

**Empirische Sonderpädagogik**, 2017, Nr. 4, S. 365-385  
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

## Die Bedeutung der Schreibflüssigkeit, des mündlichen Sprachschatzes und der orthografischen Fertigkeiten für die Aufsatzqualität bei Grundschulkindern

*Matthias Grünke, Turid Knaak & Alla Sawatzky*

*Universität zu Köln*

### Zusammenfassung

Die Schwierigkeiten, welche Kinder und Jugendliche mit dem Förderschwerpunkt Lernen aufweisen, äußern sich besonders häufig beim Abfassen von Texten. Doch trotz der weiten Verbreitung dieser Problematik erhalten die betroffenen Schülerinnen und Schüler im Unterricht oft nur wenig spezielle Unterstützung. Viele Lehrkräfte fühlen sich der Aufgabe nur unzureichend gewachsen, Kompetenzen im Verschriftlichen von Gedanken und Ideen zu vermitteln. Der damit verbundene kognitive Prozess wird als äußerst komplex angesehen. Unter Pädagoginnen und Pädagogen herrscht eine große Unsicherheit im Hinblick darauf vor, an welchen Stellen am besten anzusetzen ist, um Rückstände in der Fähigkeit zum Verfassen von Texten aufzuholen. In der vorliegenden Studie steht die Frage im Fokus, welchen Stellenwert die Schreibflüssigkeit, die Breite des mündlichen Wortschatzes und die Rechtschreibkompetenz für die Qualität von Geschichten besitzen, die von Kindern in dritten und vierten Klassen verfasst werden. Die Befunde deuten darauf hin, dass der Textmenge bzw. der Schreibflüssigkeit in diesem Kontext die größte Bedeutung zukommt. Auch die Vielfalt des zur Verfügung stehenden mündlichen Wortschatzes und die orthographischen Fähigkeiten spielen eine Rolle, allerdings in geringerem Ausmaß. Diese Erkenntnisse liefern wertvolle Hinweise bezüglich darauf, an welchen relativ leicht zu fördernden Kompetenzen im Schulunterricht anzusetzen ist, um einem sonderpädagogischen Förderbedarf im Bereich des expressiven Schreibens frühzeitig vorzubeugen bzw. ihn adäquate zu behandeln.

Schlüsselwörter: Entwicklungsstörung des expressiven Schreibens, Förderschwerpunkt Lernen, Textqualität, Vorläuferkompetenzen, Prävention.

### **The impact of writing fluency, oral lexical diversity, and spelling skills on the quality of stories written by elementary school children**

#### **Abstract**

Difficulties that children and youth with special needs in the area of learning demonstrate, often express themselves in insufficient text composition skills. But even though this is a widely spread problem, students affected by it do oftentimes not receive specific support. Many teachers do not feel up to the task of instilling in students how to put their thoughts and ideas on paper. They view the corresponding cognitive processes as extremely intricate. Among educators, there is a deeply held uncertainty about where to start when trying to remedy deficits in text pro-

duction skills. This study focuses on the effect that writing fluency, oral lexical richness and spelling abilities of third and fourth graders have on the quality of texts that they produce. The findings suggest that the length of a text (or the writing fluency respectively) play the greatest role in this regard. However, to a lesser extent, oral lexical richness and orthographic abilities can also be considered as relevant. These insights provide valuable advice on which relatively easy to teach skills during regular classroom instruction one should focus on in order to timely prevent special needs in the area of expressive writing.

Key Words: Disorder of written expression, mild learning difficulties, text quality, prerequisite skills, prevention

## Einleitung

Sich schriftlich verständlich ausdrücken zu können, ist eine Schlüsselqualifikation und eine wichtige Kulturtechnik in unserer Gesellschaft (Fayol, Alamargot & Berninger, 2012). Durch die anwachsende Kommunikation über E-Mail, SMS oder Instant Messenger gewinnt diese Kompetenz im sozialen Bereich zusehends an Bedeutung (Karadag & Kayabasi, 2013). Außerdem ist sie grundlegende Bedingung für eine erfolgreiche Laufbahn in Schule und Berufswelt. Jugendliche ohne ausreichende Schreibkompetenz scheitern beim Bemühen um einen (höheren) Bildungsabschluss relativ häufig (Grünke & Leonard-Zabel, 2015). Auch das Finden eines Arbeitsplatzes gestaltet sich vergleichsweise schwierig (Kellogg & Rauerson, 2007).

Die Fähigkeit zum Verfassen von Texten entwickelt sich mit dem Alter und mit zunehmender Übung (Midgette, Haria & MacArthur, 2008). Der Prozess beginnt mit dem Erwerb basaler Vorläuferkompetenzen (z. B. phonologische Bewusstheit, graphomotorische Fertigkeiten, Kenntnisse über das alphabetische Prinzip, orthographisches Regelwissen, Kenntnisse über Wortschatz und Grammatik, konzeptuelles und prozedurales Wissen über Schriftsprache). Später kommen höherwertige Fähigkeiten hinzu, die sich in dem wegweisenden Modell von Hayes und Flower (1980) je einem von drei Bereichen zuordnen lassen: (1) Planen (Planning) (2) Verschriftlichen (Translating) und (3) Überarbeiten (Revising).

Kinder starten ihre ersten Schreibversuche in aller Regel im Alter von etwa sieben Jahren. Sie beginnen üblicherweise mit einfachen Erzählungen über Geschehnisse, die ihnen selbst widerfahren sind. In dieser Phase bringen sie jeden Gedanken zu Papier, der ihnen in den Sinn kommt. Ein vorheriges Planen findet kaum statt. Auch die (mögliche) Zielgruppe wird während dieses Prozesses so gut wie nicht ins Kalkül gezogen (Graham, 2010). Bis zum Alter von ca. zehn Jahren verfeinert sich die Fähigkeit zum Verfassen von Geschichten. Die Kinder können im Normalfall nun auch über Begebenheiten berichten, die ihnen nicht selbst passiert oder die frei erfunden sind. Im Entwicklungsabschnitt zwischen 10 und 14 Jahren finden die potenziellen Leserinnen und Leser beim Verfassen immer mehr Berücksichtigung. Mädchen und Jungen können nun nicht nur Geschichten, sondern auch informative Darstellungen und Berichte schreiben. Im weiteren Verlauf sind sie für gewöhnlich in der Lage, nach und nach alle Textformen zu produzieren (Becker-Mrotzek, 2014; Becker-Mrotzek & Böttcher, 2012).

