügt wurden außerdem noch die *Exekutivfunktion*, die die Planung, Überwachung und Überarbeitung umfassen. Laut der *Simple View of Writing* unterstützen die Transkriptionsfähigkeiten und die Exekutivfunktionen die Textgenerierung in der Umgebung des Arbeitsgedächtnisses, das das Zentrum des Modells darstellt und als übergeordnete Instanz das Kurz- und Langzeitgedächtnis koordiniert. Zu Beginn des Schreibenlernens scheinen die Transkriptionsfähigkeiten besonders großen Einfluss auf die Textgenerierung zu haben. Mit zunehmender Automatisierung der Transkriptionsfähigkeiten wird deren Einfluss jedoch geringer und kognitive Ressourcen werden frei, sodass die Textgenerierung flüssiger und organisierter verläuft (Berninger & Amtmann, 2003). Die fortlaufende technologische Weiterentwicklung im Bereich der funktionellen Magnetresonanztomografie und deren Anwendung bei Erwachsenen und Kindern führte zu neuen Erkenntnissen in der Neuropsychologie und Schreibforschung, sodass auch die *Simple View of Writing* angepasst werden musste. So veröffentlichten Berninger und Winn im Jahre

2006 die *Not-so-Simple View of Writing* (Abbildung 4). Die Veränderungen im Modell bezogen sich hauptsächlich auf die Rolle des *Arbeitsgedächtnisses*: Während des gesamten Schreibprozesses – von der Planung über die Komposition bis hin zur Überprüfung und Überarbeitung – wird das Langzeitgedächtnis aktiviert. Demgegenüber wird das Arbeitsgedächtnis lediglich während der Phasen der Überprüfung und Überarbeitung aktiviert. Weiterhin wird das *Arbeitsgedächtnis* im überarbeiteten Modell präziser dargestellt, indem es in mehrere Teilfähigkeiten aufgeschlüsselt wird: 1) orthografischer, phonologischer und morphologischer Speicher; 2) phonologische Schleife für das Lernen von Wörtern und die aktive Beibehaltung verbaler Informationen; 3) exekutive Hilfen, die das verbale Arbeitsgedächtnis mit dem Netzwerk für Exekutivfunktionen und mit dem nonverbalen Arbeitsgedächtnis, das Informationen im visuell-räumlichen Skizzenblock speichert, verbindet (Berninger & Winn, 2006). Die *Exekutivfunktionen* wurden um die überwachende Aufmerksamkeit ergänzt, die entscheidet, welche Informationen relevant sind und irrelevante Informationen unterdrückt. Dies führt dazu, dass der Schreibende bei seiner Schreibaufgabe bleiben kann, aber dennoch zwischen den verschiedenen mentalen Aufgaben beim Schreiben wechseln kann (Berninger & Winn, 2006) (Abbildung 4).

#### **Abbildung 4**

*(Not-so-)Simple View of Writing (eigene Abbildung nach Beninger & Winn, 2006)*



### 3.5 Direct and Indirect Effects Model of Writing (2017-2022)

Erst einige Jahre später wurde mit dem *Direct and Indirect Effects Model of Writing* (DIEW) ein weiterer Versuch unternommen, den Schreibprozess bei sich in der Entwicklung befindenden Schreibenden noch präziser durch direkte und indirekte Beziehungen zwischen den einzelnen Teilfähigkeiten darzustellen (Abbildung 5). Die Autor\*innen kritisierten beispielsweise an der *Not-so-simple View of Writing*, dass eine Spezifizierung der einzelnen Teilfähigkeiten fehle. Besonders die Textgenerierung sowie die zugrundeliegenden Beziehungen seien nicht konkretisiert worden (Kim & Schatschneider, 2017), weshalb es nicht möglich sei, eine konkrete Diagnostik oder Förderung aus diesen Modellen abzuleiten (Kim & Graham, 2022). Deshalb wurde von Kim und Schatschneider im Jahre 2017 ein neues Modell entwickelt und von Kim und Park (2019) auf seine Generalisierbarkeit überprüft. Das Modell beschreibt die konkreten Teilfähigkeiten, die in den Prozess und die Entwicklung des Schreibens involviert sind. Bisherige Modelle wurden deshalb zum einen um hierarchiehohe kognitive Fähigkeiten sowie zum anderen um Hintergrundwissen wie inhaltliches und diskursives Wissen sowie Affekt und Motivation erweitert. Außerdem wurden weitere strukturelle Beziehungen zwischen den einzelnen Teilfähigkeiten hinzugefügt. Hierbei gingen die Autor\*innen davon aus, dass viele sprachliche sowie kognitive Fähigkeiten direkt und indirekt zur Produktion von Texten beitragen und es deshalb wichtig sei, sowohl direkte als auch indirekte Pfade zu berücksichtigen. Des Weiteren wurde die Hypothese aufgestellt, dass die Nutzung hierarchieniedriger Teilfähigkeiten Voraussetzung für die hierarchiehohen Teilfähigkeiten sei, weshalb sowohl von einer direkten als auch einer indirekten Beteiligung der hierarchieniedrigen und hierarchiehohen Teilfähigkeiten beim Schreiben von Texten ausgegangen wurde. Ähnlich wie in bestehenden Modellen wurden das Arbeitsgedächtnis und die Exekutivfunktionen hier als grundlegende kognitive Fähigkeit für sprachliche Fähigkeiten gesehen, welche wiederum zum Schreiben beitragen (Abbildung 5). Das latente Konstrukt der *Textgenerierung* (Discourse Oral Language) wurde im DIEW durch mündliche

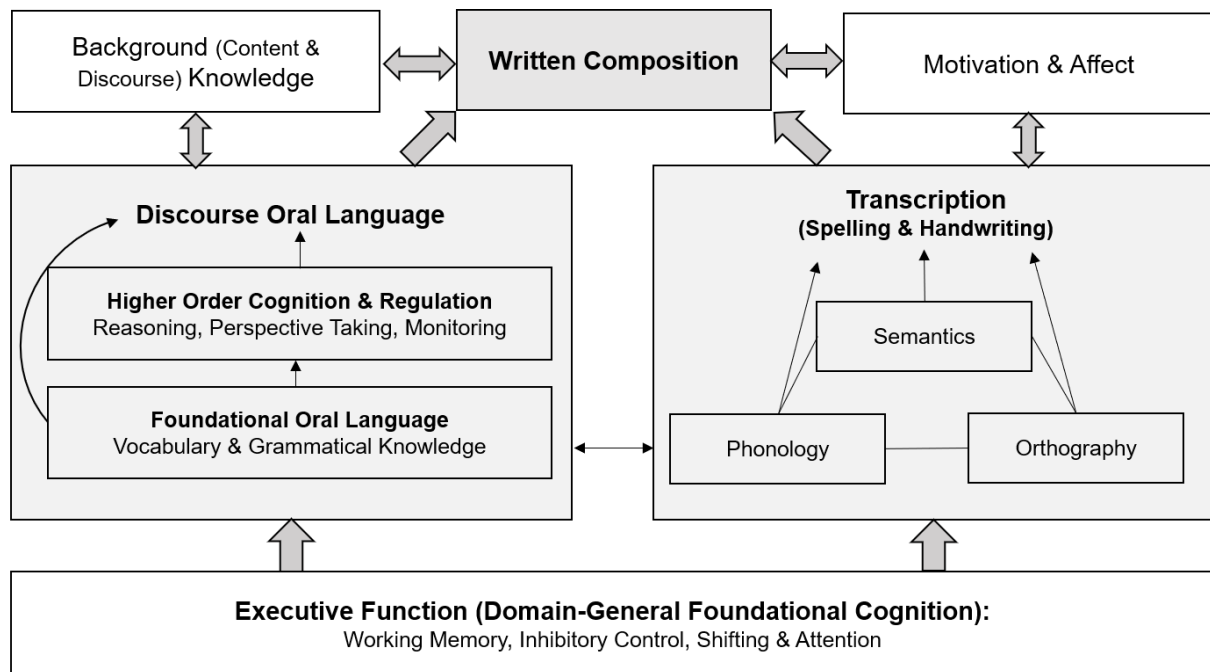
Sprachkenntnisse (Wortschatz und grammatikalische Kenntnisse) sowie durch hierarchiehohe kognitive Fähigkeiten (Inferenz, Perspektivübernahme und Monitoring) konkretisiert. Gemäß Kim et al. (2011) sollte die Textgenerierung mündlich erfasst werden, da sie der mündlichen Sprachrepräsentation, also der mentalen Darstellung sowie Verarbeitung von gesprochener Sprache, entspricht. Dies basiert auf der Annahme, dass präverbale Ideen und Gedanken in mündliche Sprache kodiert werden müssen, bevor sie in schriftliche Texte übertragen werden können. Gemeinsam mit *Rechtschreibung* und *Handschrift* (Transcription) vermittelt die *Textgenerierung* (Discourse Oral Language) die Beziehung zwischen *hierarchiehohen kognitiven Fähigkeiten*, *Arbeitsgedächtnis* und *Textqualität*. Das Modell wurde in seiner ersten Fassung mit englischsprachigen Erstklässler\*innen evaluiert, indem der Einfluss der einzelnen Teilfähigkeiten auf die Textqualität sowie deren Interaktion untereinander untersucht wurde (Kim & Schatschneider, 2017). Kim und Park (2019) bestätigten die Interaktionen der Teilfähigkeiten des *DIEWs* mit longitudinalen Daten von Kindern der 1. bis 3. Jahrgangsstufe. Basierend auf den Resultaten konnten drei Hypothesen hinsichtlich struktureller Beziehungen zwischen den Teilfähigkeiten aufgestellt werden: Die Ergebnisse der Studie legen zum einen nahe, dass sich die hierarchischen Beziehungen zwischen den einzelnen Teilfähigkeiten im Zuge der Schreibentwicklung wandeln und stützt somit die Annahmen der *Not-so-simple View of Writing* bezüglich der Automatisierung der Transkriptionsfähigkeiten und der dadurch freiwerdenden Ressourcen für hierarchiehohe Prozesse (Berninger & Winn, 2006; Kim & Park, 2019), was zur Postulierung der *Hypothese der dynamischen Beziehungen* führte. Die *Hypothese der hierarchischen Beziehungen* besagt, dass es sowohl direkte als auch indirekte Einflüsse von den hierarchieniedrigen Teilfähigkeiten auf die Textproduktion gibt, wobei die indirekten Einflüsse über die hierarchiehohen Teilfähigkeiten ausgeübt werden. Die *Hypothese der interaktiven Beziehungen* wiederum postuliert, dass die Beziehungen sowohl zwischen den einzelnen



Fähigkeiten und dem Schreiben als auch zwischen den einzelnen Fähigkeiten untereinander während der Entwicklung interaktiv und bidirektional sind (Kim & Park, 2019).

### Abbildung 5

*Direct and Indirect Effects Model of Writing (eigene Abbildung nach Kim & Park, 2019)*

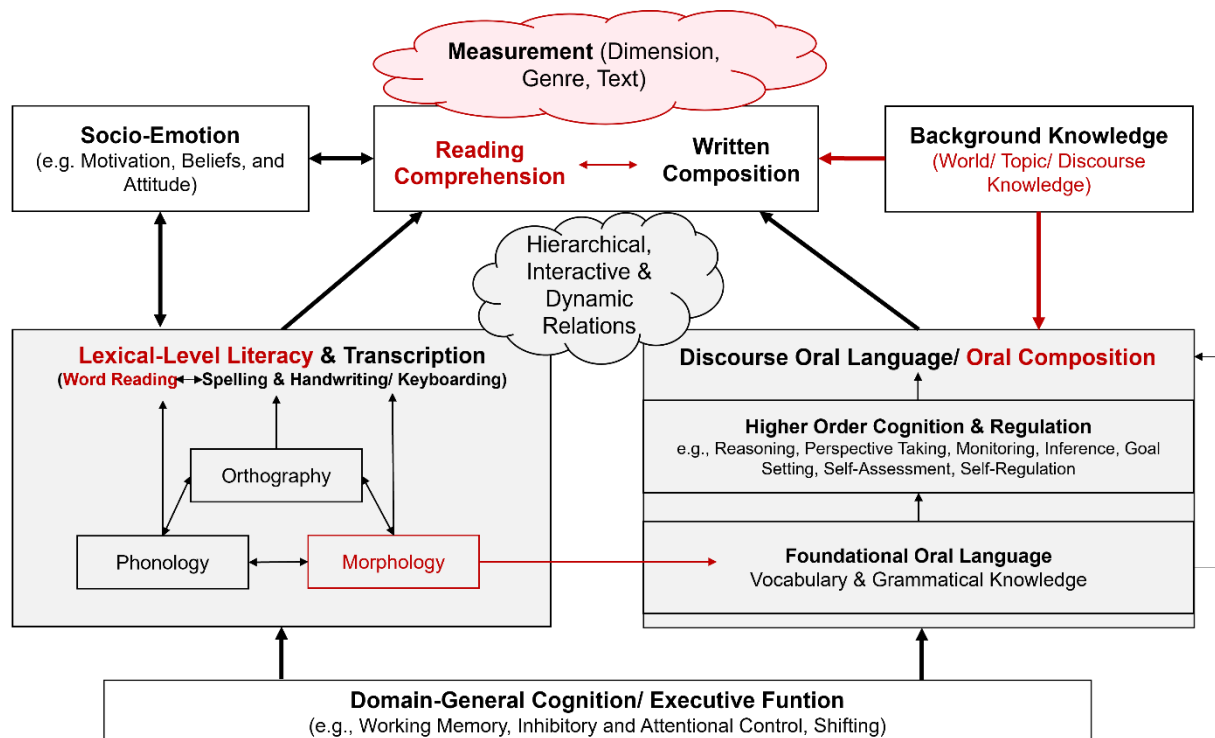


Im Jahre 2022 fand eine erneute Überarbeitung des Modells durch Kim und Graham statt, in der das *DIEW* um die *Hypothese der dynamischen oder differentiellen Beziehungen* in Abhängigkeit von der Messung und den Dimensionen des Schreibens erweitert wurde. Als Dimensionen wurden hier verschiedene Aspekte des Schreibproduktes betitelt (Textqualität, Korrektheit des Textes und Schreibproduktivität). Es wurde vermutet, dass die Beziehungen zwischen den einzelnen Teilfähigkeiten und deren Einfluss auf das Schreibprodukt je nach Dimension variieren. So könne beispielsweise in der Phase der Überarbeitung die Rechtschreibung einen größeren Einfluss auf die Korrektheit des Textes haben und der Wortschatz einen größeren Einfluss auf die Qualität des Textes. Zusätzlich wurde ebenfalls das *Leseverständnis* als eine Fähigkeit, die zum Schreiben beiträgt, in das Modell aufgenommen (Abbildung 6). Kim und Graham (2022) postulierten, dass das *DIEW* ein allgemein gültiges Modell ist, das für verschiedene Schreibaufgaben und Genres eingesetzt

werden kann. Laut den Autor\*innen bildet das Model die Textproduktionskompetenz adäquat ab, obwohl die einzelnen Teilfähigkeiten ggf. in Abhängigkeit von individuellen Faktoren wie der Entwicklung, der Schreibaufgabe und den Messfaktoren variieren können (Kim & Graham, 2022).

### Abbildung 6

*Expanded Direct and Indirect Effects Model of Writing (eigene Abbildung nach Kim & Graham, 2022)*



*Anmerkung.* Das *DIEW* wurde von Kim und Graham, 2022 erweitert. Die Erweiterungen des *DIEWs* wurden in der Abbildung rot markiert.

### 3.6 Modellannahmen über den Entwicklungsverlauf des Schreibens

Die bereits vorgestellten Schreibmodelle liefern relevante Informationen über die Interaktion von verschiedenen kognitiven, Transkriptions- und textspezifischen Fähigkeiten. Während mit der *Not-so-simple View of Writing* Theorie und dem *DIEW* die Textproduktionskompetenz von Schreibnovizen abgebildet werden können, zeigt Hayes (2012) in seinem aktuellsten Modell die Interaktion der einzelnen Prozesse des Schreibens bei erfahrenen Schreibenden, die die Schreibentwicklung bereits abgeschlossen haben. Diese

Modelle bieten jedoch keinen Überblick über den Entwicklungsverlauf der Textproduktionskompetenz. Sie liefern keine Informationen darüber, zu welchem Zeitpunkt die Kinder, welche Fähigkeiten erwerben sollten. Diese Informationen kann man sogenannten Schreibentwicklungsmodellen entnehmen, welche die normgerechte Entwicklung der Textproduktionskompetenz vom Schreibnovizen bis hin zum erfahrenen Schreibenden abbilden. Diese Modelle orientieren sich inhaltlich an den Schreibmodellen, geben aber eine bestimmte Reihenfolge vor, in denen die Teilfähigkeiten erworben werden und in der Schreibproduktion an Relevanz gewinnen.

Bereiter (1980) hat im Rahmen dessen ein fünfstufiges Modell (Tabelle 1) entworfen, das die Teilfähigkeiten der Textproduktionskompetenz nach ihrem Schwierigkeitsgrad ordnet. Diese fünf Stufen sind integrativ und werden von den Schreibenden innerhalb des Lernprozesses durchlaufen. Während das assoziative Schreiben (Stufe 1) noch die einfache Verknüpfung von Ideen beinhaltet, werden in der darauffolgenden 2. Stufe bereits hierarchieniedrige Fähigkeiten bei der Textproduktion angewandt. Die folgenden Stufen thematisieren die zusätzliche Anwendung hierarchiehoher textspezifischer und kognitiver Fähigkeiten und den Einsatz von Bearbeitungsstrategien. Die letzte Stufe beinhaltet das Schreiben als Prozess des Denkens, um Schreiben somit zur Erkenntniserweiterung einsetzen zu können. Bei dem Modell ist zu beachten, dass die einzelnen Stufen nicht als starre Reihenfolge betrachtet werden sollten. Vielmehr ist eine simultane Entwicklung einzelner Fähigkeiten möglich.

Inhaltlich ähnlich, aber etwas vereinfacht haben Bereiter und Scardamalia (1987) die Schreibentwicklung dargestellt. Sie unterscheiden in ihrem Modell zwischen zwei Arten des Schreibens: In einer ersten Phase liegt der Fokus eines jeden Schreibenden auf dem sogenannten *Knowledge Telling*. Dabei bringen die Schreibenden ihre Ideen ohne Planung und Überarbeitung ungeordnet zu Papier. Der Fokus liegt hier auf dem Reproduzieren von

vorhandenem Wissen. Dabei orientieren sie sich zunächst am Erlebten und geben dies unreflektiert wieder.

**Tabelle 1**

*Schreibentwicklungsmodell nach Bereiter (1980)*

Stufe	Bezeichnung der Stufe	Kompetenzen
1. Stufe	Das assoziative Schreiben <i>(associative writing)</i>	Ideenfindung und -verknüpfung ohne konzeptionelle Planung und Berücksichtigung der Kohärenz
2. Stufe	Das normorientierte Schreiben <i>(performative writing)</i>	Produktion eines normgerechten Textes unter Berücksichtigung von Orthografie, Syntax, Grammatik und Textmusterwissen
3. Stufe	Das leserbezogene Schreiben <i>(communicative writing)</i>	Berücksichtigung der Leserschaft durch den Einsatz von Adressatenorientierung
4. Stufe	Das kritische Schreiben <i>(unified writing)</i>	Kritische Beurteilung des eigenen Textes hinsichtlich Inhalt, Sprache und Wirkung auf den Lesenden
5. Stufe	Das erkenntnisbringende Schreiben <i>(epistemic writing)</i>	Schreiben als Bestandteil des Denkens, der durch Verarbeitungsprozesse zu einem Zuwachs an deklarativem Wissen führt

Die fortgeschrittene Form des Schreibens nennen die Autoren *Knowledge-Transforming*. Statt der reinen Wiedergabe von Wissen, findet auf dieser Ebene eine aktive Auseinandersetzung mit dem Text statt. Der Schreibende plant und überarbeitet das Schreibprodukt stetig unter der Berücksichtigung der Leserschaft. Dabei nutzt er/sie kohäsionsstiftende Mittel, um sowohl global als auch lokal Kohärenz herzustellen. Mit

zunehmender Schreiberfahrung wird immer mehr die Strategie des *Knowledge-Transforming* angewandt, wobei es gerade im Schreibentwicklungsprozess auch immer wieder zu einer Mischform aus beiden Strategien kommen kann.

Aus den beiden Modellen lassen sich Meilensteine für die Schreibentwicklung ableiten. Jedoch gibt es in den Modellen keinerlei Angaben zu Altersgrenzen oder Jahrgangsstufen.