Wenn es um die Beurteilung der Qualität von Texten im Allgemeinen und von Geschichten im Besonderen geht, so spielt in der einschlägigen Literatur ein Aspekt eine ganz besonders herausragende Rolle: die lexikalische Vielfalt. Weisen Schreibprodukte eine große Reichhaltigkeit an unterschiedlichen Wörtern auf, so gilt dies als Ausdruck einer hohen Sprachkompetenz (z. B. Zareva, Schwanenfugel & Nikolova, 2005) und führt zu positiven Bewertungen (z. B. Laufer

& Nation, 1995). Was die Erfassung der Güte von Texten angeht, so kann man im deutschsprachigen Raum zwar auf informelle Skalen wie den „Basiskatalog zur mehrdimensionalen kriterialen Textbewertung“ (Becker-Mrotzek & Böttcher, 2006) oder das „Zürcher Textanalyseraster“ (Nussbaumer, 2010) zurückgreifen. Standardisierte Tests liegen bislang jedoch noch nicht vor. Im angloamerikanischen Bereich ist dies anders. Hier existieren einige entsprechende Instrumentarien wie etwa das „International English Language Testing System“ (IELTS) (Loughheed, 2007), die „Michigan English Language Assessment Battery“ (MELAB) (Cambridge Michigan Language Assessments, 2015) oder der „Test of Written Language“ (TOWL-4) (Hammill & Larsen, 2009). Lexikalische Vielfalt ist bei allen Verfahren ein entscheidender Maßstab für die Bewertung. Exemplarisch sei an dieser Stelle auf das zentrale Kriterium für die Beurteilung der schriftlichen Teile der Prüfung im IELTS verwiesen, nämlich auf den „...range of vocabulary the candidate has used“ (Loughheed, 2007, S. 10). In der MELAB werden hohe Punktzahlen dann vergeben, wenn das folgende Kennzeichen vorliegt: „... a wide range of appropriately used vocabulary“ (Cambridge Michigan Language Assessments, 2015, S. 59).

Auch wenn es um die Förderung dessen geht, was in dem Modell von Hayes und Flower (1980) mit „Translating“ bezeichnet wird, steht die lexikalische Vielfalt häufig klar im Vordergrund. Mit „Translating“ ist in diesem Kontext gemeinhin das Überführen von Gedanken und Ideen in geschriebene Sprache gemeint (Fayol, Alamargot & Berninger, 2012). Die bei der Förderung der damit einhergehenden Fähigkeit am häufigsten zum Einsatz kommende Methode nennt sich „Sentence Combining“ (Satzkombinieren). Saddler (2013) beschreibt ihren Zweck wie folgt: „... sentence combining provides mindful practice in manipulating and rewriting basic phrases or clauses into more varied and syntactically mature forms“ (S. 240). Der Autor weist zu Recht

darauf hin, dass diese Form der Förderung die einzige ist, welche vor dem Hintergrund des hohen Stellenwerts der lexikalischen Vielfalt als Gütekriterium für gute Texte seit mehr als 50 Jahren durchgängig im Fokus der Schreibforschung steht.

Leider ist der Anteil an Kindern und Jugendlichen, der massive Rückstände in der Schriftsprachentwicklung aufweist und infolgedessen keine Erzählungen, Kommentare, Beschreibungen, Berichte, Reportagen usw. von akzeptabler Güte verfassen kann, relativ hoch. Gravierende Probleme im Sinne der Kategorie F81.8 („Entwicklungsstörung des expressiven Schreibens“) der ICD-10 (Dilling, Mombour & Schmidt, 2015) bzw. der Kategorie 315.2 („Disorder of Written Expression) des DSM-V (American Psychiatric Association, 2013) sind unter allen Lernschwierigkeiten am häufigsten vertreten. Die Prävalenzraten reichen hier bis zu 14% (Katusic, Colligan, Weaver & Barabasi, 2009).

Das kompetente Verfassen von qualitativ zufriedenstellenden Geschichten stellt für Kinder am Ende Ihrer Grundschulzeit im engeren Sinne den ersten Meilenstein für die Entwicklung einer hinreichend ausgeprägten Schreibfähigkeit dar. Um Störungen bzw. einen sonderpädagogischen Förderbedarf in dieser Hinsicht zu vermeiden, ist unbedingt sicherzustellen, dass Schülerinnen und Schüler an dieser Hürde nicht scheitern. Ein erfolgversprechendes pädagogisches Eingreifen ist jedoch nur dann möglich, wenn auf Seiten der Kinder entsprechende Vorläuferfertigkeiten in ausreichendem Maße vorhanden sind. In der einschlägigen Literatur werden in diesem Kontext drei Bedingungen besonders häufig genannt (siehe z. B. Fayol, Alamargot & Berninger, 2012; Graham, Collins & Rigby-Wills, 2017):

(1) *Hinreichende Schreibflüssigkeit*. Wie etwa Berninger und Winn (2006) oder Graham (2010) treffend feststellen, muss im Bereich der Handschrift eine sichere Automatisierung vorliegen, bevor man wohl formulierte Schreibprodukte erwarten kann. Das

Verfassen von Texten geschieht in aller Regel nicht gleichmäßig. Stattdessen wechseln sich Pausen mit Phasen ab, in denen es zu einem „Schreibschwall“ kommt. Es liegen einige Studien vor, in denen gezeigt werden konnte, dass Grundschul Kinder, die darin geübt sind, mit einem Stift Buchstaben, Wörter und Sätze zu Papier zu bringen, zeitlich ausgedehntere Schreibschwälle zeigen und insgesamt längere Texte produzieren als Mädchen und Jungen, die in dieser Hinsicht weniger versiert sind (z. B. Alves, Limpo, Fidalgo, Carvalhais, Pereira & Castro, 2016; Connelly, Dockrell, Walter & Critten, 2012).

(2) *Ausreichend breiter mündlicher Wortschatz.* Gemäß der Simple View of Reading von Gough und Tunmer (1986) stellt das Leseverständnis das Produkt aus Sprachverständnis und Worterkennung dar. Damit ist gemeint, dass ein Mensch neben ausreichenden Dekodierungsfähigkeiten auch über einen genügend großen Wortschatz verfügen muss, um einem Text Bedeutung beimessen zu können (vgl. Grünke & Strathmann, 2007). Analog ist es nötig, im Langzeitgedächtnis über ein reichhaltiges Vokabular zu verfügen, um in der Lage zu sein, das entsprechende Wissen beim Erstellen von Geschichten im Besonderen und von Schreibprodukten im Allgemeinen zu nutzen. Im klassischen Schreibmodell von Hayes und Flower (1980) spielt dieser dauerhafte Informationsspeicher eine prominente Rolle. Er enthält jegliches Wissen (z. B. über das Thema oder über die Adressatinnen und Adressaten), auf das eine Schreiberin bzw. ein Schreiber zurückgreifen muss, um einen anspruchsvollen Text verfassen zu können. Selbstredend zählt hierzu auch die mündliche lexikalische Vielfalt. Im kognitiven Schreibmodell von Hayes (1996) – es handelt sich hierbei um eine Weiterentwicklung des ursprünglichen Ansatzes von Hayes und Flower (1980) – rücken einige Aspekte noch expliziter ins Blickfeld. Zum einen misst Hayes dem Arbeitsgedächtnis eine große Bedeutung bei und zum anderen rücken hier individuelle Unterschiede ver-

schiedener Schreiber in den Blickpunkt. So spielen u.a. affektive Einflüsse, Veranlagungen, Überzeugungen und das soziale Umfeld des Schreibers bei der Textproduktion eine größere Rolle, als noch im ursprünglichen Modell. Mit der Einführung der Rubrik „Linguistic Knowledge“ im weiterentwickelten Modell betont der Autor zudem die zentrale Bedeutung von relevanten Kenntnissen über die Verwendung und die Bedeutung von Wörtern für das Verschriftlichen von Gedanken und Ideen. Je größer der Fundus einer Person in dieser Hinsicht ist, umso besser sind ihre Voraussetzungen, auch im Schriftlichen eine facettenreiche und mannigfaltige lexikalische Vielfalt an den Tag legen zu können.

(3) *Angemessene orthographische Fertigkeiten.* Auch hinreichend ausgeprägte Rechtschreibkompetenzen stellen offenkundig eine zentrale Voraussetzung für die Fähigkeit dar, abwechslungsreiche und als gut bewertete Geschichten zu verfassen. Bereits ein falsch geschriebenes Wort kann die Aussage eines ganzen Satzes verändern oder unverständlich machen. Schülerinnen und Schüler, die zu intensiv über die korrekte Schreibweise eines Wortes nachdenken müssen, sind nicht gut in der Lage, ihre Aufmerksamkeit auf die Qualität ihrer Texte zu richten (Graham, Harris & Chorzempa, 2002). Das Arbeitsgedächtnis ist in diesem Fall relativ stark belastet, so dass es nur noch geringe Kapazitäten für Planung, Erstellung und Überarbeitung eines Textes bereitstellen kann (Kellogg & Raulerson, 2007). Schülerinnen und Schüler, die Probleme in der Rechtschreibung aufweisen, sind oftmals nicht gut dazu in der Lage, ihre Ideen und Gedanken zu Papier bringen. Dies hält sie davon ab, Geschichten von akzeptabler Qualität zu produzieren (Graham, Collins & Rigby-Wills, 2017; Grünke & Leonard-Zabel, 2015). In Anbetracht des Umstandes, dass eine große lexikalische Vielfalt in aller Regel maßgeblich zu einer positiven Bewertung der Güte eines Textes beiträgt, verwundert dieser Zusammenhang nicht: Man hält sich beim freien Schreiben

eher an diejenigen Wörter, deren Orthographie man beherrscht. Ist die entsprechende Palette relativ schmal, dann schöpft man bei der Auswahl an Wörtern für das eigene Schreibprodukt auch aus einem vergleichsweise kleinen Fundus.

Trotz der praktischen Relevanz, die das Schreiben besitzt, gibt es im Vergleich zum Lesen und zu anderen schulischen Bereichen deutlich weniger Studien, die sich mit dieser Kompetenz beschäftigen (Grünke & Leonard-Zabel, 2015). Dies liegt u. a. daran, dass die Fähigkeit zur Erstellung von Texten in einem komplexen Prozess Ausdruck findet, dessen fachkundige Beurteilung zweifelsohne schwieriger ist als die Überprüfung der Leseflüssigkeit oder des arithmetischen Könnens (Rodríguez, Grünke, González-Castro, García, & Álvarez-García, 2015; Troia, 2010). Die Vielschichtigkeit der kognitiven Vorgänge bei der Textproduktion schrecken viele Lehrkräfte davon ab, sich der Vermittlung der hierfür notwendigen Kompetenzen zu widmen (Troia & Graham, 2003). Lügen ihnen jedoch fundierte Informationen darüber vor, welche Wege sie gehen könnten, um ihre Schülerinnen und Schüler beim Meistern der ersten Hürde während der Entwicklung zu einer versierten Schreiberin bzw. zu einem versierten Schreiber zu unterstützen (nämlich beim Verfassen von Geschichten), wäre es leichter möglich, das Aufkommen von Schreibstörungen zu vermeiden.

## Fragestellung

Wenn es darum geht, Vorläuferkompetenzen zu postulieren, die für die Entwicklung einer gut ausgeprägten lexikalischen Vielfalt im Schreiben als zentrales Merkmal für die Qualität von Geschichten maßgeblich sind, dann sind sicherlich viele Optionen denkbar. Mit Blick auf zentrale Schreibmodelle und einschlägige Untersuchungen (siehe oben), scheinen hierbei allerdings die drei hier thematisierten Variablen die größte Relevanz zu besitzen. In Veröffentlichungen,

in denen es um die Vorhersagekraft dieser Komponenten auf die Qualität von Texten geht, werden als Begründung – von ganz wenigen Ausnahmen abgesehen (Grünke, Büyüknarci, Wilbert & Breuer, 2015; Sumner, Connelly & Barnett, 2014) – in erster Linie logische Argumente und keine empirischen Belege angeführt. Eine Überprüfung der Zusammenhänge steht bislang noch weitgehend aus.

Das Ziel der vorliegenden Arbeit besteht somit darin, die Bedeutung der Schreibflüssigkeit, der mündlichen lexikalischen Vielfalt und der orthographischen Fertigkeiten für die schriftliche lexikalische Vielfalt, die Kinder am Ende ihrer Grundschulzeit beim Verfassen von Geschichten an den Tag legen, empirisch zu untersuchen. Es wird davon ausgegangen, dass alle drei genannten Variablen in bedeutsamer Weise die Varianz der schriftlichen lexikalischen Vielfalt von Kindern erklären.