Angelehnt an die Inhalte von Breiter (1980) sowie von Bereiter und Scardamalia (1987) bieten Becker-Mrotzek und Böttcher (2012) mit ihrem Modell, bestehend aus vier Phasen, zeitliche Orientierungswerte, sodass eine potentielle Einordnung der Schüler\*innen anhand ihrer erworbenen Kompetenzen erfolgen kann. Dies bietet somit einen entscheidenden Vorteil für die Diagnostik und die Förderplanung. Die vier Phasen erstrecken sich von der Einschulung bis hin zum Schreiben in der Adoleszenz. Zu Beginn des Schriftspracherwerbs im Alter zwischen fünf bis sieben wird das Arbeitsgedächtnis durch das Erlernen der basalen Transkriptionsfähigkeiten zunächst stark belastet, was es für Schreibanfänger\*innen schwierig macht, sich gleichzeitig auf die Textproduktion zu konzentrieren. Mit zunehmender Automatisierung im Alter zwischen sieben bis zehn Jahren, kommt es dann zur Produktion von Texten. Hierbei nutzen die Schüler\*innen vor allem Muster, die ihnen aus dem mündlichen Erzählen bekannt sind und orientieren sich beim Schreiben am Erlebten. Es kommt zu Verwendung der *Knowledge Telling* Strategie (Scardamalia & Bereiter, 1987). Mit zehn bis 14 Jahren findet eine immer differenzierte Schreibentwicklung statt. Es wird Textmusterwissen erworben sowie der Einsatz grammatikalischer und sprachlicher Mittel weiterentwickelt. Dabei wird das Geschriebene immer mehr an die Adressat\*innen angepasst, sodass es zur Verwendung der *Knowledge Transforming* Strategie kommt (Scardamalia & Bereiter, 1987). Die letzte Phase ist gekennzeichnet durch den vollständigen Erwerb der literalen Kompetenz. Hier stehen dem Schreibenden sämtliche Teilfähigkeiten des Schreibprozesses zur Verfügung, sodass er/sie in der Lage sein muss, verschiedenste Sachverhalte mittels angemessener Textart, unter Verwendung geeigneter sprachlicher Mittel,

für einen bestimmten Adressatenkreis formulieren zu können. Die in diesem Modell verwendeten Altersgrenzen stellen keine Norm dar. Sie dienen lediglich als Orientierung, da es gerade in den mittleren Phasen zu großen Varianzen und Mischformen kommen kann. Ein guter Schreibunterricht beeinflusst hier die Entwicklung maßgeblich (Becker-Mrotzek & Böttcher, 2012).

### **3.7 Möglichkeiten und Grenzen bisheriger Modelle**

Die Produktion von Texten gilt als kognitiv herausfordernder Prozess. Aus diesem Grund wurde mit der Etablierung der Schreibmodelle durch Hayes und Flower ein kognitiver Ansatz des Schreibens diskutiert, zunächst bei erfahrenen Schreibenden und anschließend mittels der *(Not-so-) Simple View of Writing* Theorie bei Kindern, deren Schreibentwicklung noch nicht abgeschlossen ist. Ein Kritikpunkt an diesen Modellen ist jedoch die unzureichende Spezifizierung der hierarchiehohe Teilfähigkeiten (z.B. Kim & Schatschneider, 2017; Philippek/ Kreutz et al., 2025). Besonders der Begriff der Textgenerierung wird in der Literatur vielfach diskutiert. Nach Berninger und Kolleg\*innen (1992) beruht die Textgenerierung definitionsgemäß sowohl auf der Generierung und Organisation von Ideen als auch auf deren Umsetzung in sprachliche Repräsentationen. Die Textgenerierung mache innerhalb des Schreibprozesses den Unterschied zwischen starken und schwachen Schreibenden aus (Berninger, 2000). Was das konkret bedeutet bzw. welche schriftsprachlichen Fähigkeiten gemeint sind, bleibt unklar. Nichtsdestotrotz ist die Textgenerierung ein wesentlicher Bestandteil vieler Studien. Die Ergebnisse sind in diesem Zusammenhang jedoch nur bedingt vergleichbar, da die Operationalisierung aufgrund der mangelnden Spezifizierung teils sehr unterschiedlich vorgenommen wird. In der Studie von Limpo und Alves (2013) wird die Textgenerierung als abhängige Variable verwendet und durch die Textqualität operationalisiert. In anderen Studien wird sie als Teilfähigkeit der Textqualität und somit als unabhängige Variable verstanden und ihr Einfluss auf die Textqualität untersucht: Salas und Silvente (2020) erfassten die Textgenerierung auf

Wortebene durch die Anzahl der Wörter pro Text und auf Satzebene durch die durchschnittliche Anzahl Wörter pro Satz. Die aktuellste Forschung zur Modellierung der Textproduktionskompetenz setzt die Textgenerierung mit mündlichen Sprachfähigkeiten (Wortschatz und grammatikalische Fähigkeiten) gleich, mit der Begründung, dass generierte präverbale Ideen in mündliche Sprache kodiert werden müssen bevor sie in eine schriftliche Repräsentation umgesetzt werden können (Berninger, 2000; Kim & Graham, 2022; Kim & Park, 2019; Kim & Schatschneider, 2017). So wurde beispielsweise in der empirischen Studie zum *DIEW* die Textgenerierung unter anderem mittels aktivem mündlichen Vokabeltest operationalisiert (Kim & Graham, 2022; Kim & Park, 2019; Kim & Schatschneider, 2017). Im Rahmen des Schriftspracherwerbs wird zwar vor allem der Übergang von der Generierung mündlicher Texte zur Produktion schriftlicher Texte fokussiert, doch müssen die Kinder im Kontext der Textgenerierung neue Prozesse erlernen, die sie dazu befähigen, die mündliche Sprache in geschriebene Sprache zu transferieren, weshalb die Textgenerierung nicht mit der mündlichen Sprache gleichgesetzt werden kann (Arfè & Pizzocaro, 2016). Daher scheint es nicht zielführend, zu hypothesieren, dass die Textgenerierung mit der mündlichen Sprachproduktion gleichzusetzen ist und es bleibt unklar, welche spezifischen Teilfähigkeiten Berninger und Kollegen mit der Erstellung des Modells beabsichtigten.

Aufgrund der unzureichenden Spezifizierung der hierarchiehohe schriftsprachlichen Teilfähigkeiten in den Modellen ist die Übertragung der bestehenden Schreibmodelle in die Schulpraxis nur schwer möglich, da sie weder explizite Hinweise für die Diagnostik noch für die Förderung im Bereich der Textproduktionskompetenz geben können. In Deutschland wurde bereits seit der PISA-Studie im Jahre 2000 und der damit verbundenen Entwicklung von Bildungsstandards eine stärkere Kompetenzorientierung gefordert (Klieme et al., 2003). Diese Kompetenzorientierung bezieht sich bislang allerdings nur auf kognitive Problemlösekompetenzen, die insbesondere in den Schreibprozessphasen des Planens,

Übersetzens und Überarbeitens verankert sind. Hierarchiehohe schriftsprachliche Teilfähigkeiten im Modell darzustellen, fällt bislang offensichtlich noch schwer.

Als weitere Forschungslücke lässt sich anführen, dass sich bisherige Modelle entweder an erwachsene und erfahrene Schreibende richten (Hayes, 2012; Hayes & Flower, 1980) oder auf Daten von Schreibanfänger\*innen der 1. – 3. Jahrgangsstufe beruhen (Kim & Graham, 2022; Kim & Park, 2019; Kim & Schatschneider, 2017). Da es sich beim Schreiben jedoch um einen hierarchischen, dynamischen und interaktiven Prozess handelt (Kim & Park, 2019), fehlen weitere Daten, die die Interaktion der Teilfähigkeiten und deren Einfluss auf die Textqualität bei älteren Schüler\*innen untersuchen, deren Automatisierung der Transkriptionsfähigkeiten schon weiter fortgeschritten ist.

Schreibentwicklungsmodelle können hier erste Hinweise auf die Teilfähigkeiten liefern, die in dieser Altersstufe die Textproduktion prägen (Becker-Mrotzek & Böttcher, 2012).

Insgesamt zeigen diese Erkenntnisse, dass die Modellierung der Textproduktionskompetenz mit den bestehenden Modellen noch lange nicht abgeschlossen ist und unterstreicht die Notwendigkeit weiterer Forschungsinitiativen in diesem Bereich. Bestehende Modelle liefern jedoch bereits wichtige Hinweise zu den relevanten Teilfähigkeiten und deren Interaktion untereinander sowie deren Einfluss auf die Textqualität, worauf neue Modelle aufbauen können.



#### 4 Teilfähigkeiten der Textproduktionskompetenz

Die bestehende Studienlage zur Textproduktionskompetenz verweist auf eine Vielzahl an Teilfähigkeiten, die einen bedeutsamen Einfluss auf den Schreibprozess und somit auch das Schreibprodukt haben. Aus den Schreibmodellen geht hervor, dass verschiedenste kognitive Aspekte relevant sind. Besonders das Arbeitsgedächtnis nimmt in allen beschriebenen Modellen einen zentralen Platz ein. Darüber hinaus ließen sich verschiedene schriftsprachliche Teilfähigkeiten identifiziert, die gemäß der *Hypothese der hierarchischen Beziehungen* (Kim & Park, 2019) auf verschiedenen Ebenen anzusiedeln sind: Die basalen Teilfähigkeiten, die bereits zu Beginn des Schriftspracherwerbs erlernt werden und im Verlauf automatisiert werden können, werden als hierarchieniedrige Teilfähigkeiten bezeichnet. In der einschlägigen Literatur sowie in existierenden Schreibmodellen werden damit in der Regel die Transkriptionsfähigkeiten assoziiert (Sturm et al., 2017).

Hierarchiehohe Teilfähigkeiten sind fortgeschrittenere Fähigkeiten, die über die basalen Fähigkeiten hinaus gehen. Sie zeichnen sich dadurch aus, dass sie im Gegensatz zu den hierarchieniedrigeren Teilfähigkeiten nicht automatisiert werden können und stets für die Reflexion zugänglich sind (Olive & Kellogg, 2002; Sturm, 2024). Erst durch diese Fähigkeiten können einzelne Sätze zu einem kohärenten und kommunikativen Gesamttext werden, der an eine bestimmte Leserschaft adressiert wird.

Im Folgenden werden die auf Basis der vorliegenden Literatur und in Modellannahmen ermittelten, für die Textproduktionskompetenz relevanten Teilfähigkeiten näher beschrieben und deren Bedeutung für die Textqualität erläutert. Bei der Interpretation der aktuellen Studienlage besteht jedoch die Problematik, dass die Teilfähigkeiten teilweise sehr unterschiedlich definiert und gemessen werden, was die Vergleichbarkeit der Studienergebnisse erschwert. Im Folgenden wird daher für jede einzelne Teilfähigkeit eine Erläuterung hinsichtlich der in dieser Arbeit verstandenen Definition und Messung vorgenommen.

#### 4.1 Hierarchieniedrigere Transkriptionsfähigkeiten

In der *Not-so-simple View of Writing* Theorie stellen die Transkriptionsfähigkeiten eine der drei zentralen Komponenten dar. Sie gehören zu den basalsten und somit hierarchieniedrigsten Fähigkeiten, die von Beginn der Schulzeit an gelehrt werden. Die Transkriptionsfähigkeiten lassen sich unterteilen in Rechtschreibung und Schreibflüssigkeit, die sich wiederum in die Bereiche der Handschrift und der Verwendung einer Tastatur unterteilen lässt. Die Transkriptionsfähigkeiten umfassen somit das Abrufen, Zusammensetzen und Auswählen orthografischer Symbole sowie die Programmierung und Ausführung motorischer Bewegungen zur Herstellung dieser orthografischen Symbole (Abbott & Berninger, 1993; Connelly et al., 2012; Limpo & Alves, 2018).

Rechtschreibung kann sowohl im Freitext, durch das Auszählen von Rechtschreibfehlern im Text, als auch mittels standardisierter Tests, in denen den Kindern meist einzelne Wörter oder Sätze diktiert werden, ermittelt werden. Der Freitext hat im Vergleich zu einem standardisierten Test den Vorteil, eine realistischere Schreibsituation zu gestalten. Denn letztendlich soll eine korrekte Orthografie auf Textebene in verschiedenen Schreibsituationen angewendet werden können (Reichardt, 2015). Magalhaes et al. (2020) stellten jedoch fest, dass in den Klassen 2, 4 und 6 in der Diktatsituation mehr Rechtschreibfehler auftraten als beim Verfassen eines Aufsatzes. Dies begründeten die Autorinnen dadurch, dass die Kinder während der Diktatsituation vordefinierte Wörter schreiben mussten, wohingegen sie beim Verfassen eines Aufsatzes die Möglichkeit hatten, die Wörter selbst auszuwählen. Dadurch konnten sie Wörter, deren Schreibung sie nicht beherrschten, vermeiden (Magalhães et al., 2020). Andere Studien konnten jedoch zeigen, dass Kinder im Freitext deutlich mehr Fehler machen, da sie die Rechtschreibung parallel zu anderen Prozessen der Textproduktion bewältigen müssen (Fay, 2010; Risel, 2011). Dies verdeutlicht, dass die Erfassung der Rechtschreibung im Text zu einem anderen Ergebnis führt als mittels eines standardisierten Tests (Diktat oder Lückentext). Basierend darauf scheint es, als könnte ein Test ein

umfassenderes Bild der Rechtschreibleistung eines Kindes kreieren, da die Kinder vorgegebene Wörter schreiben müssen, die die deutsche Rechtschreibung weitestgehend abdecken. Ein weiterer Vorteil der Erfassung der Rechtschreibung mittels Tests ist die Erfüllung der Gütekriterien und die damit einhergehende Vergleichbarkeit der Ergebnisse verschiedener Kinder, weshalb in den beiden dargestellten Studien der vorliegenden Arbeit ein standardisierter Rechtschreibtest verwendet wurde.

Bei der *Schreibflüssigkeit* stellt sich nicht nur das Problem der Operationalisierung, sondern auch der Definition. Besonders im deutschsprachigen Raum existieren verschiedene Definitionen des Begriffs „Schreibflüssigkeit“. Sturm (2024) unterscheidet in ihrer Übersichtsarbeit fünf verschiedene Konzepte der Schreibflüssigkeit: In der englischsprachigen Literatur wird sie vornehmlich als die Ausführung motorischer Bewegungen mittels bestimmten Schreibwerkzeugs verstanden (handschriftliche Geläufigkeit). Andere Autor\*innen zählen zur Schreibflüssigkeit neben den motorischen Ausführungen auch die Verwendung korrekter Orthografie. Letztere Definition wiederum wird teilweise durch die Fähigkeit zum flüssigen Formulieren ergänzt. Teilweise wird jedoch auch lediglich die Formulierung weniger zusammenhängender Sätze oder die Formulierung ganzer Texte im Sinne der textgenerierenden Flüssigkeit als Schreibflüssigkeit betrachtet. Basierend auf den zuvor vorgestellten modelltheoretischen Annahmen, die die Schreibflüssigkeit als motorische Komponente beschreiben, wird sich im Folgenden auf die erstgenannte Definition bezogen.

Bedingt durch die verschiedenen Definitionen, gibt es auch bei der Messung diverse Möglichkeiten. Ähnlich wie bei der Rechtschreibung kann Schreibflüssigkeit im Text oder mittels Testung ermittelt werden. Wird die Schreibflüssigkeit im Text gemessen, so wird die Anzahl Wörter pro Minute erfasst (Chenoweth & Hayes, 2001; Graham et al., 1997). Bei dieser Variante der Messung ist zu berücksichtigen, dass es sich nicht um eine rein motorische Aufgabe handelt, sondern dass zusätzlich Anforderungen an zahlreiche weitere

Teilfähigkeiten gestellt werden. Am weitesten verbreitet ist daher die Messung mittels sogenannter Alphabetaufgabe, bei der die Buchstaben des Alphabets so schnell und so oft wie möglich korrekt hintereinander innerhalb einer Minute aufgeschrieben werden sollen (z.B. Alves & Limpo, 2015; Connelly et al., 2012; Limpo & Alves, 2013). Bei dieser Messvariante wird kritisiert, dass es sich nicht lediglich um eine motorische Aufgabe handle, sondern dass zusätzliche Anforderungen an das orthografische Wissen gestellt werden, die den Abruf von Phonem-Graphem-Korrespondenzen erfordern. Darüber hinaus würde beim Einhalten der Reihenfolge der Buchstaben der Abruf aus dem Langzeitgedächtnis abverlangt (Nottbusch, 2017). Weitere Anforderungen sind bei dieser Aufgabe jedoch so gering wie möglich gehalten. Zudem stellt sie eine besonders gute ökonomische und valide Messvariante dar (Schmitt & Knopp, 2017), weshalb diese Messmethode in den beiden dargestellten Studien dieser Arbeit verwendet wurde. Um die Schreibflüssigkeit auch auf Wort- und Satzebene zu messen, ergänzen einige Studien die Alphabetaufgabe um weitere Kopieraufgaben von Wörtern und Sätzen (Graham & Harris, 1997; Limpo & Alves, 2013; Limpo et al., 2017).

Im Hinblick auf den Zusammenhang zwischen Transkriptionsfähigkeiten und der Textproduktionskompetenz erweisen sich die Transkriptionsfähigkeiten in der frühen Schulzeit als Schlüsselkompetenz. Sie fungieren als Prädiktor für die Textproduktionskompetenz und bilden eine wesentliche Voraussetzung für diese (Connelly et al., 2012; Stephany, 2019). Sie leisten vor allem in der 1. bis 3. Klasse einen wichtigen Beitrag zur Entwicklung des Schreibens (Harrison et al., 2016), wobei die Rechtschreibung sogar von der 1. bis zur 7. Klasse als stabiler Prädiktor für die Textproduktionskompetenz gilt. Und auch die Schreibflüssigkeit macht einen großen Teil der Varianz in der Textqualität bei Schüler\*innen der 3. Klasse aus (Harrison et al., 2016). Nach der *Not-so-simple View of Writing* Theorie ist die Automatisierung der Transkriptionsfähigkeiten notwendig, um kognitive Ressourcen für die hierarchiehohe Teilfähigkeiten freizusetzen (Berninger & Winn, 2006).

Bei nicht altersgemäßer bzw. fehlender Automatisierung könnte der Schreibprozess gehemmt werden, was auf mehrere Faktoren zurückführen ist: Zum einen beansprucht eine nicht automatisierte Schreibflüssigkeit die begrenzte Kapazität des Arbeitsgedächtnisses. Zum anderen müssen Ideen bei verlangsamtem Schreiben länger im Gedächtnis gehalten werden, wodurch sie eher verloren gehen. Zudem erhöht sich der kognitive und psychische Aufwand, sodass Schreiben als belastend erlebt und zunehmend vermieden wird – ein Teufelskreis, der zu weniger Übung und damit zu ausbleibender Weiterentwicklung führen kann (Limpo et al., 2018). Zugleich erfordert das Nachdenken über eine korrekte Rechtschreibung ebenso erhöhte Arbeitsgedächtniskapazität, die sich auf die Wortwahl, Formulierung und Ideengenerierung negativ auswirken kann (Breetvelt et al., 1994). Insbesondere gilt für den Schulanfang, dass ohne die Kenntnis über Buchstaben, Rechtschreibregeln und Schreibflüssigkeit, es gar nicht erst möglich ist, Wörter, Sätze und Gedanken zu notieren. Eine Automatisierung ist erreicht, wenn die Schüler\*innen flüssig und weitestgehend ohne Rechtschreibfehler schreiben (Graham & Santangelo, 2014; Limpo & Alves, 2013).

Je nach Grad der Automatisierung konnten verschiedenste Studien signifikante direkte und indirekte Effekte der Transkriptionsfähigkeiten auf die Textqualität von der 1. bis zu 9. Klasse belegen (z.B. Abbott et al., 2010; Cordeiro et al., 2020; Graham & Harris, 1997; Limpo & Alves, 2013). Bei älteren Schüler\*innen schienen die Transkriptionsfähigkeiten mit zunehmender Schullaufbahn und somit wahrscheinlich zunehmender Automatisierung, die Qualität des Schreibens stets weniger zu beeinflussen (Abbott et al., 2010; Berninger et al., 2010). Die freiwerdenden kognitiven Kapazitäten können dann für die hierarchiehohe Fähigkeiten genutzt werden (Berninger & Winn, 2006; Philippek/ Kreutz et al., 2025).

#### **4.2 Hierarchiehohe textspezifische Fähigkeiten**

Innerhalb der *Not-so-simple View of Writing* Theorie werden die hierarchiehohen Fähigkeiten durch die Komponente Textgenerierung repräsentiert. Wie in Kapitel 3.7 bereits beschrieben, ist die Textgenerierung jedoch unterspezifiziert und es ist nicht ausreichend

geklärt, welche konkreten Fähigkeiten dieser zugrunde liegen. Damit zukünftige Modelle jedoch besser auf die Diagnostik und Förderung übertragbar sind, sollten sie konkrete hierarchiehohe textspezifische Fähigkeiten einbeziehen, die im Folgenden auch als solche betitelt werden.

In der Schreibforschung gibt es bereits zahlreiche Studien, die den Einfluss verschiedenster hierarchiehoher textspezifischer Fähigkeiten auf die Textqualität verschiedener Textarten untersucht haben. In der Literatur gelten insbesondere zwei Fähigkeiten als zentral für die Produktion narrativer Texte: das textspezifische aktive Vokabelwissen sowie die Fähigkeit zur Herstellung von Kohärenz.

#### ***4.2.1 Textspezifisches aktives Vokabelwissen***

Vor allem im englischsprachigen Raum wurden bereits zahlreiche Fähigkeiten untersucht, die den schriftlichen Wortschatz betreffenden, wie beispielsweise die Wortfrequenz (frequency), Bedeutsamkeit der Wörter (meaningfulness), lexikalische Raffinesse (lexical sophistication) oder lexikalische Dichte (lexical density) (Crossley et al., 2014; Gómez Vera et al., 2016). Im Rahmen des textspezifischen aktiven Vokabelwissens hat sich die lexikalische Vielfalt als eine der wichtigsten Determinanten der Textqualität erwiesen, insbesondere bei narrativen Texten (Gómez Vera et al., 2016; Olinghouse & Wilson, 2013). Es konnte festgestellt werden, dass sich ein vielfältiger und kreativer Wortschatz positiv auf die Qualität narrativer Texte auswirkt, was bei Sachtexten hingegen nicht beobachtet werden konnte (Olinghouse & Wilson, 2013). Nicht nur in Kindertexten, sondern auch bei erfahrenen Schreibenden sorgt eine diversere Wortwahl für besser bewertete Texte (D. McNamara et al., 2010). Denn eine ausreichend differenzierte Wortwahl ist nicht nur für die zwischenmenschliche Kommunikation unverzichtbar, sondern auch für die Fähigkeit Gedanken in Sprache zu fassen – mündlich und schriftlich (Amrein, 2017). Der Einfluss der lexikalischen Vielfalt auf die Textqualität kann von der 3. Klasse bis hin zu

Studierenden empirisch belegt werden (Cameron et al., 1995; Gómez Vera et al., 2016; D. McNamara et al., 2010; Olinghouse & Wilson, 2013).

In Abgrenzung zu den hierarchieniedrigen Transkriptionsfähigkeiten erweist es sich für die hierarchiehohen textspezifischen Fähigkeiten als vorteilhaft, die Erhebung und Ermittlung der jeweiligen Fähigkeit innerhalb des Schreibkontextes vorzunehmen. Dieser Vorgang wird durch eine entsprechende Messung innerhalb des Textes realisiert. So wird auch die lexikalische Vielfalt im Schreibprodukt bewertet. Hierbei haben sich mehrere Möglichkeiten der Berechnung etabliert: Die bekannteste Art der Berechnung ist die Type/Token Ratio (TTR), bei der die Anzahl Types (Anzahl verschiedener Wörter) durch die Anzahl Token (Anzahl Wörter) geteilt wird. Die Berechnung wird allerdings stark von der Textlänge beeinflusst, da die Wahrscheinlichkeit der Wiederholung der Wörter steigt und sich somit das Verhältnis von Types und Token zueinander verändert. Somit gilt die TTR nicht als valides Maß für die lexikalische Vielfalt (Koizumi & In'nami, 2012). Eine Alternative wurde versucht mit dem correctedTTR<sup>6</sup> und dem Guiraud<sup>7</sup> zu finden. Die Unabhängigkeit von der Textlänge ist auch bei diesen Maßen allerdings nicht eindeutig gesichert (Hess et al., 1989; Koizumi & In'nami, 2012). Als weitere Alternative wurde das Measure of Textual Lexical Diversity (MTLD) entwickelt, wobei es sich um ein softwarebasiertes Bewertungsmaß handelt. Der Einfluss der Textlänge wird hier reduziert, indem die mittlere Länge von aufeinanderfolgenden Wortketten, die einen evidenzbasierten TTR-Wert von 0.7 beinhalten, in einem Text berechnet wird. Im Rahmen der Analyse wird jeder Text sowohl vorwärts als auch rückwärts untersucht. Der finale MTLD-Score stellt dabei den Mittelwert aus dem vorwärts und rückwärts berechneten MTLD dar. Studien zu diesem Thema haben ergeben, dass das MTLD die lexikalische Vielfalt am präzisesten ermittelt, ohne dass die Textlänge einen signifikanten Einfluss auf die Ergebnisse hat (Koizumi & In'nami, 2012; McCarthy &

---

<sup>6</sup> Quotient aus der Anzahl an Types und der Wurzel der zweifachen Anzahl Token

<sup>7</sup> Quotient aus der Anzahl Types und der Wurzel aus der Anzahl Token

Jarvis, 2010). Da das MTLD derzeit als bestes Maß für die lexikalische Vielfalt gilt, wurde dies in den beiden dargestellten Studien dieser Arbeit zur Berechnung der lexikalischen Vielfalt verwendet.

Eine weitere Teilfähigkeit, die bislang vor allem im deutschsprachigen Raum Gegenstand von Untersuchungen zum textspezifischen Vokabelwissen war, ist die Verwendung angemessener Wörter innerhalb eines Textes. Die Auswahl der korrekten Wörter ist bei Schreibanfänger\*innen nicht selbstverständlich und kann das Verständnis des Lesenden und damit auch die Textqualität enorm beeinflussen (Becker-Mrotzek et al., 2014). Die Auswahl angemessener Wörter kann von entscheidender Bedeutung für den Inhalt einer Äußerung sein (Mathiebe, 2019). Erste Ergebnisse zu diesem Maß zeigen, dass Schüler\*innen höherer Klassen und Schulformen signifikant häufiger in der Lage waren, kontextangemessene Wörter zu verwenden (Mathiebe, 2019). Die Verwendung der angemessenen Wörter wurde bislang per Ratingverfahren durch zwei Bewertende bestimmt, deren Beurteilungen gemittelt wurden. Die Interraterreliabilität war dabei ausreichend ( $\alpha = .72$ ) (Mathiebe, 2019). Da das Maß eine gewisse Aussagekraft im Hinblick auf die Textqualität zu besitzen scheint, sollte es zukünftig weiter erforscht werden und kam daher in den beiden Studien dieser Arbeit zur Anwendung.

#### **4.2.2 Kohärenz**

Die hierarchiehohe Fähigkeit der Kohärenzbildung ist ein vielfach untersuchtes Gebiet innerhalb der Schreib- und Sprachforschung (Becker-Mrotzek et al., 2015; Becker-Mrotzek et al., 2014; Stephany, 2019). Kohärenz gilt als textliches Oberflächenphänomen und bezeichnet dabei die Fähigkeit, einen Sachverhalt unter logischen, chronologischen oder funktionalen Aspekten zu strukturieren und zu versprachlichen. Ein Text ist kohärent, wenn der Lesende den Zusammenhang zwischen verschiedenen Textbestandteilen rekonstruieren kann (Storrer, 2020). Der Zusammenhang kann dabei auf verschiedene Art und Weise vom Schreibenden auf der Textoberfläche hergestellt werden: Zum einen durch grammatische Aspekte wie



Syntaxstruktur oder Tempus (grammatische Kohärenz) oder auch durch den inhaltlich-thematischen Zusammenhang eines Textes (Averintseva-Klisch, 2018; Becker-Mrotzek et al., 2015). Damit der Gesamtzusammenhang – sei es grammatikalisch oder inhaltlich-thematisch – für den Lesenden deutlich wird, müssen alle Äußerungen innerhalb eines Textes entweder implizit oder explizit miteinander verbunden werden (Becker-Mrotzek et al., 2015). Im Rahmen der Herstellung von Kohärenz wird zwischen der globalen Kohärenz und der lokalen Kohäsion unterschieden (Hennes et al., 2018). Globale Kohärenz betrifft den Zusammenhang des Gesamttextes und kann beispielsweise explizit hergestellt werden durch Verbindungen zwischen einzelnen Absätzen, grafischen Absatzmarkierungen, Zwischenüberschriften oder grafischen Markierungen von Schlüsselbegriffen (Crossley & McNamara, 2010; Storrer, 2004). Globale Kohärenz kann auch implizit durch den Kontext und Weltwissen des Lesenden hergestellt werden. Kohäsion ist die auf der Textoberfläche messbare Eigenschaft der Kohärenz (Storrer, 2004) und kann zwischen benachbarten Textteilen oder aufeinanderfolgenden Sätzen hergestellt werden. Die Herstellung erfolgt immer explizit durch Satz- und Teilsatzverbindungen durch beispielsweise Konjunktionen, Konjunkionaladverbien oder durch Rekurrenz sowie Referenzen in der direkten Rede (Crossley & McNamara, 2010; Storrer, 2004). Des Weiteren sollten auf die Stimmigkeit der Wortwahl sowie auf eine sinnvolle Beziehung zwischen den Wörtern geachtet werden (Hennes et al., 2018). Auch Tempus und Modus, die Hinweise auf eine Sequenzierung von Ereignissen geben können, dienen als Werkzeuge, um den Zusammenhang innerhalb eines Textes herzustellen (Averintseva-Klisch, 2018; Zifonun, 2000). Die Studienlage zeigt einen deutlichen Einfluss der Kohäsion auf die Textqualität von Schüler\*innen der 3. Klasse bis hin zu Studierenden (Cameron et al., 1995; Cox et al., 1990; Crossley & McNamara, 2010; MacArthur et al., 2019).

In wissenschaftlichen Studien werden unterschiedliche Methoden zur Messung der Kohärenz angewendet. Dabei wird aber vor allem die Kohäsion als messbarer Teil der

Kohärenz auf Textoberfläche berücksichtigt. Ähnlich wie bei der Rechtschreibung besteht die Möglichkeit, die Fähigkeit zur Bildung von Kohäsion mittels Lückentext, in den die passenden sprachlichen Mittel wie Konnektoren oder Rekurrenten eingetragen werden, zu beurteilen (Becker-Mrotzek et al., 2014). Diese Variante der Messung ist allerdings sehr selten, da hierbei die konkrete Anwendung der Kohäsion unter Einbezug des Kontextes außer Acht gelassen wird. Häufiger findet die Bewertung der Kohäsion im Text mittels Ratingverfahren Anwendung, bei dem die adäquat verwendeten Kohäsionsmittel (bzw. die fehlenden Kohäsionsmittel) ermittelt werden. Um den Einfluss der Textlänge zu schmälern, kann die Anzahl an Kohäsionsmittel durch die Anzahl an Propositionen innerhalb des Textes geteilt werden (Cox et al., 1990). Dieses Verfahren wurde auch in den beiden dargestellten Studien dieser Arbeit angewendet.

### **4.3 Kognitive Funktionen**

Neben den Transkriptions- und den textspezifischen Fähigkeiten, die Bestandteil der Textproduktionskompetenz sind, gibt es Fähigkeiten, die die Voraussetzung für die Textproduktionskompetenz bilden (Grabowski et al., 2014). Dazu zählen die kognitiven Funktionen.

Seit Beginn der Entwicklung der Schreibmodelle ist sich die Wissenschaft einig, dass die Produktion von Texten ein kognitiver Prozess ist. Trotz zahlreicher Unterschiede zwischen den Modellen, bilden die kognitiven Funktionen eine Konstante, weshalb sie Teil vieler Studien im Bereich der Textproduktion sind. Von Besonderem Interesse sind der Einfluss des Arbeitsgedächtnisses sowie verschiedener hierarchieniedriger und hierarchiehoher Exekutivfunktionen.

#### **4.3.1 Exekutivfunktionen**

Bei Exekutivfunktionen handelt es sich um einen Regenschirmbegriff, der mehrere Subprozesse umfasst, die dazu dienen eine Handlung zielgerichtet auszuführen. Exekutivfunktionen können auf verschiedenste Art und Weise unterteilt werden. Eine der

bekanntesten Einteilungen stammt von Miyake et al. (2000), die Inhibition, Updating und Shifting als drei grundlegende Exekutivfunktionen definieren. Inhibition bezeichnet die Unterdrückung von impulsiven oder irrelevanten Reaktionen, um eine angemessenere oder eine gezieltere Handlung durchzuführen (Diamond, 2013). In Bezug auf den Schreibprozess wurde bereits herausgefunden, dass Inhibition die Schreibflüssigkeit unterstützen kann, indem alternative Buchstaben oder Buchstabenformen unterdrückt werden oder indem konkurrierende motorische Programme gehemmt werden (Salas & Silvente, 2020). Auch im Bereich der Rechtschreibung werden durch die Inhibition alternative Rechtschreibmuster unterdrückt, besonders in Fällen von Mehrdeutigkeit (Salas & Silvente, 2020). Ebenso kann Inhibition auf hierarchiehohe Prozesse wirken und irrelevante Ideen innerhalb des Schreibprozesses unterdrücken (Rocha et al., 2022; Ruffini et al., 2024). Updating ist die Fähigkeit, Informationen innerhalb des Arbeitsgedächtnisses regelmäßig zu kontrollieren und durch neue Informationen zu ersetzen (Diamond, 2013). Sie sorgt bei Schreibanfänger\*innen, bei denen die Transkriptionsfähigkeiten noch nicht automatisiert sind, dafür, die phonologische Form im Kurzzeitgedächtnis aktiv zu halten, bis der Schreibende die passende orthografische Regel ausgewählt hat (Salas & Silvente, 2020) oder das Wort handschriftlich vollendet hat (Altemeier et al., 2008). Bei automatisierten Transkriptionsfähigkeiten sorgt die Fähigkeit zum Updating dafür, Ideen während der Textproduktion aufrechtzuerhalten und mit anfänglichen Plänen sowie dem Schreibziel stetig abzugleichen und gegebenenfalls anzupassen (Ruffini et al., 2024). Die Forschung hat außerdem gezeigt, dass sowohl Inhibition als auch Updating, einen Einfluss auf die Textlänge im Schreibprozess haben, indem unangemessene Wörter unterdrückt und relevante Wörter während des Schreibprozesses aktiv ausgewählt werden (Drijbooms et al., 2015). Shifting bezeichnet die Fähigkeit, flexibel zwischen verschiedenen Aufgaben oder mentalen Anforderungen zu wechseln (Diamond, 2013). Bei Schreibanfänger\*innen ist Shifting auf hierarchieniedriger Ebene für den schnellen Wechsel zwischen Buchstaben verantwortlich, mit dem Ziel eine flüssigere Handschrift zu

erlangen. Auf hierarchiehoher Ebene ist es verantwortlich für den Wechsel zwischen verschiedenen Ideen innerhalb des Textes (Hooper et al., 2021) oder für den Wechsel zwischen den Phasen des Planens, Übersetzens und Überarbeitens, was charakteristisch für eine gute Textqualität ist (Hayes & Flower, 1980; Olive, 2021).

Studien, die den Einfluss der Exekutivfunktionen auf die Textqualität untersuchten, konnten sowohl direkte signifikante Effekte bei Kindern der 2.-8. Klasse (Cordeiro et al., 2020; Salas & Silvente, 2020) sowie indirekte signifikante Effekte über die Schreibflüssigkeit bei Kindern der 4. Klasse feststellen (Drijbooms et al., 2015).

Neben den drei bereits angeführten basalen Exekutivfunktionen existieren auch hierarchiehohe Exekutivfunktionen, zu denen das Schlussfolgern (reasoning), das Planen (planning) und das Überwachen (monitoring) zählen (Diamond, 2013). Diese stellen ebenfalls essentielle Fähigkeiten dar, die beim Schreiben von Texten benötigt werden, was die entscheidende Bedeutung von Exekutivfunktionen für den Schreibprozess unterstreicht (Olive, 2021). Trotz unbestrittener Relevanz der hierarchiehohen Exekutivfunktionen für den Schreibprozess konzentriert sich die vorliegende Arbeit auf die basalen Exekutivfunktionen, da diese in der Literatur klarer definiert sind und deren Einfluss auf die Textqualität bei Kindern gut belegt ist (z.B. Drijbooms et al., 2015; Rocha et al., 2022; Salas & Silvente, 2020). Ausschließlich deren Interaktion mit mehreren Fähigkeiten verschiedener hierarchischer Ebenen ist noch nicht ausreichend geklärt. Letztlich sind die basalen Exekutivfunktionen in die hierarchiehohen Exekutivfunktionen involviert, sodass auch darauf Rückschlüsse gezogen werden können (Berninger & Richards, 2010).

Die Messung der Exekutivfunktion ist durch standardisierte neuropsychologische Testungen oder Testbatterien möglich. Hierbei ist zu berücksichtigen, dass eine Vielzahl an Tests existiert, die zum Teil zwischen den diversen Unterfunktionen der Exekutivfunktionen differenzieren und zum Teil mehrere Unterfunktionen gemeinsam testen. In der Studie

Philippek/ Kreutz et al. (2025) wurde eine neuropsychologische Gruppentestung verwendet, welche die drei basalen Exekutivfunktionen (nach Miyake et al., 2000) adressiert.

## 5 Philippek/ Kreutz et al. (2025): Forschungsfragen und Untersuchungsaufbau

Wie in Kapitel 2 beschrieben, ist die Textproduktionskompetenz von großer Relevanz für den schulischen und beruflichen Erfolg und stellt eine Schlüsselkompetenz sowohl im Alltag als auch im sozialen Miteinander dar. Die Vermittlung dieser Kompetenz an die Schülerschaft ist eine grundlegende Aufgabe der Schulen, die frühestmöglich zu realisieren ist (Graham et al., 2023). Die Förderung dieser Kompetenz stellt bislang eine Herausforderung dar, da es sich um ein komplexes Konstrukt handelt, das aus mehreren Teilfähigkeiten besteht, die sowohl auf hierarchieniedriger als auch auf hierarchiehoher Ebene anzusiedeln sind. Mehrere Schreibmodelle haben bereits den Einfluss der verschiedensten Teilfähigkeiten auf die Textproduktionskompetenz – operationalisiert durch die Textqualität – untersucht. Dabei wurde festgestellt, dass sowohl kognitive Komponenten wie die Exekutivfunktionen als auch die hierarchieniedrigen Transkriptionsfähigkeiten einen signifikanten Einfluss auf die Textqualität haben (z.B. Berninger & Winn, 2006; Kim & Graham, 2022). Der Einfluss hierarchiehoher textspezifischer Fähigkeiten auf die Textqualität konnte abseits von Modellen ebenfalls bestätigt werden (z.B. Gómez Vera et al., 2016; MacArthur et al., 2019; D. McNamara et al., 2010). Allerdings wurden letztere Teilfähigkeiten noch nicht gemeinsam mit kognitiven Funktionen und hierarchieniedrigen Transkriptionsfähigkeiten modelliert. Des Weiteren bezogen sich bestehende Modelle bislang auf Schreibanfänger\*innen oder -expert\*innen. Gemäß der *Hypothese der dynamischen Beziehungen* (Kim & Park, 2019) verändert sich jedoch die Interaktion zwischen den Teilfähigkeiten und deren Einfluss auf die Textproduktionskompetenz, weshalb es besonders interessant erscheint, die Phase, in der die Transkriptionsfähigkeiten weitestgehend automatisiert sind, die Schreibentwicklung jedoch nicht vollständig abgeschlossen ist, näher zu betrachten. Mit der Erstellung und Verifizierung des *Cascaded Model of Writing (CASMOW)* wurde versucht, diese Forschungslücken zu schließen. *CASMOW* modelliert Exekutivfunktionen, hierarchieniedrige Transkriptionsfähigkeiten sowie hierarchiehohe textspezifische Fähigkeiten, die gemäß der

*Hypothese der hierarchischen Beziehungen* nach Kim und Park (2019) in drei Kaskaden angeordnet wurden. Die Aufteilung der jeweiligen Teilfähigkeiten auf die Kaskaden basierte auf der Annahme, dass die Exekutivfunktionen als zentrale Basis für alle Teilfähigkeiten fungieren. Sie nehmen somit sowohl einen direkten Einfluss auf die Transkriptionsfähigkeiten als auch auf die textspezifischen Fähigkeiten und koordinieren diese. Sie haben jedoch keinen direkten Einfluss auf die Textproduktionskompetenz. Weiterhin wurde angenommen, dass die hierarchieniedrigen Fähigkeiten durch eine weitestgehende Automatisierung einen indirekten Einfluss über die hierarchiehohen Teilfähigkeiten auf die Textproduktionskompetenz haben und die hierarchiehohen Teilfähigkeiten einen direkten Einfluss haben.