## Methode

### Stichprobe

Für die vorliegende Studie wurden Daten an einer Grundschule aus einem sozialen Brennpunktviertel in einer niederrheinischen Großstadt erhoben. Die Kontaktaufnahme erfolgte telefonisch über die Schulleitung. Letztendlich konnten 59 Schülerinnen und Schüler in die Untersuchung einbezogen werden, von denen zum Zeitpunkt der Erhebung 14 eine dritte (24%) und 45 eine vierte Klasse (76%) besuchten. Von den Kindern waren 37 weiblich (63%) und 22 männlich (37%). Acht Schülerinnen und Schüler waren 8 Jahre, 29 waren 9 Jahre, 18 waren 10 Jahre und zwei waren 11 Jahre alt ( $M = 9.2$  Jahre,  $SD = 0.76$  Jahre,  $Md = 9$  Jahre,  $IQB = [9$  Jahre; 10 Jahre]). Laut Auskunft der Schulleitung stammte bei 30 Mädchen und Jungen mindestens ein Elternteil gebürtig nicht aus Deutschland (51%). Sechs Kinder wiesen einen türkischen (10%), drei einen iranischen (5%) und drei



einen marokkanischen (5%) Migrationshintergrund auf. Alle übrigen nicht in Deutschland geborenen Eltern stammten jeweils aus einer anderen Nation (31%).

### Vorgehen

Aus dem Buch von Hirmer und Hirmer (2007) wurden insgesamt fünf als geeignet erscheinende Themenstellungen für Geschichten ausgewählt („Ein Erlebnis auf dem Schulweg“, „Verschlafen“, „Glück im Unglück“, „Hatten wir einen Spaß“ und „Den Schlüssel vergessen“). Jedem Kind wurde per Zufall ein Thema zugewiesen. Es folgte eine randomisierte Festlegung im Hinblick darauf, ob eine jeweilige Schülerin bzw. ein jeweiliger Schüler zunächst eine Geschichte erzählen oder aufschreiben sollte. Dies hatte den Zweck, mögliche systematische Effekte zu kontrollieren, die auf die Reihenfolge hinsichtlich der Art der Aufgabenstellung zurückzuführen sind. Zwei Studierende der Sonderpädagogik (Master of Education) nahmen die Datenerhebungen vor. Sie wurden vorab im Rahmen zweier 45minütiger Treffen mit dem Erstautor ausgiebig in ihre Aufgaben eingewiesen. Die Studierenden führten die Kinder am Beginn eines Schultags einzeln in einen separaten Raum, stellten sich vor und versuchten, ihnen durch Small Talk jegliche Nervosität oder Unsicherheit in Anbetracht der ungewohnten Situation zu nehmen. Im Anschluss präsentierten sie den Mädchen bzw. Jungen die zugeloste Überschrift und ließen sie hierzu entweder eine Geschichte erzählen oder aufschreiben. Eine Zeitbegrenzung für ein vorheriges Überlegen und für die Ausführung der Aufgabe gab es nicht. Gegen Ende des Schultages wurden dieselben Kinder von den Studierenden abermals in einen separaten Raum geführt, um dort zum jeweils gleichen Thema eine Geschichte aufzuschreiben (falls sie bereits erzählt wurde) oder zu erzählen (falls sie bereits aufgeschrieben wurde). Auch in diesem Fall gab es keine Zeitbegrenzung. Die erzählten Begebenheiten wurden mit einem Diktiergerät

aufgenommen und im Anschluss wörtlich transkribiert. Hierbei schwankte die pro Kind investierte Zeit zwischen fünf und 20 Minuten. Die Datenerhebung erstreckte sich insgesamt über eine Schulwoche.

### Abhängige und unabhängige Variablen

**Abhängige Variable.** Als Indikator für die lexikalische Vielfalt als ein Aspekt der Güte von schriftlichen Erzählungen wurde im Einklang mit den entsprechenden Ausführungen in der Einleitung die Breite des verwendeten Wortschatzes herangezogen. Es liegen zahlreiche Möglichkeiten vor, um lexikalische Vielfalt zu bestimmen (z. B. Daller, van Hout & Treffers-Daller, 2003; Vermeer, 2000). Die einfachste stellt die Type-Token-Relation (TTR) dar (auch Diversifikationsquotient genannt). Hierbei wird ermittelt, welche verschiedenen Wörter in einem Text Verwendung finden. Diese sind im Anschluss durch die Anzahl der Gesamtwörter zu dividieren (McCarthy & Jarvis, 2010). Allerdings sind die Ergebnisse einer solchen Analyse mitunter schwer zu interpretieren, da der Quotient stark von der Länge des Textes abhängt. Es existieren mehrere Alternativen zur TTR, die den Anspruch erheben, diese Dependenz zu kontrollieren (Koizumi, 2012). Zu den bekanntesten zählt der Guiraud Index (GI; Guiraud, 1960; Vermeer, 2000). Der GI hat bereits in zahlreichen Untersuchungen mit Schülerinnen und Schülern Anwendung gefunden, u. a. in der viel beachteten Studie von Sumner, Connelly und Barnett (2014) zum Einfluss der Rechtschreibkompetenz auf die Wortschatzbreite. Er wird folgendermaßen ermittelt:

$$GI = \frac{v}{\sqrt{N}} .$$

Hierbei steht  $v$  für die Anzahl der verschiedenen Wörter (die jeweils nur einmal gezählt werden) und  $N$  für die Anzahl aller verwendeten Wörter. Für eine genaue Beschreibung der Auswertung – z. B. im Hin-

blick auf die Berücksichtigung verschiedener Tempora und Konjunktionen – sei an dieser Stelle auf die Ausführungen von Daller (2010) verwiesen. Anders als bei der TTR, bei der eine hohe Anzahl an Wörtern erfahrungsgemäß sehr häufig mit einem hohen Index für lexikalische Vielfalt einhergeht, ist dieser Zusammenhang beim GI deutlich geringer (ebd.). Darüber hinaus hat er sich in Studien als geeignetes Kriterium für die Beurteilung der Qualität von Aufsätzen erwiesen (z. B. Mellor, 2011). Angesichts der Vorteile der Verwendung des GI gegenüber der TTR bei gleichem Erhebungsaufwand wurde als Indikator für lexikalische Vielfalt der erste Index verwendet.