In der folgenden Studie wurde die Überprüfung von *CASMOW* hinsichtlich der Textproduktionskompetenz bei narrativen Texten von Schüler\*innen in der Sekundarstufe I anhand der folgenden Forschungsfragen vorgenommen:

1. *Inwieweit kann die Textqualität in Erzählungen durch das Cascaded Model of Writing vorhergesagt werden?*
2. *Wie interagieren die Teilfähigkeiten des Cascaded Model of Writing miteinander?*

Um diese Fragen beantworten zu können, wurden 186 Schüler\*innen (89 Mädchen; durchschnittliches Alter: 11,75 Jahre) der 5. bis 7. Jahrgangsstufe deutscher Gymnasien und Gesamtschulen untersucht. Die Exekutivfunktionen und die hierarchieniedrigen Teilfähigkeiten aus dem *CASMOW* wurden mit Hilfe von standardisierten Tests ermittelt. Um die Textproduktionskompetenz zu bestimmen, wurden die Schüler\*innen gebeten, eine Geschichte zu einem bestimmten Thema zu produzieren. Die Textqualität der Geschichten wurde mittels Comparative Judgement Methode geratet. Dieselben Schüler\*innentexte wurden zur Bestimmung der hierarchiehohen textspezifischen Teilfähigkeiten genutzt, da es als sinnvoll erachtet wurde, den Abruf dieser Fähigkeiten auf der Textoberfläche innerhalb des Schreibkontextes zu ermitteln (Cameron et al., 1995). Zur Beantwortung der Forschungsfragen wurden sämtliche im Rahmen von *CASMOW* erhobenen Teilfähigkeiten in

ein Strukturgleichungsmodell überführt. Hierzu wurden alle Kaskaden sowie Pfade von *CASMOW* im Strukturgleichungsmodell nachgebildet und auf Signifikanz überprüft. Nach Ermittlung des Modelfits wurde die Varianz der Teilfähigkeiten innerhalb der Textqualität analysiert. Obwohl die Textlänge per se keinen Einfluss auf die Textqualität hat und somit auch nicht in *CASMOW* enthalten ist, wurde sie dennoch in einem zweiten Schritt in das Modell integriert. Dies erfolgte, da sie in der Vergangenheit häufig als Einflussfaktor auf die Textqualität genannt wurde. So konnte beobachtet werden, dass längere Texte meist besser beurteilt werden (Fleckenstein et al., 2020; MacArthur et al., 2019).

Auf Basis der Literatur wurde angenommen, dass durch die gemeinsame Modellierung von kognitiven als auch Transkriptions- und textspezifischen Teilfähigkeiten ein Großteil der Varianz in der Textqualität erklärt werden kann.

Die hierarchiehohen textspezifischen Teilfähigkeiten wurden in der folgenden Forschungsarbeit in hierarchiehohe kompositionsspezifische Fähigkeiten (higher-level composing-specific skills) umbenannt, um zu betonen, dass es sich um Fertigkeiten handelt, die für die Produktion eines funktionalen Textes erforderlich sind und somit dem englischen Begriff der Textproduktionskompetenz (composing competence) zugeordnet werden. Sie entsprechen jedoch denselben Fähigkeiten.

Die Studie Philippek/ Kreutz et al. (2025) wird im folgenden Kapitel dargestellt.



## **6 The Contributions of Executive Functions, Transcription Skills and Text-specific Skills to Text Quality in Narratives**

**Source:** Philippek, J.<sup>8</sup>, Kreutz, R.M.<sup>8</sup>, Hennes, A.-K. & Schabmann, A. (2025). The contributions of executive functions, transcription skills and text-specific skills to text quality in narratives. *Reading and Writing*, 38, 651-670. <https://doi.org/10.1007/s11145-024-10528-5>

**Disclosures and Acknowledgements:** We have no conflict of interest to disclose. We received no financial support for the research. We would like to thank all the children and students who participated in the study.

**Publication Link:** <https://link.springer.com/article/10.1007/s11145-024-10528-5>

**Licence:** <https://creativecommons.org/licenses/by/4.0/>

**Changes:** The references cited in this research paper have been adapted and included in the references list at the end of the dissertation. In addition, the appendix of this paper can be found in the appendix of the dissertation. The formatting has also been adapted.

---

<sup>8</sup> Julie Philippek and Rebecca M. Kreutz have contributed equally to this work and share first authorship.

### **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

## 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, 2014). 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, 2014). 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; D. 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, 2012). 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. D. 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, 2000). 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 D. McNamara (2010), that cohesive elements (as rated by experts) in persuasive essays correlated positively with text quality. In addition, Crossley and D. 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 (Arfè & 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.

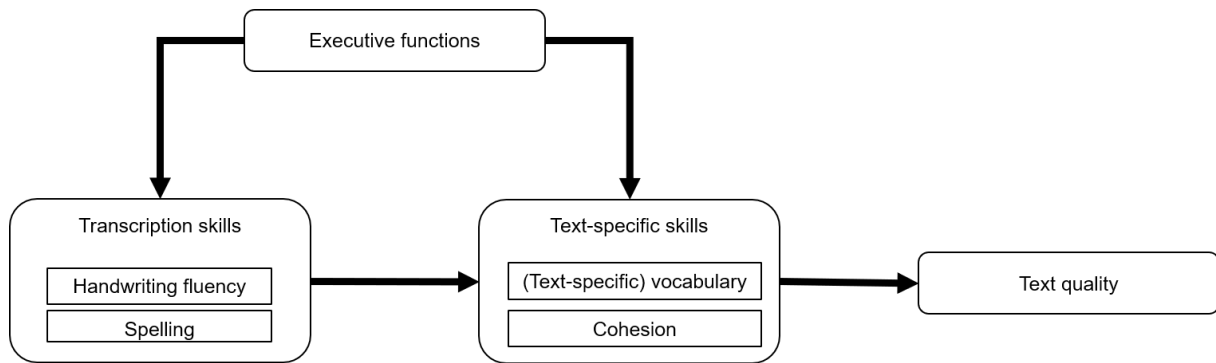
In arranging these subskills, we referred on the hierarchical relations hypothesis from the expanded Direct and Indirect Effects Model of Writing (DIEW) by Kim and Graham (2022), which states that not all subskills are directly related to text quality and that low-level



transcription skills are related to text quality via higher-level subskills. Specifically, DIEW assumes that transcription skills have an influence on written composition among others via discourse oral language.

**Figure 1**

*The Cascaded Model of Writing*



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 school<sup>9</sup>: 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

---

<sup>9</sup> Comprehensive secondary school encompassing all possible tracks.

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 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 D. 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, 2012b). 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., 2024; Verhavert et al., 2019).

In this study, comparative judgements were conducted using the online comparing tool Comproved ([www.comproved.com](http://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 Chi<sup>2</sup>-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 ( $\text{Chi}^2(10) = 12.62$ ,  $p = .246$ ,  $\text{RMSEA} = 0.038$ ,  $\text{CFI} = 0.989$ ,  $\text{NFI} = 0.952$ ).

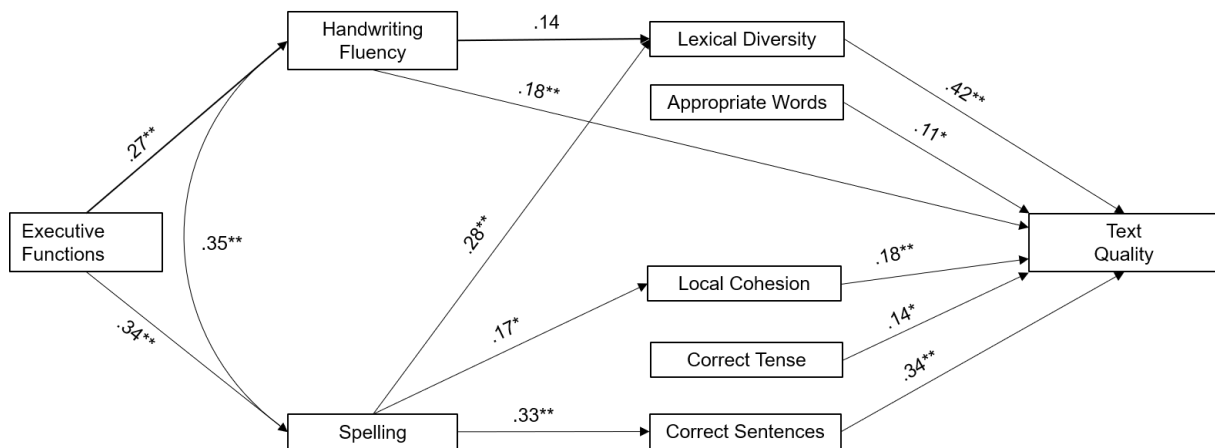
**Table 1***Descriptive Statistics*

Measures	Mean	Standard deviation	Range
Executive functions	2.64	1.48	0 – 6 (6) <sup>a</sup>
Handwriting fluency	63.25	23.00	8 – 138
Spelling	36.94	8.24	13 – 49 (49) <sup>a</sup>
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

<sup>a</sup>Theoretical maximum**Table 2***Correlations of Variables*

	1	2	3	4	5	6	7	8	9
1. Executive functions									
2. Handwriting fluency	.27**								
3. Spelling	.34**	.41**							
4. Lexical diversity	.08	.25**	.34**						
5. Appropriate words	.15*	.01	.07	.18*					
6. Local cohesion	.09	.00	.17*	.19**	.06				
7. Correct tense	.01	-.03	-.05	-.07	-.01	.02			
8. Correct sentences	.18*	.04	.33**	.05	-.01	.03	-.06		
9. Text length	.29**	.47**	.44**	.48**	.15*	.22**	-.01	.21**	
10. Text quality	.20**	.30**	.44**	.53**	.20**	.28**	.09	.36**	.60**

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

**Figure 2***Structural equation model of the Cascaded Model of Writing (1B)*

*Note:* 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:  $\Delta \chi^2(13) = 14.559$ ,  $p = .336$ ).

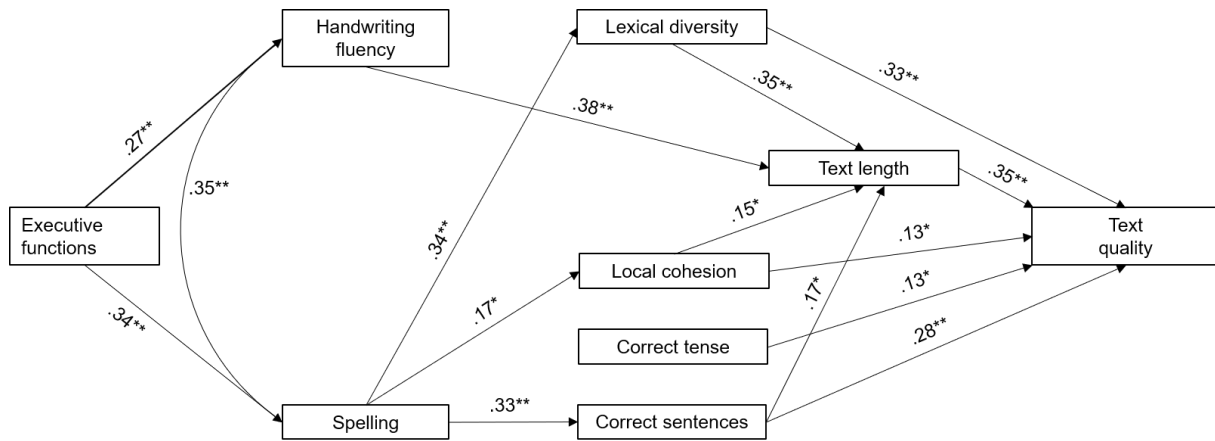
In this model, significant direct effects could be found from all text-specific skills and from handwriting fluency to text quality. Moreover, there was an indirect effect from handwriting fluency to text quality via lexical diversity ( $\beta = 0.06$ ;  $p = .048$ ); thus, handwriting fluency had a 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 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 ( $\chi^2(21) = 23.58, p = .314, RMSEA = 0.026, CFI = 0.992, NFI = 0.937$ ).

**Figure 3**

*Structural equation model of the Cascaded Model of Writing (2C)*



*Note:* Standardized path coefficients of the interactions between executive functions, transcription skills, text-specific skills, text length and text quality. \* $p < .05$ ; \*\* $p < .01$

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; D. 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 is 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 (D. 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.

### 7 Philippek/ Kreutz et al. (2025): Zwischenfazit

In der Studie Philippek/ Kreutz et al. (2025) wurde ein Modell konzipiert, das die Interaktion von Teilfähigkeiten verschiedener hierarchischer Ebenen und deren Einfluss auf die Textproduktionskompetenz darstellt. Das theoriegeleitete *CASMOW* konnte empirisch für Schüler\*innen der 5. – 7. Jahrgangsstufe bestätigt werden:

Es zeigte sich, dass Exekutivfunktionen, hierarchieniedrige Transkriptionsfähigkeiten und hierarchiehohe textspezifische Fähigkeiten (textspezifisches Vokabelwissen und Fähigkeiten zur Herstellung von Kohäsion) die Qualität narrativer Texte beeinflussen (Forschungsfrage 1). Zudem wurde deutlich, dass die einzelnen Teilfähigkeiten miteinander interagieren (Forschungsfrage 2). Die Ergebnisse stehen im Einklang mit bestehenden Schreibmodellen. So konnte – wie in der *Not-so-Simple-View of Writing* beschrieben (Berniger & Winn, 2006) – ein Einfluss kognitiver Funktionen, Transkriptionsfähigkeiten und hierarchiehoher Fähigkeiten auf die Textqualität bestätigt werden. Darüber hinaus konnte die im *DIEW* (Kim & Park, 2019; Kim & Graham, 2022) postulierte *Hypothese der hierarchischen Beziehungen* empirisch belegt werden: Die Transkriptionsfähigkeiten beeinflussten die Textqualität nicht unmittelbar, sondern indirekt über die hierarchiehohen textspezifischen Fähigkeiten bzw. über die Textlänge. Ebenfalls übereinstimmend mit bisherigen Modellannahmen fungierten die Exekutivfunktionen als Basis aller Teilfähigkeiten (Berninger & Winn, 2006; Kim & Graham, 2022). Ihr Einfluss auf die Textqualität wurde dabei durch die Transkriptionsfähigkeiten sowie die hierarchiehohen textspezifischen Fähigkeiten moderiert. *CASMOW* kann somit grundsätzlich als notwendige Ergänzung vorheriger Modelle um die Komponenten der hierarchiehohen textspezifischen Teilfähigkeiten gesehen werden und ermöglicht ein differenziertes Verständnis der Interaktion zugrundeliegender Teilfähigkeiten narrativer Texte.

Aufgrund der Untersuchung expliziter Teilfähigkeiten durch *CASMOW* können nun Implikationen für die Praxis abgeleitet werden: Die hierarchiehohen textspezifischen

Fähigkeiten sollten bereits vor, aber spätestens ab der Sekundarstufe I in den Fokus des Schreibunterrichts rücken und explizit trainiert werden. Auf Basis bisheriger Modellannahmen (z.B. *Not-so-simple View of Writing*) könnte angenommen werden, dass dies lediglich erfolgsversprechend ist, sobald die Automatisierung der hierarchieniedrigen Transkriptionsfähigkeiten weitestgehend abgeschlossen ist, sodass kognitive Kapazität für die hierarchiehohen Fähigkeiten frei wird. Demnach könnten beispielsweise Schüler\*innen mit einer Lese-Rechtschreibstörung erst im fortgeschrittenen Alter nach andauernder Förderung in der Lage sein, funktionale Texte zu produzieren. Aus der Praxis geht hervor, dass eine Lese-Rechtschreibstörung die Textproduktion zwar beeinflussen kann, die Produktion funktionaler Texte aber dennoch möglich ist. Eine eingeschränkte Fähigkeit zur Textproduktion wird daher nicht als typisches Symptom oder diagnostisches Kriterium einer Lese-Rechtschreibstörung angesehen (Reichardt, 2015; Schulte-Körne, 2022). Auch im schulischen Alltag zeigt sich, dass es Kinder gibt, die trotz schwacher Transkriptionsfähigkeiten in der Lage sind, gute Texte zu produzieren (Reichardt, 2015). Demnach wird hypothetisiert, dass die Textqualität andere Mechanismen impliziert als die basalen Transkriptionsfähigkeiten und somit doppelte Dissoziationen zwischen den beiden Konstrukten vorliegen könnten.

## 8 Doppelte Dissoziationen

Bei Dissoziationen handelt es sich um ein bekanntes Konzept aus der Neuropsychologie und der Kognitionsforschung, das erstmals in den 1960er Jahren beschrieben wurde (Dunn & Kirsner, 2003). Anhand von Dissoziationen lässt sich untersuchen, ob eine bestimmte kognitive Funktion an der Ausführung einer Aufgabe beteiligt ist, an der Ausführung einer anderen Aufgabe jedoch nicht. Durch Dissoziationen ist es also möglich, die neuronal strukturelle und funktionelle Trennung verschiedener kognitiver Funktionen, die vermeintlich interagieren, näher zu untersuchen. Ziel ist es jedoch nicht, die konkrete Lokalisation innerhalb des Gehirns zu bestimmen (Teuber, 1955; Young et al., 2000). Dazu bedarf es weiterführender Untersuchungen mittels bildgebender Verfahren.

In der Neuropsychologie unterscheidet man zwischen der einfachen und doppelten Dissoziation: Bei der einfachen Dissoziation ist durch eine Läsion oder experimentelle Manipulation eines bestimmten Hirnareals, die zu einem Defizit einer bestimmten kognitiven Funktion führt, eine kognitive Aufgabe signifikant stärker beeinträchtigt als eine andere Aufgabe. Ein klassisches Beispiel für eine einfache Dissoziation ist die Broca Aphasie, bei der durch eine Läsion im sogenannten Broca Areal (benannt nach seinem Entdecker Paul Broca) eine Sprachstörung vorliegt, die sich in einer gestörten Sprachproduktion aber einem intakten Sprachverständnis äußert. Einfache Dissoziationen stellen meist nur einen schwachen Hinweis für das Vorhandensein unabhängiger kognitiver Funktionen dar, denen zwei bestimmte Aufgaben zugrunde liegen. Die Begründung dafür ist, dass es möglich sein könnte, dass die unbeeinträchtigte Funktion weniger empfindlich auf die experimentelle Manipulation oder die Hirnschädigung reagiert (Dunn & Kirsner, 2003).

Doppelte Dissoziationen hingegen sind ein anerkanntes Konzept, dem in den 2000er Jahren besondere Aufmerksamkeit gewidmet wurde (Dunn & Kirsner, 2003). Der Begriff wurde erstmals von Teuber (1955) eingeführt. Doppelte Dissoziationen liegen vor, wenn zwei verschiedene Aufgaben unterschiedliche Verarbeitungsanforderungen an zwei oder mehr

funktionell dissoziierbare Subsysteme stellen (Shallice, 1979). Sie sind besonders im Bereich der neurologischen Sprachstörungen von großer Relevanz und liefern beispielsweise den Beweis, dass bei einer Broca Aphasie und einer Wernicke Aphasie unterschiedliche Funktionen in unterschiedlichen Hirnarealen beeinträchtigt sind. Während bei einer Broca Aphasie – wie bei der einfachen Dissoziation beschrieben – die Sprachproduktion beeinträchtigt und das Sprachverständnis intakt ist, ist bei einer Wernicke Aphasie die Sprachproduktion flüssig und das Sprachverständnis beeinträchtigt. Dies zeigt, dass Sprachproduktion und Sprachverständnis verschiedene Subsysteme zu Grunde liegen und die Ursache für eine Broca und eine Wernicke Aphasie in Läsionen in unterschiedlichen Hirnarealen liegen müssen. Diese Annahmen wurden sowohl durch post-mortem Untersuchungen sowie in aktuelleren Studien mittels bildgebender Verfahren bestätigt. Im Bereich der Schriftsprache konnten Landerl und Moll (2009) bestätigen, dass doppelte Dissoziationen zwischen Lese- und Rechtschreibdefiziten durch unterschiedliche Beeinträchtigungen, die im Zusammenhang mit Lese- und Schreibfähigkeiten stehen, erklärt werden können. So profitieren Kinder mit isolierten Lesedefiziten nicht von ihrem intakten wortspezifischen Wissen. Ihre langsame Lesegeschwindigkeit kann durch eine Beeinträchtigungen des visuell-verbalen Zugangs erklärt werden, während bei Kindern mit isolierter Rechtschreibschwäche hingegen das Defizit beim Aufbau von wortspezifischen Wissen durch einen intakten visuell-verbalen Zugang kompensiert werden kann (Moll & Landerl, 2009). Doppelte Dissoziationen innerhalb der Textproduktionskompetenz wurden bislang noch nicht untersucht.

## 9 Kreutz et al. (2025): Forschungsfragen und Untersuchungsaufbau

Die Bedeutung der basalen Transkriptionsfähigkeiten für die Textproduktion ist sowohl in der Forschung als auch in der schulischen Praxis bislang nicht abschließend geklärt. Während Schreibentwicklungsmodelle (Kapitel 3.6) darauf hinweisen, dass sich bereits ab der Primarstufe die hierarchiehohe Teilfähigkeiten wie die Fähigkeit zur Kohärenzherstellung und Adressatenorientierung entwickeln und somit auch gefördert werden müssen, scheint es in der Praxis so, als würde weiterhin der Fokus auf die Transkriptionsfähigkeiten gelegt werden (Cordeiro et al., 2020; Graham & Harris, 1997; Moats et al., 2006).

Empirische Studien haben gezeigt, dass bei ausbleibender Automatisierung der basalen Fähigkeiten und dadurch stark beanspruchter kognitiver Ressourcen die Produktion guter Texte gehemmt werden könnte. Gleichzeitig zeigen Studien, die die Entwicklung der Transkriptionsfähigkeiten von Kindern der Primarstufe mit Kindern der Sekundarstufe I vergleichen, dass bei den älteren Kindern geringere Zusammenhänge zwischen Transkriptionsfähigkeiten und Textqualität bestehen (Limpo & Alves, 2013; Nottbusch, 2017). Dies bestätigt, dass mit zunehmendem Alter und somit zunehmender Automatisierung der Transkriptionsfähigkeiten der Einfluss auf die Textqualität abnimmt. Jedoch gibt es Belege dafür, dass die Transkriptionsfähigkeiten auch noch in der Mittelstufe das Arbeitsgedächtnis belasten und eine Verbesserung der basalen Fähigkeiten noch bis zur 9. Klasse möglich ist (Alves & Limpo, 2015; Limpo et al., 2018).

Obwohl empirische Studien einen Zusammenhang zwischen Transkriptionsfähigkeiten und Textproduktionskompetenz nahelegen, zeigt die schulische Praxis, dass Kinder mit schwachen Transkriptionsfähigkeiten durchaus funktionale Texte verfassen können. Ebenso gibt es Kinder, die trotz altersadäquater Transkriptionsfähigkeiten nicht in der Lage sind, gute Texte zu produzieren (Reichardt, 2015). Dies stützt die Annahme der doppelten Dissoziationen, bei der zwei Fähigkeiten zwar korrelieren, jedoch auch unabhängig voneinander auftreten und auf unterschiedliche kognitive Prozesse zurückzuführen sind.

Folglich wird eine differenzierte Betrachtung des Zusammenhangs von Transkriptionsfähigkeiten und Textproduktionskompetenz benötigt, weshalb sich für die Studie Kreutz et al. (2025) folgende erste Forschungsfrage ergibt:

1. *Gibt es bei Schüler\*innen der 5. bis 7. Klasse klar definierbare doppelte Dissoziationen in der Leistung der hierarchieniedrigen Transkriptionsfähigkeiten und der Textqualität (altersadäquate Textqualität/ defizitäre Transkriptionsfähigkeiten oder defizitäre Textqualität/ altersadäquate Transkriptionsfähigkeiten)?*

Es wird davon ausgegangen, dass der komplexe Prozess der Textproduktion andere zugrundeliegende Fähigkeiten impliziert als die hierarchieniedrigen Transkriptionsfähigkeiten und sich beide Konstrukte unabhängig voneinander entwickeln können. Dementsprechend wird vermutet, dass es Kinder mit verschiedenen Leistungsprofilen gibt. Die resultierenden Leistungsprofile könnten durch doppelte Dissoziationen erklärt werden. Sollte dies bestätigt werden können, stellt sich die Frage, welche Faktoren bei Schüler\*innen mit defizitären Transkriptionsfähigkeiten dazu beitragen, dass sie dennoch gute Texte produzieren können. *CASMOW* hat gezeigt, dass die Textqualität vornehmlich durch die hierarchiehohe textspezifischen Teilfähigkeiten (textspezifisches Vokabelwissen und die Fähigkeit zur Herstellung von Kohäsion) beeinflusst wird, während die Transkriptionsfähigkeiten nur einen indirekten Einfluss haben (siehe Philippek/ Kreutz et al., 2025). Demnach lautet die zweite Forschungsfrage:

2. *Falls doppelte Dissoziationen vorliegen: Unterscheiden sich die Leistungen der jeweiligen Schüler\*innen hinsichtlich der textspezifischen*

*Formulierungsfähigkeiten<sup>10</sup>, so dass diese Fähigkeiten die Unterschiede in der Leistung der Textproduktion erklären können?*

Wie in der Studie Philippek/ Kreutz et al. (2025) dargelegt, üben die hierarchiehohe Teilfähigkeiten einen signifikanten Einfluss auf die Textqualität aus. Somit kann davon ausgegangen werden, dass sie einen entscheidenden Beitrag zur Textqualität leisten. Demnach besteht die Hypothese, dass diese Fähigkeiten bei den Schüler\*innen, die defizitäre Leistungen in den Transkriptionsfähigkeiten aufweisen, aber gute Texte produzieren, für die bessere Textqualität verantwortlich sind und trotz fehlender Automatisierung zur Anwendung kommen. Dies könnte Implikationen für den Schreibunterricht liefern, in dem die Transkriptionsfähigkeiten bislang noch stark im Fokus stehen (Cordeiro et al., 2020; Moats et al., 2006).

Für die Studie Kreutz et al. (2025) wurde die gleiche Stichprobe wie in der Studie Philippek/ Kreutz et al. (2025) herangezogen. Die 186 Schüler\*innen (durchschnittliches Alter = 11,75 Jahre; 48% Mädchen) wurden gebeten, einen Erzähltext zu schreiben sowie Tests zur Schreibflüssigkeit und zur Rechtschreibung durchzuführen. Die Schüler\*innentexte wurden im Anschluss dazu verwendet, die Textqualität mittels der Comparative Judgement Methode zu bestimmen, sowie das Measure of Textual Lexical Diversity (MTLD) zu berechnen und die Maße für die angepasste Wortwahl und für die Kohäsion (lokale Kohäsion, korrekte Syntax, korrektes Tempus) zu ermitteln. Für die Berechnungen von Studie 2 wurden die Variablen der Textqualität, Rechtschreibung und Schreibflüssigkeit weiterverarbeitet. Dazu wurde jeweils eine Rangfolge bestimmt. Im Anschluss wurden die Schüler\*innenleistungen in schwach (< 20%) und altersadäquat (> 25%) umkodiert. Zur Erstellung der Profile mit doppelten Dissoziationen wurden die einzelnen Maße

---

<sup>10</sup> Innerhalb des Begutachtungsprozesses des Journals of Writing Research wurden die hierarchiehohe textspezifischen Fähigkeiten in textspezifische Formulierungsfähigkeiten (text-specific formulation skills) umbenannt. Sie entsprechen jedoch denselben Teilkompetenzen.



(Rechtschreibung/ Schreibflüssigkeit und Textqualität) in Verbindung zueinander gesetzt. Für eine gezieltere Analyse der Ergebnisse wurden die Schreibflüssigkeit und die Rechtschreibung getrennt voneinander analysiert. Auf Basis der vier resultierenden Profile wurden im folgenden ANOVAs mit den textspezifischen hierarchiehohe Teilfähigkeiten als abhängige Variable berechnet. Die Studie Kreutz et al. (2025) wird im folgenden Kapitel dargestellt.

## **10 Good text quality despite poor transcription skills – An analysis of double dissociation**

**Source:** Kreutz, R.M., Philippek, J., Hennes, A.-K. & Schabmann, A. (2025). Good text quality despite deficient transcription skills – An analysis of double dissociation. *Journal of Writing Research*, 300–321.

**Disclosures and Acknowledgements:** We have no conflict of interest to disclose. We received no financial support for the research. We would like to thank all the children and students who participated in the study.

**Licence:** <https://creativecommons.org/licenses/by-nc-nd/4.0/>

**Changes:** The references cited in this research paper have been adapted and included in the references list at the end of the dissertation. The formatting has also been adapted.

### **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 5<sup>th</sup> to 7<sup>th</sup> grade students had to complete transcription skills tests and write a narrative. Text quality and the 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

## **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, 2007b; 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/ Kreutz et al., 2025). 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/ Kreutz et al., 2025) 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.

## **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, 2014). 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).

For decades, various writing models have sought to explain the subskills which are needed for producing a functional text (e.g., Berninger & Winn, 2006; Hayes, 2012; Kim & Graham, 2022). 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 is 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 underlying 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.

### **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 (Graham, 1990; Hayes & Berninger, 2010; Scardamalia & Bereiter, 1987). Empirical studies support these assumptions by confirming that transcription skills have a significant influence on text quality from 1<sup>st</sup> to 9<sup>th</sup> 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 1<sup>st</sup> to 7<sup>th</sup> grade, and that handwriting fluency had a direct effect on text quality from 3<sup>rd</sup> to 4<sup>th</sup> grade. Cordeiro et al. (2020) were also able to prove the relevance of transcription skills for text quality in 2<sup>nd</sup> graders, while Limpo and Alves (2013) found a direct effect of transcription skills on text quality for 4<sup>th</sup> to 6<sup>th</sup> 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; Hirschler Lichtsteiner et al., 2018; Limpo & Graham, 2020) and weaker intermediate school students (Limpo et al., 2017). Meta-analyses on writing instruction from 1<sup>st</sup> to 12<sup>th</sup> 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 high-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).

### **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/ Kreutz et al. (2025) 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 5<sup>th</sup> to 7<sup>th</sup> 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/ Kreutz et al., 2025).

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).

### **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 these double dissociations 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 5<sup>th</sup> to 7<sup>th</sup> 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?

## **Method**

### **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.<sup>11</sup> Students were between nine and 14 years old ( $M = 11.75$ ;  $SD = 1.05$ ; 48% female). Fifty-seven (30.6%) of these were 5<sup>th</sup> graders, 63 (33.9%) 6<sup>th</sup> graders and 66 (35.5%) 7<sup>th</sup> graders from three classes each.

---

<sup>11</sup> Corresponds to the German school type “Gymnasium” and “Gesamtschule”.

## 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  $\alpha = .94$  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 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 (Jones et al., 2015; Pollitt, 2012a). Therefore, the students' texts 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 (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](http://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 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  $\kappa = .87$  ( $p < .001$ ) for local cohesion errors,  $\kappa = .89$  ( $p < .001$ ) for sentence errors, and  $\kappa = .92$  ( $p < .001$ ) 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.

### **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.

### **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) <sup>a</sup>	-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

<sup>a</sup> theoretical maximum

## Results

### 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*

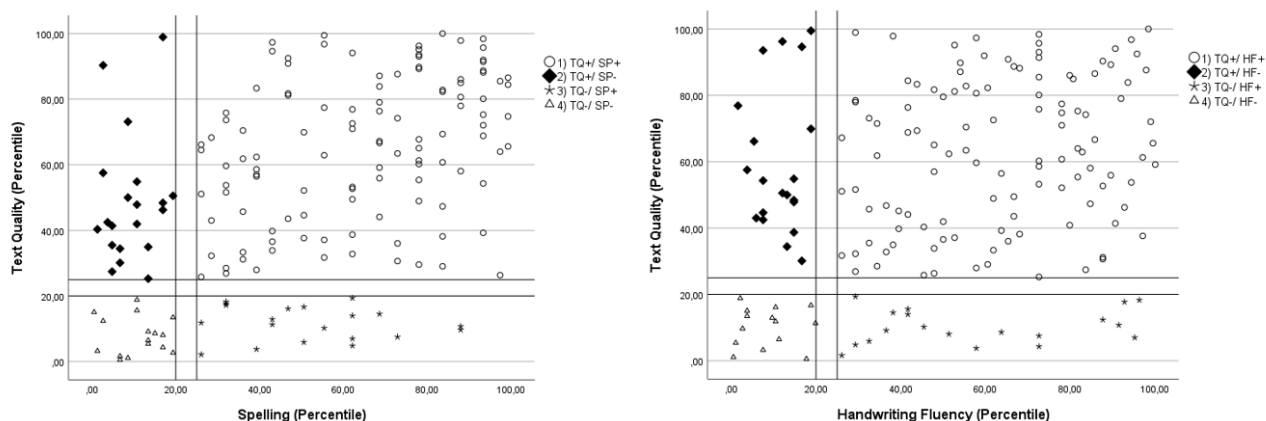
	Handwriting fluency	Spelling	Lexical diversity	Appropriate words	Local cohesion	Correct tense	Correct sentences
Spelling	.41**						
Lexical diversity	.25**	.34**					
Appropriate words	.01	.07	.18*				
Local cohesion	.01	.17*	.19**	.06			
Correct tense	.03	.05	.07	.01	.02		
Correct sentences	.04	.33**	.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 ( $F(3, 166) = 10.37$ ;  $p < .001$ ;  $\eta^2 = .16$ ), correct sentences ( $F(3, 166) = 7.97$ ;  $p < .001$ ;  $\eta^2 = .13$ ), and local cohesion ( $F(3, 166) = 4.12$ ;  $p = .008$ ;  $\eta^2 = .07$ ), 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).

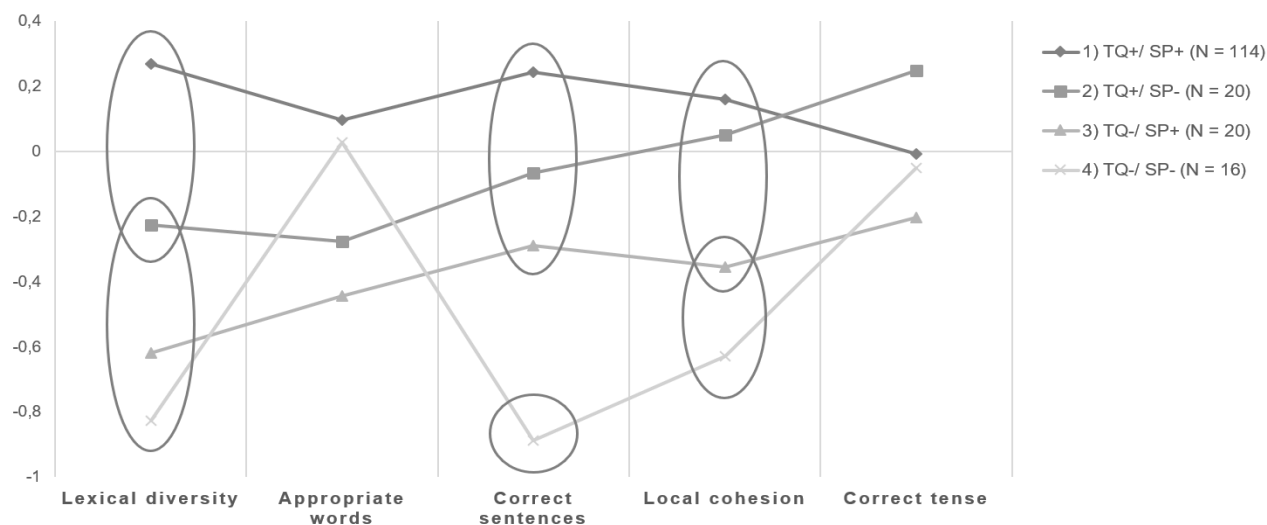
**Table 4**

*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	.27 <sup>a</sup>	-.23 <sup>ab</sup>	-.62 <sup>b</sup>	-.83 <sup>b</sup>	10.37**
Appropriate words	.10	-.028	-.45	.03	2.12
Correct sentences	.24 <sup>a</sup>	-.07 <sup>a</sup>	-.29 <sup>a</sup>	-.89 <sup>b</sup>	7.97**
Local cohesion	.16 <sup>a</sup>	.05 <sup>a</sup>	-.35 <sup>ab</sup>	-.63 <sup>b</sup>	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 Skills*

*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 ( $F(3, 164) = 10.43$ ;  $p < .001$ ;  $\eta^2 = .16$ ), correct sentences ( $F(3, 164) = 8.03$ ;  $p < .001$ ;  $\eta^2 = .13$ ), and local cohesion ( $F(3, 164) = 3.51$ ;  $p = .017$ ;  $\eta^2 = .06$ ), 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*

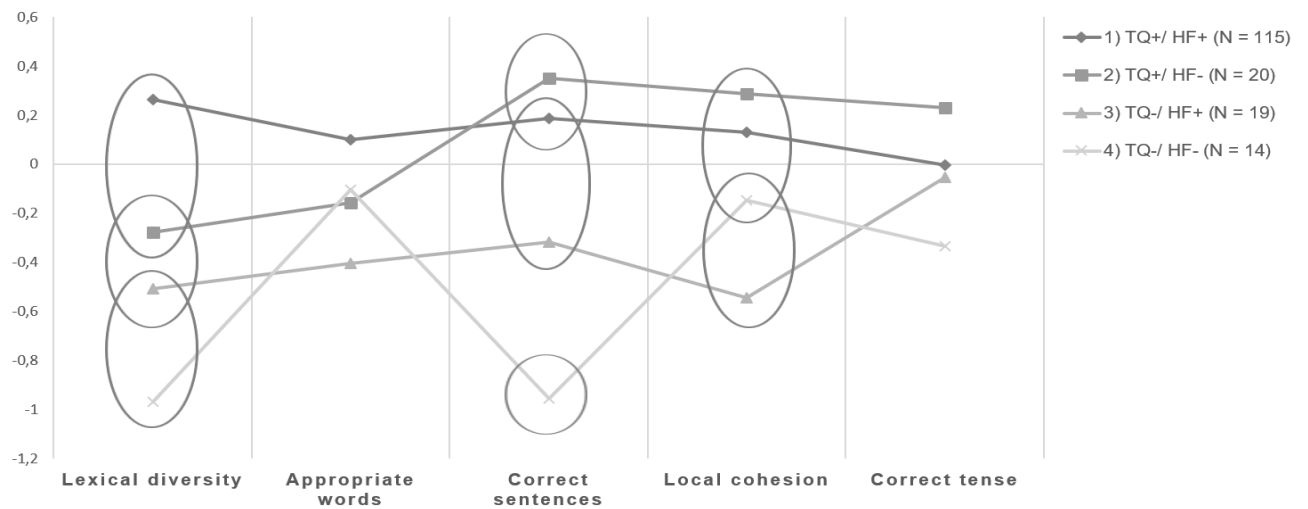
	(1) TQ+/ HF+	(2) TQ+/ HF-	(3) TQ-/ HF+	(4) TQ-/ HF-	F (3, 164)
Lexical diversity	.26 <sup>a</sup>	-.28 <sup>ab</sup>	-.50 <sup>bc</sup>	-.97 <sup>c</sup>	10.43**
Appropriate words	.10	-.16	-.40	-.10	1.65
Correct sentences	.19 <sup>ab</sup>	.35 <sup>b</sup>	-.32 <sup>a</sup>	-.95 <sup>c</sup>	8.03**
Local cohesion	.13 <sup>a</sup>	.29 <sup>a</sup>	-.55 <sup>b</sup>	-.15 <sup>ab</sup>	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.

## Discussion

In this study, we examined dissociations 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 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/ Kreutz et al., 2025), 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 double dissociations were 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 nondissociation 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 high-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.

### **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 5<sup>th</sup> to 7<sup>th</sup> 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.

### **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 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.

### **11 Kreutz et al. (2025): Zwischenfazit**

Die Ergebnisse der Studie Kreutz et al. (2025) zeigen eindeutig, dass die eingangs formulierte Hypothese über das Bestehen doppelter Dissoziationen zwischen Transkriptionsfähigkeiten und Textproduktionskompetenz bestätigt werden kann. Schreibflüssigkeit und Rechtschreibung wurden getrennt betrachtet, zeigen aber ähnliche Ergebnisse. In der Diskussion wird deshalb von Transkriptionsfähigkeiten im Allgemeinen berichtet.