**Unabhängige Variablen.** Schreibflüssigkeit wird in aller Regel über die Textlänge erfasst (Hosp, Hosp & Howell, 2016). Zuweilen bezieht man sowohl die Menge der geschriebenen Wörter als auch die hierfür benötigte Zeit mit ein (etwa bei der Composing Rate, Latif, 2013). Allerdings lassen sich Phasen des Planens, Verschriftlichens und Überarbeitens aus dem Modell von Hayes und Flower (1980) nur schwer voneinander trennen. Denkt ein Kind über den weiteren Verlauf einer zu verfassenden Erzählung immer wieder kurz nach, so verlängert dies die Zeit und verringert die Composition Rate (obschon diese Verzögerung nichts mit der Fähigkeit zu tun hat, flüssig zu schreiben, siehe z. B. Polio & Friedman, 2017). In der vorliegenden Studie wurde daher die Anzahl der geschriebenen Wörter (AW (S)) als Indikator für die Schreibflüssigkeit herangezogen. Es ist davon auszugehen, dass Kinder, denen die Mechanik des Verschriftlichens vergleichsweise schwer fällt, relativ schnell ermüden und eher kurze Texte formulieren (unabhängig von Zeitvorgaben oder Zeiterfassungen) (Graham, 2010).

Die mündliche lexikalische Vielfalt, wurde auf der Basis der transkribierten Erzählungen mit Hilfe des oben beschriebenen GI operationalisiert, da die Vorteile dieses Indikators auch bei der Erfassung der mündlichen lexikalischen Vielfalt gelten. Im Fol-

genden wird die Abkürzung „GI (S)“ für den GI beim Schreiben und die Abkürzung „GI (E)“ für den GI beim Erzählen verwendet.

Zur Ermittlung der Rechtschreibkompetenz (bzw. des Ausmaßes von Schwierigkeiten in diesem Bereich) wurde der Fehlerquotient (FQ) herangezogen. Dieser errechnet sich wie folgt:

$$FQ = \frac{NFehler \times 100}{NWörter}$$

Der FQ ist ein gängiges Maß zur Erfassung orthographischer Probleme und findet beispielsweise im Rahmen der Bewertung von Abiturprüfungen in Hessen Anwendung (Kultusministerium Hessen, 2016). Berücksichtigt werden bei der Ermittlung des FQ alle orthographischen Fehler.

**Kontrollvariablen und sonstige Variablen.** Da das Geschlecht und das Alter sich häufig als Kovariaten von sprachlichen Variablen zeigen (Morrow, 2014), wurden diese als Kontrollvariablen mit in die multiplen Regressionsanalysen einbezogen. Die Anzahl erzählter Wörter (AW (E)) als Grundlage für die Ermittlung des GI (E) wurde ebenfalls erhoben und ist in Tabelle 1 und Tabelle 2 mitaufgeführt.

**Ermittlung der sprachlichen Variablen.** Die gleichen Studierenden, welche die Erhebung durchführten, ermittelten unabhängig voneinander für alle Schülerinnen und Schüler den GI (S), die AW (S), (auf Basis der Transkriptionen) den GI (E), die AW (E) sowie (unter Rückgriff auf die Kategorien der Oldenburger Fehleranalyse für die Klassen 3-9 nach Thomé & Thomé, 2016) den FQ. Im Anschluss wurden die Ergebnisse miteinander verglichen. In jedem Fall eines diskrepanten Urteils erfolgte eine kurze Diskussion mit dem Ziel, einen Konsens zu erreichen. Die Bewertungen stimmten bei GI (S), GI (E) und FQ zu über 95% überein. Im Falle von AW (S) und AW (E) konnte sogar eine vollständige Korrespondenz erreicht werden. Der Abgleich und Austausch bei unterschiedlichen Beurteilungen führte in allen Fällen zu schnellen einvernehmlichen Einigungen.

## Ergebnisse

Die Mittelwerte, Standardabweichungen, Varianzen, Variationskoeffizienten, Varianzverhältnisse und standardisierten Mittelwertsdifferenzen finden sich jeweils nach den beiden Gruppen (1. erst Erzählen, dann Schreiben und 2. erst Schreiben, dann Erzählen) getrennt in Tabelle 1.

Es fällt auf, dass sich die Streuungen für die Variablen AW (E) und FQ in den beiden Gruppen stark unterscheiden. Nach Cohen (1988, zitiert nach Eid, Gollwitzer & Schmitt, 2015, S. 349) handelt es sich ab einem Varianzverhältnis von  $v \geq 2$  um hohe Effekte. In der Gruppe, in der erst erzählt und danach geschrieben wurde, ist die Unterschiedlichkeit in Bezug auf die Anzahl der Wörter (erzählt und geschrieben) mehr als doppelt bzw. mehr als viermal so groß wie in der Gruppe, in der erst geschrieben und danach erzählt wurde. Umgekehrt verhält es sich mit dem Streuungsunterschied im FQ: Hier ist die Varianz in der Gruppe, in der erst geschrieben und danach erzählt wurde, annähernd dreimal so groß wie in der Gruppe, in der erst erzählt und dann geschrieben wurde. Entsprechend fällt der Variationskoeffizient – also die Größe der Standardabweichung in Prozenten des Durchschnitts ( $VK = 100\% \cdot SD/M$ ) – über beide Gruppen hinweg mit  $VK = 91\%$  für die Anzahl der geschriebenen Wörter und  $VK = 90\%$  für den Fehlerquotienten recht hoch aus. Die Mittelwertdifferenzen für die beiden Versuchsbedingungen übersteigen in keiner der Variablen einen kleinen Effekt ( $d \leq |0.16|$ , Cohen, 1988, zitiert nach Eid, Gollwitzer & Schmitt, 2015, S. 338) und alle  $p$ -Werte der Gruppenvergleiche liegen über der Signifikanzgrenze von 5%. Auch Gegenüberstellungen einer robusten Statistik der zentralen Tendenz (Mann-Whitney-Prüfung der Ränge) ergeben keine statistisch bedeutsamen Unterschiede. Schließlich übt auch die zufällig getroffene Themenauswahl keinen signifikanten Einfluss auf AW (E), AW (S), FQ, GI (E) oder GI (S) aus.