Hinsichtlich der Leistungen in der Transkription und der Textproduktion konnten vier Gruppen definiert werden: Es zeigte sich 1) eine leistungsstarke Gruppe mit sehr guten bis altersadäquaten Leistungen in Transkription und Textproduktion und 2) eine leistungsschwächere Gruppe mit defizitären Leistungen in beiden Bereichen. Zusätzlich wurden zwei Gruppen mit doppelten Dissoziationen identifiziert, nämlich mit 3) schlechten Leistungen in der Transkription, aber guter Textproduktion und 4) dem umgekehrten Leistungsprofil (Forschungsfrage 1). Doppelte Dissoziationen konnte trotz bestätigter Einflüsse der Transkriptionsfähigkeiten auf die Textqualität nachgewiesen werden (Abbott et al., 2010; Cordeiro et al., 2020; Graham & Harris, 1997; Limpo & Alves, 2013). Die Forschung zeigt, dass die Einflüsse mit zunehmender Entwicklungsstufe abnehmen (Nottbusch, 2017; Sturm, 2017). Vielleicht könnte dies nicht alleine durch die fortschreitende Automatisierung begründet werden, sondern zusätzlich – gerade bei den Schüler\*innen, die eine gute Leistung in der Textproduktion zeigen – durch den Erwerb und die zunehmend sichere Verwendung hierarchiehoher textspezifischer Fähigkeiten. Denn Studie 2 konnte ebenfalls zeigen, dass sich die Leistungsprofile in der Verwendung der hierarchiehoher textspezifischen Fähigkeiten lexikalische Vielfalt, lokale Kohäsion und korrekte Syntax signifikant unterscheiden (Forschungsfrage 2). Die Resultate stimmen mit Studien überein, die den isolierten Einfluss auf die Textqualität bereits belegen konnten (Cox et al., 1990; Gómez Vera et al., 2016; MacArthur et al., 2019; Olinghouse & Wilson, 2013).

## 12 Diskussion

Betrachtet man die Ergebnisse der beiden dargestellten Studien gemeinsam, zeigen beide Studien, dass die hierarchieniedrigen Transkriptionsfähigkeiten für die Textqualität in der Sekundarstufe I relevant, aber nicht maßgeblich sind. In Studie Philippek/ Kreutz et al. (2025) manifestiert sich dies in der indirekten Wirkung auf die Textqualität. Diese Ergebnisse decken sich mit der *Hypothese der hierarchischen Beziehungen* aus dem *DIEW* (Kim & Graham, 2022; Kim & Park, 2019), die besagt, dass nicht alle Teilfähigkeiten einen direkten Einfluss auf die Textqualität haben und dass hierarchieniedrige Fähigkeiten sich über die hierarchiehohen Fähigkeiten auf die Textqualität auswirken. Wie die *Hypothese der dynamischen Beziehungen* postuliert, könnte dies darin begründet sein, dass bei Schüler\*innen der Sekundarstufe I die hierarchieniedrigen Fähigkeiten weitestgehend automatisiert sind (Kim & Graham, 2022; Kim & Park, 2019). Es lässt sich vermuten, dass die meisten Schüler\*innen in dieser Transitionsphase, in der der Wechsel von der Primarstufe zur Sekundarstufe I stattfindet, die basalen Fähigkeiten bereits vollständig automatisiert haben oder dass Schüler\*innen mit defizitären Leistungen in der Stichprobe durch die Schulwahl möglicherweise unterrepräsentiert waren. In Studie Kreutz et al. (2025) wurde jedoch gezeigt, dass es durchaus Schüler\*innen gab, die Defizite in den basalen Fähigkeiten aufwiesen. Dennoch waren einige von ihnen in der Lage, gute Texte zu verfassen. Die beiden vorliegenden Studien weisen also übereinstimmend darauf hin, dass der Einfluss der hierarchiehohen Fähigkeiten für die Textqualität von maßgeblicher Relevanz ist. Innerhalb des Modells aus Studie 1 konnte gezeigt werden, dass die hierarchiehohen Fähigkeiten einen direkten Einfluss auf die Textqualität haben und einen Großteil der Varianz dieser erklären konnten. Studie Kreutz et al. (2025) verdeutlicht die Relevanz der hierarchiehohen textspezifischen Fähigkeiten zusätzlich. Hier konnte gezeigt werden, dass die Textqualität von Schüler\*innen mit defizitären Transkriptionsfähigkeiten maßgeblich von der Verwendung der hierarchiehohen textspezifischen Fähigkeiten beeinflusst wird. Schüler\*innen, die zwar

Schwächen in den basalen Fähigkeiten zeigten, aber gute Texte produzierten, schrieben lexikalisch vielfältigere Texte, verwendeten häufiger eine korrekte Syntax und stellten zwischen den Sätzen mehr Kohäsion her. Konkludierend kann festgestellt werden, dass es die hierarchiehohen Fähigkeiten sind, die eine Ansammlung von Wörtern und Sätzen in einen kohärenten und kommunikativ wirksamen Gesamttext verwandeln. Ihre Weiterentwicklung bildet die Grundlage für eine höhere Textqualität.

### **12.1 Limitationen**

Die beiden Studien, die im Rahmen der vorliegenden Arbeit durchgeführt wurden, umfassen einige Aspekte, die in Folgestudien beachtet werden sollten.

Die Textqualität fungiert in der Schreibforschung häufig als Indikator für die (allgemeine) Textproduktionskompetenz. Diese ist, wie alle Kompetenzkonzepte, als latentes Konstrukt nicht direkt beobachtbar (Grabowski et al., 2014). In beiden dargestellten Studien wurde die Textqualität lediglich innerhalb eines Textproduktes (Narrativ) bestimmt und auf Basis dessen auf die Textproduktionskompetenz der Schüler\*innen geschlossen. Studien zeigen zum einen jedoch, dass die Textqualität innerhalb verschiedener Textprodukte nicht immer vergleichbar ist, sodass die Resultate von einem Text nicht allgemeingültig sind und somit nicht generalisiert werden können. Zum anderen ist bekannt, dass die Einflüsse der verschiedenen Teilfähigkeiten unabhängig von der hierarchischen Ebene je nach Genre variieren (Beers & Nagy, 2009; Kim & Graham, 2022). Selbst Transkriptionsfähigkeiten zeigen je nach Genre unterschiedlich starke Einflüsse auf die Textqualität (Berninger et al., 2010; Salas & Silvente, 2020). Graham et al. (2016) plädieren daher dafür, für Forschungszwecke mehr als eine Schriftprobe der Schüler\*innen zu erheben und bewerten zu lassen und empfehlen, die Schreibaufgaben hinsichtlich der Genres und Themen möglichst heterogen zu gestalten.

Des Weiteren beziehen sich die Ergebnisse der beiden dargestellten Studien lediglich auf Schüler\*innen der 5. bis 7. Klasse, die das Abitur anstreben. Für die vorliegenden Studien

wurde die Wahl diese Stichprobe als sinnvoll erachtet. Denn für die Validierung eines neuen Modells eignet sich in erster Instanz eine normalentwickelte Stichprobe, die keine Lernstörungen wie eine Lese-Rechtschreib-Störung aufweist. Außerdem sollten die textspezifischen Fähigkeiten, deren Anwendung einiger Übung bedarf, im Unterricht bereits Berücksichtigung finden, was bei der vorliegenden Stichprobe anzunehmen war. Beide Voraussetzungen wurden also durch die vorliegende Stichprobe erfüllt. Da aber nun in beiden dargestellten Studien der Ist-Stand erhoben wurde, wäre es natürlich in weiteren Studien interessant, andere Stichproben hinzuzuziehen, um die Allgemeingültigkeit des Modells und der doppelten Dissoziationen hinsichtlich Transkriptionsfähigkeiten und Textproduktionskompetenz beurteilen zu können. Hierzu würden sich für weitere Studien sowohl jüngere Schüler\*innen aus der Primarstufe als auch ältere Schüler\*innen aus der Sekundarstufe II sowie erfahrene erwachsene Schreibende eignen, da die Interaktionen der Teilfähigkeiten entwicklungsbedingt dynamisch ist (Kim & Park, 2019). Ebenso wäre es interessant, die Analysen mit Daten von anderen Schulformen oder sogar von Schüler\*innen mit Lernschwierigkeiten zu wiederholen, um unterschiedliche Entwicklungsverläufe miteinander vergleichen zu können. Gerade die Untersuchung letzterer Stichprobe wäre von großer Relevanz, da es sich dabei um förderbedürftige Schüler\*innen handelt.



### 13 Fazit und Ausblick

Die Studie Philippek/ Kreutz et al. (2025) liefert ein neues Modell für den Forschungsbereich der Schreibforschung, welches Teilfähigkeiten verschiedener hierarchischer Ebenen, die Exekutivfunktionen sowie die Textproduktionskompetenz gemeinsam abbildet. Das Modell postuliert einen kaskadierenden Einflusspfad, bei dem basale Transkriptionsfähigkeiten die Textproduktionskompetenz indirekt über die hierarchiehohen Teilfähigkeiten beeinflussen, während diese wiederum einen direkten Einfluss auf die Textproduktionskompetenz ausüben.

Die Studie Kreutz et al. (2025) betrachtete den Zusammenhang von Transkriptionsfähigkeiten und Textproduktionskompetenz differenzierter und konnte darüber hinaus zeigen, dass der Textproduktionskompetenz – gemessen anhand der Textqualität – andere Teilfähigkeiten als die Transkriptionsfähigkeiten zugrunde liegen und doppelte Dissoziationen vorliegen. Außerdem konnte festgestellt werden, dass die hierarchiehohen Teilfähigkeiten individuelle Leistungsunterschiede in der Textqualität erklären können.

Zusammenfassend tragen beide Studien zu einem differenzierteren Verständnis der Textproduktionskompetenz im Hinblick auf die zugrundeliegende Interaktion ihrer Teilfähigkeiten bei und liefern wichtige Implikationen für die schulische Praxis und Forschung.

#### 13.1 Implikationen für die Praxis

Ein großer Vorteil des *CASMOW* ist, dass die Ergebnisse der Studie Philippek/ Kreutz et al. (2025) Implikationen für die Praxis aufzeigen, da textspezifische Fähigkeiten aufgenommen wurden, die sich innerhalb des Schreibunterrichts explizit trainieren lassen. In der Studie Philippek/ Kreutz et al. (2025) wurde bei der Validierung des Modells bereits gezeigt, dass die hierarchiehohen Fähigkeiten einen Großteil der Varianz der Textqualität ausmachen und somit auch für den Schreibunterricht von größerer Bedeutung sein sollten. Basierend auf diesen Ergebnissen scheint es also nicht ratsam, sich bei der Förderung der

Textproduktionskompetenz nur auf die hierarchieniedrigen Transkriptionsfähigkeiten zu fokussieren. Natürlich bilden die Transkriptionsfähigkeiten die Basis für die Förderung der hierarchiehohen textspezifischen Fähigkeiten und es ist wahrscheinlich, dass die Förderung dieser effektiver verläuft, wenn die Schüler\*innen mit der Handschrift und der Orthografie vertraut sind (Glaser et al., 2011). Studie Kreutz et al. (2025) konnte jedoch ebenfalls zeigen, dass die textspezifischen Fähigkeiten ausschlaggebend für die Qualität der Schüler\*innentexte sind, selbst bei schwachen Transkriptionsfähigkeiten. Es kristallisierte sich eine Gruppe von Schüler\*innen heraus, die trotz schwacher Leistung in den Transkriptionsfähigkeiten in der Lage waren, gute Texte zu produzieren. Dies geht mit anderen Studien einher, die belegen, dass Schüler\*innen bereits in der Lage sind, Texte zu produzieren, bevor sie über die Kompetenz verfügen, eine korrekte Rechtschreibung konsequent anzuwenden (Palincsar & Klenk, 1993).

Weiterführend konnte Studie Kreutz et al. (2025) zeigen, dass verschiedene Gruppen von Schüler\*innen bestehen: Für die Förderung relevant sind insbesondere diejenigen, die zwar gute Texte produzieren, jedoch Defizite in den Transkriptionsfähigkeiten zeigen; jene, die schlechte Texte produzieren trotz guter Leistung in den Transkriptionsfähigkeiten; und jene, die sowohl in den Transkriptionsfähigkeiten als auch in der Textproduktion Defizite aufweisen. Liegt der Fokus im Schreibunterricht beispielsweise auf der isolierten Förderung der Transkriptionsfähigkeiten, kommt es zu einem unausgewogenen Verhältnis zwischen hierarchieniedrigen und -höheren Teilfertigkeiten (Cordeiro et al., 2020; Graham & Harris, 1997; Moats et al., 2006). Es besteht die Gefahr, dass die Differenzen innerhalb der Textproduktionskompetenz zwischen den Schüler\*innen immer größer werden, da die Entwicklung der hierarchiehohen textspezifischen Teilfähigkeiten nur sehr verzögert erfolgt. Das Vorkommen verschiedener Gruppen verdeutlicht die Relevanz der individuellen Anpassung der Förderung auf die jeweiligen Schüler\*innen. In diesem Kontext wäre die Entwicklung eines Förderprogramms denkbar, das eine Vielzahl individuell kombinierbarer

Module umfasst, bestehend aus Trainingsmöglichkeiten für sowohl hierarchieniedrige als auch hierarchiehohe Teilfähigkeiten. Die Verwendung eines solchen sogenannten Baukastenprinzips, das individuell an die Schüler\*innen angepasst werden kann, ist bereits aus anderen kognitiven Interventionsbereichen bekannt, wie beispielsweise aus kognitiven Gedächtnistrainings oder der Dyskalkulietherapie (z.B. Neurotraining: Schweizer & Wachter-Müller, 2017; Kalkulie: Gerlach et al., 2007).

### **13.2 Weitere Forschungsvorhaben**

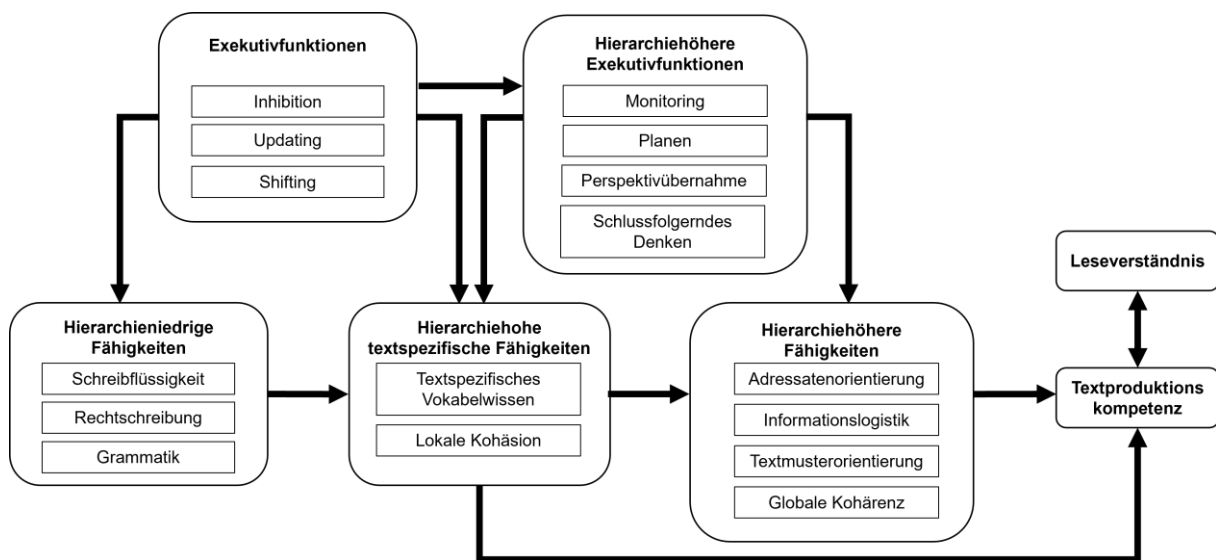
Das Forschungsgebiet der Schreibforschung ist zum einen noch relativ jung und zum anderen sehr komplex und umfangreich, wodurch noch zahlreiche Forschungsfragen unbeantwortet sind. Die Ergebnisse der beiden dargestellten Studien eröffnen die Möglichkeit, interessante Folgestudien zu entwickeln: In der Studie Philippek/ Kreutz et al. (2025) wurde mit *CASMOW* gezeigt, wie die Exekutivfunktionen, die hierarchieniedrigen Transkriptionsfähigkeiten und die hierarchiehohen textspezifischen Fähigkeiten miteinander interagieren und wie groß deren Einfluss auf die Textqualität bei Schüler\*innen der Sekundarstufe I ist. Mit Hilfe des Modells konnten 54% der Varianz in der Textqualität erklärt werden, jedoch bleibt somit fast die Hälfte der Varianz noch unerklärt. Dieses Erkenntnis ist insofern nicht überraschend, als dass die Schreibforschung bereits weitere textspezifische sowie schreibunspezifische Teilfähigkeiten identifiziert hat, die in Isolation oder in Kombination mit einzelnen Variablen ebenfalls einen Einfluss auf die Textqualität ausüben. Das Konvergenzmodell (Hennes et al., 2018; siehe Anhang B) umfasst weitere relevante Fähigkeiten, die im *CASMOW* ergänzend berücksichtigt werden sollten (Abbildung 7). Im Konvergenzmodell sind konkrete Teilfähigkeiten verortet, auf die die Schreibenden während des Schreibprozesses zurückgreifen können. Im Rahmen der zukünftigen Erweiterung von *CASMOW* sollen die *Adressatenorientierung*, *Informationslogistik*, *Textmusterorientierung* und *globale Kohärenz* ergänzt werden. Die Relevanz der einzelnen Kompetenzen für die Produktion funktionaler Texte konnte bereits bestätigt werden

(Adressatenorientierung: Becker-Mrotzek & Schindler, 2007; Informationslogistik: Schnotz, 2006; Textmusterorientierung: Schmitt & Knopp, 2017; Kohärenz: Averintseva-Klisch, 2018). Da ein angemessenes Vokabular und Kohäsion zwischen den einzelnen Sätzen beispielsweise dazu eingesetzt werden könnte, globale Kohärenz über den ganzen Text hinweg herzustellen oder Informationen innerhalb des Textes zu strukturieren, wird davon ausgegangen, dass sich die genannten Fähigkeiten auf einer weiteren, hierarchiehöheren Kaskade in *CASMOW* befinden. Nicht nur die hierarchiehohen Teilfähigkeiten bedürfen einer Ergänzung: Auf der hierarchieniedrigeren Ebene können *Grammatikkenntnisse* (insbesondere morphologische Kenntnisse) zum Modell hinzugefügt werden (Neumann & Lehmann, 2008). Denn bereits im *DIEW* konnte ein indirekter Einfluss dieser über die hierarchiehohen kognitiven und sprachlichen Teilfähigkeiten gezeigt werden (Kim & Graham, 2022; Kim & Park, 2019). Auch die Exekutivfunktionen können detaillierter dargestellt werden. Wolf et al. (2025) konnte beispielsweise zeigen, dass die Exekutivfunktionen *Inhibition*, *Updating* und *Shifting* bei geübten Schreibenden unterschiedlich mit den hierarchiehohen Teilfähigkeiten und der Textqualität zusammenhängen. So hatte *Inhibition* einen signifikanten Einfluss auf Kohäsion und Informationsgehalt, während *Updating* die angemessene Wortwahl und *Shifting* die Textqualität signifikant beeinflussten. Aber auch bei sich in der Entwicklung befindenden Schreibenden konnte gezeigt werden, dass die drei basalen Exekutivfunktionen als Prädiktor für die Textqualität gelten (Rocha et al., 2022), einen Einfluss auf die hierarchieniedrigen Transkriptionsfähigkeiten (Altemeier et al., 2008) sowie auf verschiedene hierarchiehohe Fähigkeiten haben (z.B. *Inhibition* indirekt über Schreibflüssigkeit auf Informationsgehalt: Drijbooms et al., 2015; *Updating* auf lexikalische Vielfalt und Kohärenz: Bourke & Adams, 2003). Ebenso nicht textspezifisch, aber dennoch einflussreich sind die Schreibstrategien. Limpo und Alves (2013) konnten in ihrer empirischen Studie zeigen, dass insbesondere Planen und Überarbeiten die Textqualität signifikant beeinflussen. Diese Strategien stehen in engem Zusammenhang mit den hierarchiehöheren Exekutivfunktionen. Entsprechend können

diese können als Stellvertretervariable für die Schreibstrategien dem Modell hinzugefügt werden. Kim und Graham (2022) konnten bereits im *DIEW* zeigen, dass hierarchiehöhere kognitive Fähigkeiten wie das *schlussfolgernde Denken* und die *Perspektivübernahme* besonders wichtig für eine kohärente, klare Formulierung und Anordnung von Ideen und somit für die Textqualität sind.

### Abbildung 7

#### *Cascaded Model of Writing Expanded*



*Anmerkung.* Die Abbildung zeigt die Erweiterung des *CASMOW* und veranschaulicht, wie die jeweiligen Teilfähigkeiten in die Kaskaden integriert werden können.