Tabelle 1: Deskriptive Maße der Stichprobe

	E-S (n = 32)				S-E (n = 27)				V*	d <sup>+</sup>
	M	SD	s <sup>2</sup>	VK%	M	SD	s <sup>2</sup>	VK%		
Alter	9.19	0.78	0.61		9.22	0.75	0.56		1.09 <sup>a</sup>	-0.05 <sup>a</sup>
AW (E)	117.59	92.86	8623.35	79%	112.56	60.18	3622.03	53%	2.38 <sup>c</sup>	0.06 <sup>a</sup>
AW (S)	67.97	74.01	5477.39	109%	62.26	35.95	1292.05	58%	4.24 <sup>c</sup>	0.10 <sup>a</sup>
FQ	15.75	11.13	123.96	71%	17.27	18.37	337.37	106%	2.72 <sup>c</sup>	-0.10 <sup>a</sup>
GI (E)	5.62	1.13	1.29	20%	5.78	1.24	1.53	21%	1.19 <sup>a</sup>	-0.14 <sup>a</sup>
GI (S)	5.03	1.145	2.10	29%	5.25	1.27	1.61	24%	1.30 <sup>b</sup>	-0.16 <sup>a</sup>

Anmerkungen. N = 57. Gruppe E-S (erst Erzählen, dann Schreiben), Gruppe S-E (erst Schreiben, dann Erzählen), V: Varianzverhältnis, AW (E): Anzahl Wörter erzählt, AW (S): Anzahl Wörter geschrieben, FQ: Fehlerquotient, GI (E): Guiraud Index für Erzählen, GI (S): Guiraud Index für Schreiben, VK%: Variationskoeffizient in %; \* alle  $p > .15$ , <sup>a</sup> alle  $p > .50$ , <sup>b</sup> kleiner bis mittlerer Effekt, <sup>c</sup> großer Effekt



Tabelle 2

	Geschlecht <sup>a</sup>	Alter	AW (E)	AW (S)	FQ	GI (E)
Alter	.07					
AW (E)	.04	.36**				
AW (S)	.26*	.40**	.53***			
FQ	-.06	-.45***	-.23	-.25		
GI(E)	-.01	.23	.65***	.54***	-.25	
GI(S)	.27*	.38**	.57***	.84***	-.37**	.68***

Anmerkungen. N = 59, AW (E): Anzahl Wörter erzählt, AW (S): Anzahl Wörter geschrieben, FQ: Fehlerquotient, GI (E): Guiraud Index für Erzählen, GI (S): Guiraud Index für Schreiben, <sup>a</sup> mit 0 = weiblich und 1 = männlich, \*p < .05, \*\*p < .01, \*\*\*p < .001.

Tabelle 2 zeigt eine Übersicht der Korrelationen zwischen den erhobenen Variablen. Die Beziehung zwischen AW (S) und GI (S) kann als sehr eng bezeichnet werden ( $r(57) = .84$ ). Da beim GI die Anzahl der Wörter in die Berechnung einfließt, könnte es sich bei der sehr hohen Korrelation um einen redundanten Zusammenhang handeln. Aus diesem Grund wurde die Korrelation zwischen AW (S) und GI (S) mit der Korrelation zwischen AW (E) und GI (E) hinsichtlich der Höhe verglichen. Falls die Korrelation von  $r_{AW(S), GI(E)}(57) = .84$  in hohem Maße der Tatsache geschuldet sein sollte, dass bei der Ermittlung von GI (S) die Anzahl der geschriebenen Wörter berücksichtigt wird, dann dürfte sich die Enge der Beziehung zwischen GI (E) und AW (E) von der Höhe des zuvor genannten Zusammenhangs nicht nennenswert unterscheiden. Der z-Test für den Vergleich zweier Korrelationen  $r_{ab}$  und  $r_{cd}$  aus einer Stichprobe ergab mit  $z = -2.63$ ,  $p = .009$  (zweiseitig) einen signifikanten Unterschied (zur Berechnungsvorschrift siehe Bortz & Schuster, 2010). Zudem resultierte die Gegenüberstellung der Fisher Z-Werte der beiden Korrelationen Fisher  $Z(r_{AW(S), GI(S)}) - \text{Fisher } Z(r_{AW(E), GI(E)})$  in einer Differenz von  $Z = -.446$ . Dieser Wert entspricht rücktransformiert mit  $r = -.42$  einer mittleren Effektgröße (Lenhard & Lenhard, 2014) und dient als weiterer Beleg dafür, dass sich die beiden Korrelationen  $r_{AW(S), GI(S)}$  und  $r_{AW(E), GI(E)}$  deutlich unterscheiden. Abgesehen

davon ist an dieser Stelle zu erwähnen, dass die Korrelation zwischen AW (S) und GI (S) mit  $r = .34$  in der Untersuchung von Sumner, Connelly und Barnett (2014) deutlich geringer ausfiel als im vorliegenden Fall und damit lediglich 11.56% der gemeinsamen Varianz aufklärte. Auf der Basis dieser Ergebnisse ist daher anzunehmen, dass der Zusammenhang zwischen AW (S) und dem mit Hilfe dieser Anzahl bestimmten GI (S) zumindest nicht ausschließlich redundanter Natur ist und beide Variablen getrennt voneinander in der statistischen Analyse verwendet werden können.

Um die zentrale Forschungsfrage der vorliegenden Studie nach der Bedeutung von AW (S), FQ und GI (E) für GI (S) zu beantworten, wurde eine multiple Regression in Schritten durchgeführt. Mit diesem Verfahren lässt sich der Beitrag einer jeden einzelnen unabhängigen Variable für die Varianzaufklärung im Kriterium über den Beitrag anderer unabhängigen Variablen hinaus bestimmen. In Tabelle 3 werden jeweils der statistische Einfluss in bivariater Form ( $r$  oder das standardisierte Regressionsgewicht für die einfache lineare Regressionsgleichung) und zum Vergleich der statistische Einfluss in multivariater Form (standardisiertes Regressionsgewicht im jeweiligen multiplen Modell, genannt „Schritt“) sowie die standardisierten Regressionsgewichte in den unterschiedlichen Modellen für die Prädiktoren und die Kontrollvariablen dargestellt.