Während sich bei der Erstellung von *CASMOW* bewusst dafür entschieden wurde nur textspezifische Fähigkeiten mit in das Modell aufzunehmen, da somit ein stärkerer Bezug zur Praxis hergestellt werden kann, gibt es andere Modelle und Studien, die bereits gezeigt haben, dass auch textunspezifische Fähigkeiten einen wichtigen Beitrag zur Produktion funktionaler Texte leisten. Das *DIEW* operationalisierte die Textgenerierung durch den expressiven Wortschatz und erhob das mündliche Vokabelwissen (oral language skills) (Kim & Graham, 2022; Kim & Park, 2019). Trotz der berechtigten Kritik, die Textgenerierung mündlich zu erheben, könnte es dennoch interessant sein, den Einfluss *mündlicher Sprachkenntnisse* als Prädiktor in *CASMOW expanded* aufzunehmen. Schließlich ist bekannt, dass mündliche

Sprachkenntnisse in direktem Zusammenhang mit der inhaltlichen Qualität von Schreibprodukten stehen und dass Kinder, gerade zu Beginn des Schriftspracherwerbs, besser in der Produktion mündlicher als schriftlicher Erzählungen sind und die mündliche Produktion als Prädiktor der Verschriftlichung gilt (Harrison et al., 2016; Kim et al., 2013).

Eine weitere nicht textspezifische Modalität, die in *CASMOW expanded* ergänzt werden sollte, ist das *Lesen*. Lesen und Schreiben beruhen auf gemeinsamen Fähigkeiten wie beispielsweise dem prozeduralen Wissen sowie Textmusterwissen und Wissen über universelle Textmerkmale wie die Orthografie und Syntax (Fitzgerald & Shanahan, 2000). Außerdem sind sowohl die Leseflüssigkeit als auch das Leseverständnis essentielle Fähigkeiten für den Prozess des Überarbeitens während der Textproduktion und somit auch entscheidende Faktoren für die Qualität des Textes (Breetvelt et al., 1996; Deane et al., 2008). So konnten Kim und Graham (2022) im *DIEW* bereits zeigen, dass das Leseverständnis einen signifikanten Einfluss auf die Textqualität hat.

Des Weiteren wäre es von wissenschaftlichem Interesse, *CASMOW* für andere Genres zu validieren. In diesem Kontext ist eine Veränderung der Einflüsse der einzelnen Teilfähigkeiten auf die Textqualität sowie deren Interaktion zu erwarten, da verschiedene Genres und Schreibaufgaben verschiedene Fähigkeiten erfordern (Beers & Nagy, 2009; Plakans & Gebril, 2013). Im Hinblick auf die Analyse von Essays zeigte sich beispielsweise, dass die syntaktische Komplexität, die lexikalische Vielfalt sowie die Kohäsion als signifikante Teilfähigkeiten für eine hohe Qualität zu erachten sind (Beers & Nagy, 2009; Crossley & McNamara, 2010). Im Rahmen von Sachtexten hingegen untersuchte Wolf et al. (2025) bei erwachsenen Schreibenden die Interaktion der basalen Exekutivfunktionen und hierarchiehoher textspezifischer Fähigkeiten sowie deren Zusammenhang mit der Textqualität. Die Ergebnisse zeigten, dass Informationsgehalt und Kohäsion signifikant mit der Textqualität zusammenhängen, während die lexikalische Vielfalt, die angemessene Wortauswahl und die syntaktische Komplexität keinen Zusammenhang zeigten (Wolf et al.,

2025). Die Ergebnisse verdeutlichen die Unterschiede zwischen verschiedenen Genres, womit bestätigt werden kann, dass sowohl *CASMOW* als auch *CASMOW expanded* für andere Genres validiert werden müssen.

Das bislang validierte *CASMOW* gilt für Schüler\*innen der 5. bis 7. Jahrgangsstufe. Die Textproduktionskompetenz ist jedoch kein starres Konstrukt, das immer gleichbleibend aus bestimmten Fähigkeiten besteht (Crossley et al., 2014). Der Einfluss auf die Textqualität und die Interaktion verschiedenster textspezifischer und kognitiver Fähigkeiten verändert sich im Verlauf der Schreibentwicklung (Kim & Graham, 2022; Kim & Park, 2019). In der Primarstufe sind die Schüler\*innen hauptsächlich damit beschäftigt, die basalen Transkriptionsfähigkeiten zu erlernen und zu automatisieren. Deshalb nehmen sie hier auch in der Produktion von Texten einen größeren Stellenwert ein, welcher jedoch mit zunehmender Entwicklung im Schreiben immer geringer wird, während sich die hierarchiehohe Fähigkeiten zunehmend ausbilden. Laut Schreibentwicklungsmodellen kommt es in der Mittelstufe (7.-9. Jahrgangsstufe) zu einer Verwendung der sprachlichen und grammatikalischen Mittel nach dem Ökonomieprinzip, was bedeutet, dass kohäsive Mittel beispielsweise deutlich häufiger als üblich verwendet werden. Dies normalisiert sich in der Adoleszenz wieder (Becker-Mrotzek & Böttcher, 2012). Daher wäre es jedoch interessant, sowohl *CASMOW* als auch das *CASMOW expanded* im Entwicklungsverlauf von der Primarstufe bis hin zur Oberstufe zu validieren und die Veränderungen der Einflüsse der einzelnen Teilfähigkeiten auf die Textqualität und Veränderungen der Interaktion zwischen den einzelnen Teilfähigkeiten zu modellieren. Eine Querschnittsanalyse von Kindern der ersten Klasse bis hin zu erwachsenen Schreibende könnte einen ersten Eindruck vermitteln. Eine Längsschnittstudie, in welcher festgestellt werden könnte, welche Fähigkeiten zu Beginn des Schreibens als Prädiktor für eine spätere gute Textqualität gelten, wäre sicherlich noch aufschlussreicher.

Die Studie Kreutz et al. (2025) konnte bestätigen, dass die basalen Teilfähigkeiten bei 5.-7. Klässler\*innen nicht mehr essentiell für die Produktion von Texten sind und dass Schüler\*innen mit Defiziten darin dennoch in der Lage sind, gute Texte zu produzieren. In Bezug auf die Studie Kreutz et al. (2025) wäre es von Interesse, die auf den doppelten Dissoziationen basierenden Gruppen über die Schreibentwicklung hinweg zu betrachten, um die Veränderungen der Teilfähigkeiten innerhalb der vier Gruppen zu analysieren. Dabei sollten auch verschiedene Klassen aus verschiedenen Schulen einbezogen werden, um den Einfluss schulischer Lehrmethoden im Schreibunterricht zu berücksichtigen – insbesondere, da die Schreibentwicklung in der Sekundarstufe I je nach Lehrmethode stark variieren kann (Becker-Mrotzek & Böttcher, 2012). Während einige Lehrkräfte über die Primarstufe hinaus einen Schwerpunkt auf die Förderung der Transkriptionsfähigkeiten legen (Cordeiro et al., 2020; Graham & Harris, 1997; Moats et al., 2006), sollten gemäß der Schreibentwicklungsmodelle bereits ab der Primarstufe auch die hierarchiehohe Teilfähigkeiten gelehrt und in der Sekundarstufe I vertieft werden (Kapitel 3.6). Dabei wird einerseits für eine explizite Instruktion der einzelnen Teilfähigkeiten plädiert und andererseits wird die Auffassung vertreten, dass die Teilfähigkeiten implizit während der Textproduktion trainiert werden sollten (Graham, 2024; Sturm, 2017). Ein interessanter Forschungsansatz bestünde darin, die Zusammenhänge zwischen Lehrmethoden und den Ausprägungen der Leistungen in den vier Profilen systematisch zu untersuchen.

Auch methodisch ist eine weitere Forschungsfrage offen: In den vorliegenden Studien wurden die hierarchiehohen Teilfähigkeiten auf Textebene analysiert. Dies hatte den Vorteil, dass die hierarchiehohen Teilfähigkeiten kontextabhängig erfasst werden konnten. Die Erfassung der Maße mittels standardisierter Tests könnte, wie in Kapitel 4.1 beschrieben, auch einige Vorteile haben (z.B. bessere Vergleichbarkeit zwischen Schüler\*innen). Die Frage, die sich daraus ergibt, ist, welche Art der Messung die Konstrukte besser erfassen kann. Zukünftig



sollten also die Messungen der einzelnen Konstrukte im Test mit den Messungen im Text verglichen werden und in Bezug zur Textqualität gestellt werden.

## Anhang A

**Table 3**

*Overview of measurement in cited studies<sup>12</sup>*

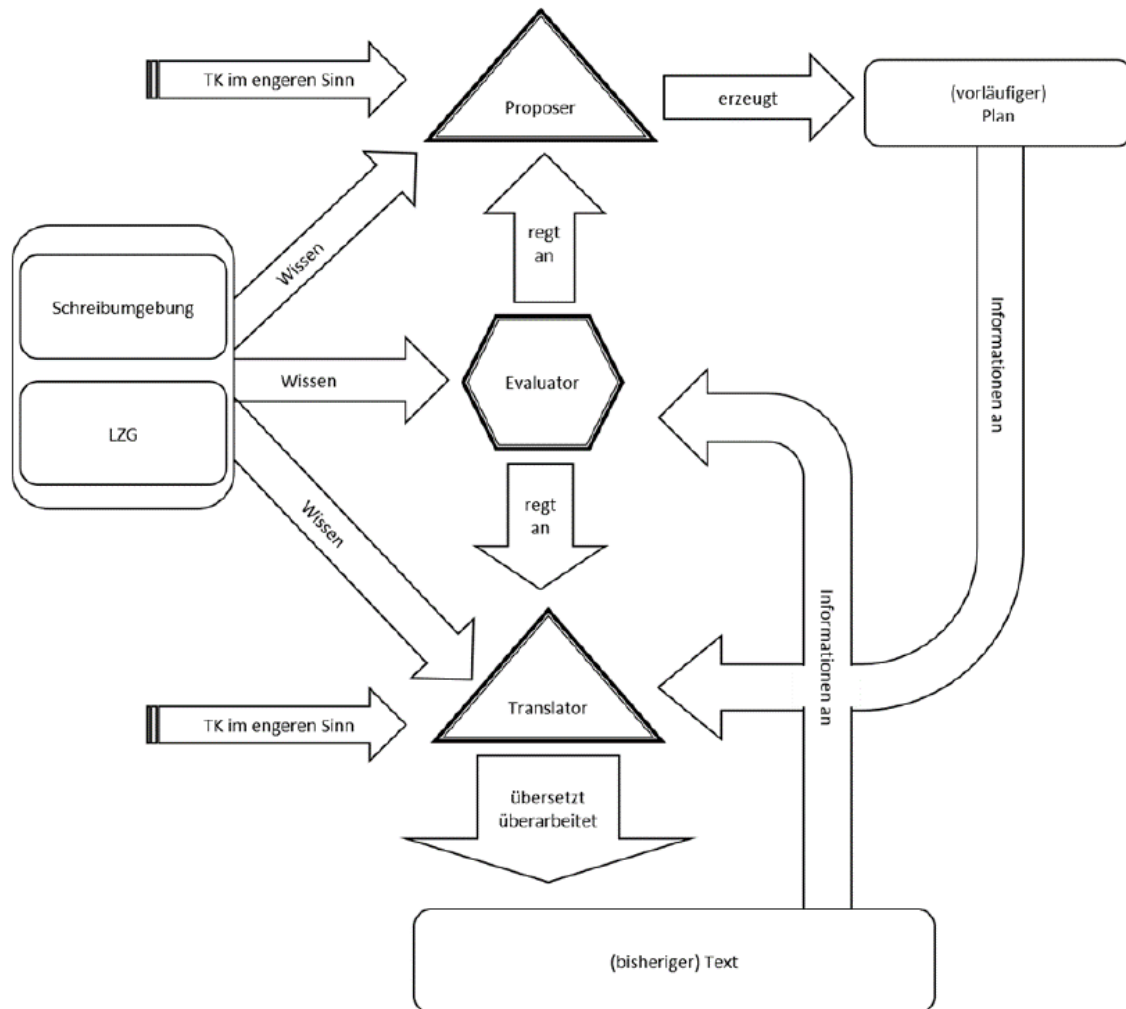
Subskill	Measurement
Handwriting Fluency	<ul style="list-style-type: none"> <li>Standardized alphabet task (Abbott et al., 2010; Berninger et al., 2010; Connelly et al., 2012; Cordeiro et al., 2020; Graham et al., 1997; Salas &amp; Silvente, 2020)</li> <li>Standardized sentence copying task (Berninger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

<sup>12</sup> Anhang aus der Studie Philippek/ Kreutz et al. (2025)

## Anhang B

## Abbildung 8

Das Konvergenzmodell (Hennes et al., 2018)



**Literaturverzeichnis**

- 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>
- Altemeier, L. E., Abbott, R. D. & Berninger, V. W. (2008). Executive functions for reading and writing in typical literacy development and dyslexia. *Journal of clinical and experimental neuropsychology*, 30(5), 588–606.  
<https://doi.org/10.1080/13803390701562818>
- 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>
- Amrein, A. (2017). *Förderung der Schreibkompetenz in der gymnasialen Oberstufe mit kreativen Schreibmethoden* [Dissertation, Universität Bremen]. <https://dnb.info/1162273224/34>
- Arfé, B. & Pizzocaro, E. (2016). Sentence Generation in Children With and Without Problems of Written Expression. In J. Perera, M. Aparici, E. Rosado & N. Salas (Hrsg.), *Written and Spoken Language Development across the Lifespan* (S. 327–344). Springer. [https://doi.org/10.1007/978-3-319-21136-7\\_19](https://doi.org/10.1007/978-3-319-21136-7_19)
- Averintseva-Klisch, M. (2018). *Textkohärenz* (2., aktualisierte Aufl.). *Kurze Einführungen in die germanistische Linguistik - KEGLI: Band 14*. Universitätsverlag Winter.
- Baer, M., Fuchs, M., Reber-Wyss, M., Jurt, U. & Nussbaum, T. (1995). Das „Orchester-Modell“ der Textproduktion. In J. Baurmann & R. Weingarten (Hrsg.), *Schreiben:*

- Prozesse, Prozeduren und Produkte* (S. 173–200). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-322-97050-3\\_9](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/10.1348/000709909x421937>
- Becker-Mrotzek, M. (2022). Schreibkompetenz: Überlegungen zu einem didaktischen Konstrukt. In M. Becker-Mrotzek & J. Grabowski (Hrsg.), *Schreibkompetenz in der Sekundarstufe: Theorie, Diagnose und Förderung* (S. 9–28). Waxmann.
- Becker-Mrotzek, M. & Böttcher, I. (2012). *Schreibkompetenz entwickeln und beurteilen: Sekundarstufe I/II* (4., überarb. Neuaufl.). *Scriptor Praxis*. Cornelsen.
- Becker-Mrotzek, M., Brinkhaus, M., Grabowski, J., Hennecke, V., Jost, J., Knopp, M., Schmitt, M., Weinzierl, C. & Wilmsmeier, S. (2015). Kohärenzherstellung und Perspektivenübernahme als Teilkomponenten der Schreibkompetenz: von der diagnostischen Absicherung zur didaktischen Implementierung. In A. Redder, J. Naumann & R. Tracy (Hrsg.), *Forschungsinitiative Sprachdiagnostik und Sprachförderung - Ergebnisse* (S. 177–205). Waxmann.
- 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>

- Becker-Mrotzek, M. & Schindler, K. (2007). Schreibkompetenz modellieren. In M. Becker-Mrotzek & K. Schindler (Hrsg.), *Kölner Beiträge zur Sprachdidaktik Reihe A: Bd. 5. Texte schreiben* (S. 7–26). Gilles & Francke.
- 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. (1980). Development in Writing. In L. Gregg & E. R. Steinberg (Hrsg.), *Cognitive Processes in Writing* (S. 21). Routledge.
- Berman, R. & Nir-Sagiy, B. (2007). Comparing Narrative and Expository Text Construction Across Adolescence: A Developmental Paradox. *Discourse Processes*, 2(43), 79–120.
- Berninger, V. W. (1996). *Reading and Writing Acquisition*. Routledge.  
<https://doi.org/10.4324/9780429303821>
- Berninger, V. W. (2000). Development of Language by Hand and Its Connections with Language by Ear, Mouth, and Eye. *Topics in Language Disorders*, 20(4), 65–84.
- 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))
- Berninger, V. W. & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/ or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris & S. Graham (Hrsg.), *Handbook of learning disabilities* (S. 345–363). Guilford Press.
- 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. & 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., 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 (Hrsg.), *Handbook of Writing Research* (S. 96–114). Guilford Press.
- Berninger, V. W., Yates, C., Cartwright, A., Rutberg, J., Remy, E. & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing*, 4(3), 257–280. <https://doi.org/10.1007/BF01027151>
- Birkel, P. & Birkel, C. (2002). Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 49(3), 219–224.
- Böhme, K., Bremerich-Vos, A. & Robitzsch, A. (2009). Aspekte der Kodierung von Schreibaufgaben. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 290–329). Beltz.
- 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, 1–11. <https://doi.org/10.1016/j.jcomdis.2019.105969>

- Bourke, L. & Adams, A. M. (2003). The relationship between working memory and early writing assessed at the word, sentence and text level. *Educational and Child Psychology*, 20(3), 19–36.
- Bouwer, R., Koster, M. & van den Bergh, H. (2023). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. *Assessment in Education: Principles, Policy & Practice*, 30(3-4), 302–319.  
<https://doi.org/10.1080/0969594X.2023.2241656>
- Bouwer, R., Lesterhuis, M., Smedt, F. de, van Keer, H. & Maeyer, S. de (2024). Comparative approaches to the assessment of writing: Reliability and validity of benchmark rating and comparative judgement. *Journal of Writing Research*, 15(3), 497–518.  
<https://doi.org/10.17239/jowr-2024.15.03.03>
- 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](https://doi.org/10.1207/s1532690xci1202_2)
- Breetvelt, I., van den Bergh, H. & Rijlaarsdam, G. (1996). Rereading and generating and their relationship to text quality. An application of multilevel analysis on writing process data. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Hrsg.), *Theories, models, and methodology in writing research* (S. 10–20). Amsterdam University Press.
- 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>
- Chenoweth, N. A. & Hayes, J. R. (2001). Fluency in Writing. *Written Communication*, 18(1), 80–98. <https://doi.org/10.1177/0741088301018001004>
- 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>
- Cooksey, R. W., Freebody, P. & Wyatt-Smith, C. (2007). Assessment as Judgment-in-Context: Analysing how teachers evaluate students' writing. *Educational Research and Evaluation*, 13(5), 401–434. <https://doi.org/10.1080/13803610701728311>
- 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. (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. (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. (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)
- Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C. & Bivens-Tatum, J. (2008). Cognitive Models of Writing: Writing Proficiency as a Complex Integrated Skill. *ETS*

- Research Report Series*. Vorab-Onlinepublikation. <https://doi.org/10.1002/j.2333-8504.2008.tb02141.x>
- Diamond, A. (2013). Executive Functions. *Annual review of psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- 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>
- Dunn, J. C. & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex*, 39(1), 1–7. [https://doi.org/10.1016/S0010-9452\(08\)70070-4](https://doi.org/10.1016/S0010-9452(08)70070-4)
- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), 155–185. <https://doi.org/10.1177/0265532207086780>
- Erberich, M. (2022). Die Wortwahl. In M. Erberich (Hrsg.), *Einfach und verständlich schreiben: Techniken von Profis für Beruf und Studium* (S. 5–15). Springer. [https://doi.org/10.1007/978-3-662-66276-2\\_2](https://doi.org/10.1007/978-3-662-66276-2_2)
- Fay, J. (2010). *Die Entwicklung der Rechtschreibkompetenz beim Textschreiben: Eine empirische Untersuchung in Klasse 1 bis 4. Theorie und Vermittlung der Sprache: Bd. 53*. Peter Lang. <https://doi.org/10.3726/978-3-653-00718-3>
- Feenstra, H. (2014). *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>
- Fitzgerald, J. & Shanahan, T. (2000). Reading and Writing Relations and Their Development. *Educational Psychologist*, 35(1), 39–50. [https://doi.org/10.1207/S15326985EP3501\\_5](https://doi.org/10.1207/S15326985EP3501_5)