Tabelle 3

Variablen	GI (S)					$R^2$	$\Delta R^2$
	$r$ bzw. $\beta^1$	$b$	$SE\ b$	$\beta$			
1. Schritt						.18**	
Alter	.38***	.65	.21	.36**			
Geschlecht <sup>a</sup>	-.27*	-.69	.33	-.25*			
2. Schritt						.70***	.51***
Alter	.38***	.11	.14	.06			
Geschlecht	-.27*	-.17	.21	-.06			
AW (S)	.84***	.02	.00	.80***			
3. Schritt						.72***	.02*
Alter	.38***	-.02	.15	-.01			
Geschlecht	-.27*	-.17	.20	-.06			
AW (S)	.84***	.02	.00	.79***			
FQ	-.37**	-.02	.01	-.17*			
4. Schritt						.79***	.07***
Alter	.38***	.01	.13	.00			
Geschlecht	-.27*	-.31	.18	-.11 <sup>+</sup>			
AW (S)	.84***	.01	.00	.60***			
FQ	-.37**	-.01	.01	-.13 <sup>+</sup>			
GI (E)	.68***	.37	.09	.32***			

Anmerkungen.  $N = 59$ ; GI (S): Guiraud Index für Schreiben, AW (S): Anzahl Wörter geschrieben, FQ: Fehlerquotient, GI (E): Guiraud Index für Erzählen, \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ , +  $.05 < p < .10$ , <sup>a</sup> mit 0 = weiblich und 1 = männlich, <sup>1</sup> dargestellt ist  $r$  (Nullkorrelation, jeweils immer mit  $df = 57$ ) bzw.  $\beta$  für den bivariaten Regressionsfall.

Zunächst verdeutlicht Tabelle 3, dass der Einfluss der Kontrollvariablen „Alter“ und „Geschlecht“ bereits im zweiten Schritt bzw. Modell stark an Bedeutung verliert (verglichen mit der statistischen Bedeutung im ersten Modell). Der Prädiktor AW (S) zeigt einen konsistent starken und den stärksten statistischen Einfluss auf das Kriterium GI (S). Lediglich im letzten Modell lässt sich ein nennenswerter mit dem Prädiktor GI (E) geteilter und damit leicht reduzierter statistischer Einfluss auf das Kriterium erkennen (das standardisierte Regressionsgewicht sinkt von  $\beta \approx .80$  im zweiten und dritten Modell auf  $\beta = .60$  im vierten Modell).

Für den Prädiktor FQ ist zu beobachten, dass der im bivariaten Fall mindestens mitt-

lere Effekt ( $r(57) = -.37$ ) bei Berücksichtigung der Kontrollvariable „Alter“ merklich auf einen höchstens mittleren Effekt absinkt (teilpartielle Korrelation  $r_{GI(S)}(FQ, Alter)(1;56) = -.22$ ). Bei Berücksichtigung der beiden anderen Prädiktoren verringert sich der statistische Einfluss von FQ weiter auf  $\beta = -.13$  mit  $p = .076$ . Damit weist die Rechtschreibkompetenz den geringsten statistischen Einfluss der drei Prädiktoren auf die Textqualität auf. Auch die Größenordnung des Effekts der Varianzaufklärung selbst lässt sich bei Berücksichtigung der anderen Prädiktoren als gering einordnen (für  $\Delta R^2 = .02$  ergibt sich  $f^2 = r^2/(1-r^2) = .02$ , Cohen, 1988, zitiert nach Eid, Gollwitzer & Schmitt, 2015, S. 610, modifiziert für post-hoc Bestimmung).

Der Prädiktor GI (E) zeigt, ähnlich wie der Prädiktor AW (S), mit  $r(57) = \beta = .68$  zunächst (im bivariaten Fall) einen sehr starken statistischen Einfluss. Bei Konstanthaltung aller anderen relevanten Variablen (Alter, Geschlecht, AW [S], FQ) reduziert sich die statistische Bedeutung dieses Prädiktors allerdings deutlich auf  $\beta = .32$ ,  $\Delta R^2 = .07$  und  $f^2 = .07$ . Ein Effekt dieser Größenordnung ist als relativ klein zu bezeichnen (kleiner Effekt  $f^2 = .02$ , mittlerer Effekt  $f^2 = .15$ , großer Effekt  $f^2 = .35$ , Cohen, 1988, zitiert nach Eid, Gollwitzer & Schmitt, 2015, S. 610). Das Absinken der Steigung durch das schrittweise Auspartialisieren der anderen Prädiktoren fällt bei GI (E) sehr deutlich aus. So ist der Unterschied zwischen der bivariaten Korrelation im Vergleich zur Semipartialkorrelation in Schritt 4 statistisch bedeutsam (überschneidungsfreie Konfidenzintervalle für die Nullkorrelation  $r(57) = .68$  mit 95%-Konfidenzintervall  $KI = [.57; 1]$  und für die Semipartialkorrelation des vierten Modells  $r_{AW(S) (GI [E], alle Prädiktoren und Kontrollvariablen)} = .26$  mit 95%-Konfidenzintervall  $KI = [.00; .53]$ ). Darüber hinaus lässt sich ein Wechsel der Größenordnung des Effekts (von „stark“ auf „höchstens mittel“) beobachten.

Insgesamt erweist sich bei gemeinsamer Betrachtung von Alter, Geschlecht, AW (S), FQ und GI (E) die AW (S) als stärkster und robustester Prädiktor für den GI (S). An dieser Stelle muss allerdings ausdrücklich darauf hingewiesen werden, dass sich diese statistische Bedeutung nur im Zusammenspiel mit den anderen in das Modell einbezogenen Prädiktoren zeigt.

Wie bereits an den Korrelationen in Tabelle 2 abschätzbar, liegt zwischen den Prädiktoren keine hohe Multikollinearität vor. Dies bestätigen auch die Toleranz- und Varianzinflationswerte für die unabhängigen Variablen ( $TOL \geq .57$  vs.  $TOL_{plu} = 1$  bei einer perfekten linearen Unabhängigkeit (plu) eines Prädiktors von den anderen Prädiktoren,  $VIF \leq 1.8$  vs.  $VIF_{plu} = 1$  bei vollständiger linearer Unabhängigkeit (plu)).