- Fix, M. (2010). *Texte schreiben: Schreibprozesse im Deutschunterricht* (2. Aufl.). *Standard Wissen Lehramt*. Schöningh.
- 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, 1–10. <https://doi.org/10.3389/fpsyg.2020.562462>
- Forkarth, C. & Manzel, S. (2024). Diagnoseinstrumente zur Beurteilung und Förderung von Schreibkompetenzen im sozialwissenschaftlichen Fachunterricht. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Sprachliche Bildung: Bd. 11. Texte schreiben in allen Unterrichtsfächern: Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 131–148). Waxmann.
- Gegenfurtner, K. R., Goldstein, E. B. & Cacciamani, L. (Hrsg.). (2022). *Lehrbuch. Wahrnehmungspsychologie: Der Grundkurs* (10. vollständig überarbeitete und aktualisierte Aufl.). Springer. <https://doi.org/10.1007/978-3-662-65146-9>
- Gerlach, M., Fitz, A., Ricken, G. & Schmidt, S. (2007). *KALKULIE: Diagnose- und Trainingsprogramm für rechenschwache Kinder* (1. Aufl.). College Entrance Examination Board.
- Glaser, C., Keßler, C. & Palm, D. (2011). *Aufsatztraining für 5. bis 7. Klassen: Ein Manual für Lehrkräfte mit Unterrichtsmaterialien*. Hogrefe Verlag.
- Gniewosz, B. (2011). Kompetenzentwicklung. In H. Reinder, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Empirische Bildungsforschung: Gegenstandsbereiche* (S. 69–82). Springer.
- Goblirsch, G. (2017). Was zum Textverständnis nötig ist. In G. Goblirsch (Hrsg.), *Gebrauchstexte schreiben: Systemische Textmodelle für Journalismus und PR* (S. 3–17). Springer Fachmedien. [https://doi.org/10.1007/978-3-658-17601-3\\_2](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 (Hrsg.), *Executive functions and writing* (S. 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>
- Grabowski, J., Becker-Mrotzek, M., Knopp, M., Jost, J. & Weinzierl, C. (2014). Comparing and combining different approaches to the assessment of text quality. In D. Knorr, C. Heine & J. Engberg (Hrsg.), *Methods in Writing Process Research* (S. 147–165). Peter Lang.
- 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. (2024). Writing Instruction and Intervention Research: A Historical and Theoretical Perspective. In F. de Smedt, R. Bouwer, T. Limpo & S. Graham (Hrsg.), *Studies in writing: volume 40. Conceptualizing, designing, implementing, and evaluating writing interventions* (S. 11–34). Brill. [https://doi.org/10.1163/9789004546240\\_003](https://doi.org/10.1163/9789004546240_003)
- 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 (Hrsg.), *Studies in writing: volume 34. Design principles for teaching effective writing: Theoretical and empirical grounded principles* (S. 13–37). Brill.  
[https://doi.org/10.1163/9789004270480\\_003](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, G. 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. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476.  
<https://doi.org/10.1037/0022-0663.99.3.445>
- Graham, S. & Perin, D. (2007b). *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>

- Hales, L. W. & Tokar, E. (1975). The Effect of the Quality of Preceding Responses on the Grades Assigned to Subsequent Responses to an Essay Question. *Journal of Educational Measurement*, 12(2), 115–117. <https://doi.org/10.1111/j.1745-3984.1975.tb01014.x>
- Harris, K. R. (2024, 13. Juni). *Four Reasons Writing Instruction Needs a Makeover - And Evidence-Based Solutions*. Learning Without Tears. <https://www.lwtears.com/blog/four-reasons-writing-instruction-needs-makeover-and-evidence-based-solutions>
- Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K. & Spurling, J. (2016). Predictors of spelling and writing skills in first- and second-language learners. *Reading and Writing*, 29(1), 69–89. <https://doi.org/10.1007/s11145-015-9580-1>
- Harsch, C., Neumann, A., Lehmann, R. & Schröder, K. (2007). Schreibfähigkeit. In E. Klieme & B. Beck (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messungen*. (S. 42–62). Beltz. <https://doi.org/10.25656/01:3232>
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Randall (Hrsg.), *The science of writing: Theories, methods, individual differences, and applications* (S. 1–27). Erlbaum.
- 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 (Hrsg.), *Traditions of Writing Research* (S. 166–180). Taylor & Francis/Routledge.
- Hayes, J. R. & Flower, L. S. (1980). Identifying the Organization of Writing Processes. In L. Gregg & E. R. Steinberg (Hrsg.), *Cognitive Processes in Writing* (S. 3–30). Routledge. <https://doi.org/10.4324/9781315630274-3>

- Hayes, J. R. & Flower, L. S. (1986). Writing research and the writer. *American Psychologist*, 41(10), 1106–1113. <https://doi.org/10.1037/0003-066X.41.10.1106>
- Hennes, A.-K. (2020). *Schreibprodukte bewerten: die Rolle der Expertise bei der Bewertung der Textproduktionskompetenz* [Dissertation, Universität zu Köln]. KUPS.  
<http://kups.ub.uni-koeln.de/id/eprint/11414>
- 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 Experts Teachers Can Learn from or Do We Face a More Fundamental Problem? *Psychological Test and Assessment Modeling*, 64(3), 272–303.
- 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.
- Hess, C. W., Haug, H. T. & Landry, R. G. (1989). The reliability of type-token ratios for the oral language of school age children. *Journal of speech and hearing research*, 32(3), 536–540. <https://doi.org/10.1044/jshr.3203.536>
- Hiebert, E. H. & Cervetti, G. N. (2012). What Differences in Narrative and Informational Texts Mean for the Learning and Instruction of Vocabulary. *Vocabulary Instruction: Research to Practice*, 2, 322–344. <https://eric.ed.gov/?id=ed518047>
- Hooper, S., Costa, L., Fernandez, E., Barker, A., Valdes, C., Catlett, S. & Green, M. (2021). executive functions and writing skills in children and adolescents. In T. Limpo & T. Olive (Hrsg.), *Executive Functions and Writing* (S. 139–159). Oxford University Press.
- Hughes, D. C., Keeling, B. & Tuck, B. F. (1980). The Influence of Context Position and Scoring Method on Essay Scoring. *Journal of Educational Measurement*, 17(2), 131–134. <https://doi.org/10.1111/j.1745-3984.1980.tb00821.x>

- 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. & Karadeniz, I. (2016, 3. August). *An alternative approach to assessing achievement* [Vortrag]. Proceedings of the 2016 40th Conference of the International Group for the Psychology of Mathematics Education, Szeged, Hungary.
- 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>
- Jost, J. (2022). Schreibaufgaben zur Indikation von Schreibkompetenz. In M. Becker-Mrotzek & J. Grabowski (Hrsg.), *Schreibkompetenz in der Sekundarstufe: Theorie, Diagnose und Förderung* (S. 117–132). Waxmann.
- Juel, C., Griffith, P. L. & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78(4), 243–255. <https://doi.org/10.1037/0022-0663.78.4.243>
- Kent, S. C. & Wanzek, 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, G. Y.-S., Al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L. & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences*, 21(5), 517–525. <https://doi.org/10.1016/j.lindif.2011.06.004>
- Kim, 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>
- Kim, G. Y.-S., Otaiba, S. A., Sidler, J. F. & Grulich, L. (2013). Language, literacy, attentional behaviors, and instructional quality predictors of written composition for first graders. *Early Childhood Research Quarterly*, 28(3), 461–469. <https://doi.org/10.1016/j.ecresq.2013.01.001>
- Kim, G. Y.-S. & Park, S.-H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing*, 32(5), 1319–1343. <https://doi.org/10.1007/s11145-018-9913-y>
- Kim, G. Y.-S. & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, 109(1), 35–50. <https://doi.org/10.1037/edu0000129>
- Kim, G. Y.-S., Schatschneider, C., Wanzek, J., Gatlin, B. & Al Otaiba, S. (2017). Writing Evaluation: Rater and Task Effects on the Reliability of Writing Scores for Children in Grades 3 and 4. *Reading and Writing*, 30(6), 1287–1310. <https://doi.org/10.1007/s11145-017-9724-6>
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* (2. Aufl.). *Bildungsreform*. <https://doi.org/10.25656/01:20901>
- Klotz, P. (2022). Grammatik und Textualität. In H. Gornik (Hrsg.), *Deutschunterricht in Theorie und Praxis Series. Sprachreflexion und Grammatikunterricht* (2. überarbeitete und erweiterte Aufl., S. 138–154). W. Verlag Bertelsmann GmbH & Co. KG.

- 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>
- Kreutz, R. M., Philippek, J., Hennes, A.-K. & Schabmann, A. (2025). Good text quality despite deficient transcription skills - An analysis of double dissociation. *Journal of Writing Research*, 300–321.
- Kultusministerkonferenz (Hrsg.). (2022). *Bildungsstandards für das Fach Deutsch Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA): Beschluss der Kultusministerkonferenz vom 15.10.2004 und vom 04.12.2003, i.d.F. vom 23.06.2022*.  
[https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2022/2022\\_06\\_23-Bista-ESA-MSA-Deutsch.pdf](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 (Hrsg.), *Innovative Practices for Higher Education Assessment and Measurement* (S. 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>
- Limpo, T., Parente, N. & Alves, R. A. (2018). Promoting handwriting fluency in fifth graders with slow handwriting: a single-subject design study. *Reading and Writing*, 31(6), 1343–1366. <https://doi.org/10.1007/s11145-017-9814-5>
- Lindauer, N. (2024). Effektivität und Effizienz von holistischen Benchmarkratings. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Sprachliche Bildung: Bd. 11. Texte schreiben in allen Unterrichtsfächern: Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 91–112). Waxmann.
- 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 (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 20–35). Beltz Juventa.
- MacArthur, C. A., Graham, S. & Fitzgerald, J. (Hrsg.). (2016). *Handbook of writing research* (2nd edition). Guilford Publications Inc.  
<https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=4000626>
- 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>
- Magalhães, S., Mesquita, A., Filipe, M., Veloso, A., Castro, S. L. & Limpo, T. (2020). Spelling Performance of Portuguese Children: Comparison Between Grade Level,

- Misspelling Type, and Assessment Task. *Frontiers in Psychology*, 11, 1–9.  
<https://doi.org/10.3389/fpsyg.2020.00547>
- Mason, J. M. & Allen, J. (1986). A Review of Emergent Literacy with Implications for Research and Practice in Reading. In E. Rothkranz (Hrsg.), *Review of research in education* (S. 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.
- Mathiebe, M., Grabowski, J. & Becker-Mrotzek, M. (2024). Reliabilität bei der Beurteilung von Textqualität. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Sprachliche Bildung: Bd. 11. Texte schreiben in allen Unterrichtsfächern: Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 253–276). Waxmann.
- 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, D., Crossley, S. A. & McCarthy, P. M. (2010). Linguistic Features of Writing Quality. *Written Communication*, 27(1), 57–86.  
<https://doi.org/10.1177/0741088309351547>
- McNamara, T., Knoch, U. & Fan, J. (2019). *Fairness, Justice and Language Assessment*. Oxford University Press.
- 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>
- Merz-Grötsch, J. (2022). *Texte schreiben lernen: Grundlagen, Methoden, Unterrichtsvorschläge* (5. Aufl.). *Praxis Deutsch*. Klett | Kallmeyer.
- 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>
- 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>
- MSB NRW (Hrsg.). (2019). *Schule in NRW. Kernlehrplan für Sekundarstufe I Gymnasium in Nordrhein-Westfalen*. <https://www.schulentwicklung.nrw.de/lehrplaene/>
- 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 Assessment of Educational Progress. (2011). *Writing Framework for the 2011 National Assessment of Educational Progress*. U.S. Department of Education.

- National Commission on Writing. (2003). *The neglected "r": The need for a writing revolution*. College Entrance Examination Board.
- Neumann, A. (2007). *Briefe schreiben in Klasse 9 und 11.: Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen* [Dissertation, Universität Hamburg].
- Neumann, A. (2017). Zugänge zur Bestimmung von Textqualität. In M. Becker-Mrotzek, J. Grabowski & T. Steinhoff (Hrsg.), *Forschungshandbuch empirische Schreibdidaktik* (S. 203–219). Waxmann.
- Neumann, A. & Lehmann, R. (2008). Schreiben Deutsch. In E. Klieme (Hrsg.), *Beltz Pädagogik. Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (S. 89–103). Beltz.
- Nottbusch, G. (2017). Graphomotorik. In M. Becker-Mrotzek, J. Grabowski & T. Steinhoff (Hrsg.), *Forschungshandbuch empirische Schreibdidaktik* (S. 125–138). Waxmann.
- 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 (Hrsg.), *Cognitive psychology. Writing: A mosaic of new perspectives* (S. 125–140). Psychology Press.
- 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>
- Olive, T. (2021). Exekutive Functions in Skilled Writers. In T. Olive & T. Limpo (Hrsg.), *Executive Functions and Writing* (S. 207–226). Oxford University Press.

- Olive, T. & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory & cognition*, 30(4), 594–600.  
<https://doi.org/10.3758/BF03194960>
- 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>
- Peschel, C. (2024). Kann man das Beurteilen von (Schüler-)Texten lernen? Einblicke in diagnostisches Vorgehen von Lehrkräften. In I. Petersen, R. Reble & J. Kilian (Hrsg.), *Sprachliche Bildung: Bd. 11. Texte schreiben in allen Unterrichtsfächern: Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 39–60). Waxmann.
- Philippek, J., Kreutz, R. M., Hennes, A.-K., Schmidt, B. M. & Schabmann, A. (2025). The contributions of executive functions, transcription skills and text-specific skills to text quality in narratives. *Reading and Writing*(38), 651–670.  
<https://doi.org/10.1007/s11145-024-10528-5>
- Plakans, L. & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217–230. <https://doi.org/10.1016/j.jslw.2013.02.003>
- 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/>.

- Reichardt, A. (2015). *Rechtschreibung im Textraum - Modellierung der Schreibkompetenz in der Grundschule. Kölner Beiträge zur Sprachdidaktik Reihe A: Bd. 9.* Gilles & Francke.
- Risel, H. (2011). *Arbeitsbuch Rechtschreibdidaktik.* Schneider-Verlag Hohengehren.
- 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>
- RStudioTeam. (2020). *RStudio: integrated development for R [Computer software]*. RStudio, PBC. <http://www.rstudio.com>
- Ruffini, C., Osmani, F., Martini, C., Giera, W.-K. & Pecini, C. (2024). The Relationship Between Executive Functions and Writing in Children: A Systematic Review. *Child neuropsychology: a journal on normal and abnormal development in childhood and adolescence*, 30(1), 105–163. <https://doi.org/10.1080/09297049.2023.2170998>
- 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>
- 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>
- 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>
- Scardamalia, M. & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Hrsg.), *Advances in applied psycholinguistics:*



- Vol.2. Reading, writing, and language learning* (S. 142–175). Cambridge University Press.
- Schipolowski, S. & Böhme, K. (2016). Assessment of writing ability in secondary education: comparison of analytic and holistic scoring systems for use in large-scale assessments. *L1 Educational Studies in Language and Literature*, 16, 1–22.  
<https://doi.org/10.17239/L1ESLL-2016.16.01.03>
- Schmitt, M. & Knopp, M. (2017). Prädiktatoren der Schreibkompetenz. In M. Becker-Mrotzek, J. Grabowski & T. Steinhoff (Hrsg.), *Forschungshandbuch empirische Schreibdidaktik* (S. 239–252). Waxmann.
- Schmölzer-Eibinger, S. & Thürmann, E. (Hrsg.). (2015). *Fachdidaktische Forschungen: Bd. 8. Schreiben als Medium des Lernens: Kompetenzentwicklung durch Schreiben im Fachunterricht*. Waxmann.
- Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. In H. Blühdorn, E. Breindl & H. Waßner (Hrsg.), *Text - Verstehen. Grammatik und darüber hinaus* (S. 222–238). de Gruyter.
- Schoonen, R. (2012). The Validity and Generalizability of Writing Scores: The Effect of Rater, Task and Language. In E. van Steendam (Hrsg.), *Studies in writing: volume 27. Measuring writing: Recent insights into theory, methodology and practices* (S. 1–22). Brill.
- Schulte-Körne, G. (2022). Lese- und/oder Rechtschreibstörung im Kindes- und Jugendalter. In J. Fegert (Hrsg.), *Psychiatrie und Psychotherapie des Kindes- und Jugendalters* (S. 1–10). Springer Reference Medizin.
- Schwarz, M. (2001). Establishing Coherence in Text. Conceptual Continuity and Textworld Models. *Logos and Language*, 2(1), 15–24.

- Schweizer, V. & Wachter-Müller, S. (2017). *Neurotraining: Therapeutische Arbeit im kognitiven Bereich mit hirngeschädigten Erwachsenen* (5. Aufl.). Springer.
- Scott, C. M. & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of speech, language, and hearing research*, 43(2), 324–339. <https://doi.org/10.1044/jslhr.4302.324>
- Shallice, T. (1979). Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology*, 1(3), 183–211. <https://doi.org/10.1080/01688637908414450>
- Skar, G. B. & Jølle, L. J. (2017). Teachers as raters: Investigation of a long term writing assessment program. *L1 Educational Studies in Language and Literature*, 17, 1–30. <https://doi.org/10.17239/L1ESLL-2017.17.01.06>
- 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>
- Stephany, S. (2019). *Schreibflüssigkeit*. Mercator Institut für Sprachförderung und Deutsch als Zweitsprache. Basiswissen. [https://mercator-institut.uni-koeln.de/sites/mercator/user\\_upload/PDF/05\\_Publikationen\\_und\\_Material/201124\\_Basiswissen\\_Schreibfluessigkeit.pdf](https://mercator-institut.uni-koeln.de/sites/mercator/user_upload/PDF/05_Publikationen_und_Material/201124_Basiswissen_Schreibfluessigkeit.pdf)
- Storrer, A. (2004). Kohärenz in Hypertexten. *Zeitschrift für germanistische Linguistik*, 31(2). <https://doi.org/10.1515/zfgl.2004.006>
- Storrer, A. (2020). Textqualität digital: Ein Modell zur Qualitätsbewertung digitaler Texte. *Deutsche Sprache*(2), 101–125. <https://doi.org/10.37307/j.1868-775X.2020.02.03>
- Sturm, A. (2017). Förderung hierarchieniedriger Schreibprozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 266–284). Beltz Juventa.

- Sturm, A. (2018). *Empfehlungen zur Sprachförderung im Pilotprojekt ALLE*. Pädagogische Hochschule FHNW.
- Sturm, A. (2024). Hierarchieniedrige Schreibfähigkeiten und Schreibflüssigkeit: Ein mehrdeutiges Konstrukt. *Lernen und Lernstörungen*, 13(2), 63–74.  
<https://doi.org/10.1024/2235-0977/a000436>
- Sturm, A., Lindauer, N. & Sommer, T. (2018). Rückmelden. Aufgaben- und lernzielbezogenes Feedback. *Der Deutschunterricht*, 70(3), 80–91.
- Sturm, A., Nänny, R. & Wyss, S. (2017). Entwicklung hierarchieniedriger Schreibprozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 84–104). Beltz Juventa.
- Swanson, H. L. & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of experimental child psychology*, 63(2), 358–385.  
<https://doi.org/10.1006/jecp.1996.0054>
- Teuber, H. L. (1955). Physiological psychology. *Annual review of psychology*, 6, 267–296.  
<https://doi.org/10.1146/annurev.ps.06.020155.001411>
- Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review*, 34(4), 273–286.
- 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. <https://doi.org/10.3389/feduc.2022.1100095>

- van den Bergh, H., de Maeyer, S., van Weijen, D. & Tillema, M. (2012). Generalizability of Text Quality Scores. In E. van Steendam, M. Tillema, G. Rijlaarsdam & H. van den Bergh (Hrsg.), *Measuring Writing: Recent Insights Into Theory, Methodology and Practices* (S. 23–32). Brill.
- 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>
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge University Press.
- Weinert, F. (2001). *Leistungsmessungen in Schulen. Pädagogik*. Beltz Verlag.
- Weston, J. L., Crossley, S. A., McCarthy, P. M. & McNamara, D. (2011, 11. März). *Number of words versus number ideas: Finding a better predictor of writing quality* [Vortrag]. Twenty-Fourth International FLAIRS Conference, Florida, USA.  
<https://cdn.aaai.org/ocs/2618/2618-11356-1-pb.pdf>
- Wolf, J., Kreutz, R. M., Hennes, A.-K. & Schabmann, A. (2025). Relevance of Executive Functions for Writing Competence in Skilled Writers. *Journal of Writing Research*, 300–321.
- Young, M. P., Hilgetag, C. C. & Scannell, J. W. (2000). On imputing function to structure from the behavioural effects of brain lesions. *Philosophical transactions of the Royal Society of London*, 355, 147–161. <https://doi.org/10.1098/rstb.2000.0555>
- Zifonun, G. (2000). Textkonstitutive Funktionen von Tempus, Modus und Genus Verbi. In K. Brinker (Hrsg.), *Text- & Gesprächslinguistik* (S. 315–330). Walter de Gruyter.  
<https://doi.org/10.1515/9783110194067-035>