Bei den drei in das Regressionsmodell eingehenden Variablen GI (S), AW (S) und FQ lassen sich jedoch deutlich schiefe Verteilungen beobachten (Schiefen von jeweils 2.61, 13.29 und 7.4,  $SE_{Sch} = 0.61$ ). Obwohl die Regressionsanalyse auf nicht normale (hier schiefe) Verteilungen robust reagiert, wurde das berichtete Regressionsmodell mit durch Logarithmierung zur Basis 10 annähernd normalisierten Werten der entsprechenden Variablen erneut berechnet (zur Robustheit der linearen Regression und Transformationen schiefer Verteilungen siehe Bortz & Schuster, 2010). Die Regressionsgewichte und die Determinationskoeffizienten bleiben dabei stabil und in den Größenordnungen vergleichbar. Lediglich für die Variable FQ ergibt sich bereits im dritten Schritt ein  $p$ -Wert von über 5% ( $\beta = -.08$  und  $\Delta R^2 = .01$  mit  $p = .27$ ) und dann, vergleichbar mit dem oben berichteten Regressionsmodell, kein statistisch signifikanter Beitrag im letzten Schritt (Schritt 4:  $\beta = -.08$  mit  $p = .19$ ).

Die Residuenanalyse zeigt keine besonderen Auffälligkeiten. Zu berücksichtigen sind allerdings ein paar Ausreißerwerte in den Variablen AW (S) ( $n = 4$ ), FQ ( $n = 5$ ) und GI (S) ( $n = 3$ ). Da Ausreißerwerte die statistischen Beziehungen besonders bei eher kleinen Stichproben stark verzerren können, wurde die Regressionsanalyse ohne Ausreißer (insgesamt betrifft das  $n = 8$  Kinder) erneut berechnet. Bei Ausschluss der acht Kinder ergibt die Regressionsanalyse sehr vergleichbare Werte zu den in Tabelle 3 dargestellten. Lediglich leichte Änderungen zeigen sich für den Anteil der durch die Kontrollvariablen aufgeklärten Varianz (bei Ausschluss der acht Ausreißerwerte  $R^2 = .10$ ,  $p = .076$ ) sowie für den Anteil der Varianz, der inkrementell durch den GI (E) aufgeklärt wird ( $\Delta R^2 = .09$ ,  $p < .001$ , ebenfalls acht ausgeschlossene Werte). Insgesamt lassen sich die Ergebnisse der vorgelegten Analysen daher als robust einschätzen.

## Diskussion

### *Beantwortung der Fragestellung*

Das Ziel der vorliegenden Untersuchung bestand darin, die Bedeutung der Fähigkeiten, flüssig zu schreiben, auf einen breiten mündlichen Wortschatz zurückzugreifen und wenig orthographische Fehler zu begehen, für die Qualität von Aufsätzen zu bestimmen. Da das Verfassen von Geschichten als die erste wichtige Hürde im Hinblick auf die Entwicklung von expressiven Schreibkompetenzen im engeren Sinne gilt und die Produktion anderer Textsorten erst später an Relevanz gewinnt, wurden solche Mädchen und Jungen in die Studie mit einbezogen, die gerade dabei sind, sich diese basale Form der Textproduktion anzueignen: Kinder aus 3. und 4. Grundschulklassen. Wenn die drei genannten basalen Komponenten (Schreibflüssigkeit, mündlicher Wortschatz und orthographische Kompetenz) einen signifikanten prognostischen Wert für die Varianz der Aufsatzqualität besitzen, so ließe sich hieraus eine Begründung für ihre gezielte Prävention und Intervention ableiten.

Es ist an dieser Stelle zu konstatieren, dass die Länge der verfassten Geschichten den stärksten Einfluss auf die Textqualität (operationalisiert über die lexikalische Vielfalt im Schreiben) ausübt. Dieses Ergebnis korrespondiert mit den Befunden von Grünke et al. (2015). In dieser Studie mit 10- bis 13jährigen Schülerinnen und Schülern wurde ein anderer Indikator für die Textqualität verwendet als die Breite des Wortschatzes, nämlich die durch acht unterschiedliche Raterinnen und Rater vorgenommenen Einschätzungen mittels einer deutschen Version der NAEP Writing Scale (Canz, 2015). Auch hier zeigte sich, dass die Varianz der Anzahl der geschriebenen Wörter enger mit der Varianz der Bewertungen korrespondiert als jeder andere Faktor (Lesbarkeit des Textes, Einheitlichkeit der Handschrift, Fehleranteil, etc.).

Die mündliche lexikalische Vielfalt übt ebenfalls einen statistischen Effekt auf die

Güte der Geschichten aus. Wenn Kinder beim Erzählen in der Lage sind, auf ein abwechslungsreiches Repertoire an Wörtern zurückzugreifen, dann ist auch das Vokabular ihrer geschriebenen Texte relativ mannigfaltig. Allerdings muss der Einfluss dieser Variable als eher gering angesehen werden. Noch weniger wesentlich ist die (statistische) Wirkung des Fehleranteils. Dieser spielt lediglich eine untergeordnete Rolle. Zwar wirken sich Probleme in der Orthographie negativ auf die Qualität der Geschichten aus, allerdings nur in kleinem Ausmaß.

Wie oben bereits betont, ist die mehr oder minder relevante Bedeutung der einzelnen Prädiktoren unbedingt vor dem Hintergrund des Zusammenspiels der einzelnen Variablen zu interpretieren. Betrachtet man die Bedeutung der mündlichen lexikalischen Vielfalt und der Rechtschreibung isoliert, so ist ihr Effekt auf die Textqualität merklich größer als im vollständigen Modell der linearen multiplen Regression. Die Annahme, dass die Schreibflüssigkeit, die Breite des mündlichen Wortschatzes und die orthographischen Fähigkeiten die Qualität von Geschichten vorhersagen können, die von Kindern am Ende ihrer Grundschulzeit verfasst wurden, lässt sich folglich mit Einschränkungen als bestätigt ansehen. Alle drei Variablen üben ganz offenkundig einen Einfluss auf die Güte der Texte aus, allerdings spielt lediglich die Schreibflüssigkeit eine wirklich gewichtige Rolle.

Einige Nebenergebnisse bedürfen kurzer Kommentare: Die Streuungen bei den Variablen AW (E) und FQ unterscheiden sich in den beiden Gruppen markant. Wurde zu Beginn des Schultages erst erzählt und erst später geschrieben, so schwankte die Menge der Wörter stärker als bei den Kindern, die erst schreiben sollten. Demgegenüber variierte der Anteil an orthographischen Fehlern bei den zuerst schreibenden Mädchen und Jungen deutlicher als bei den zuerst erzählenden. Diese Beobachtung lässt sich möglicherweise durch zufällig zustande gekommene Ausreißerwerte erklä-

ren. In der Gruppe der zuerst schreibenden Schülerinnen und Schüler gab es niemanden, dessen erzählte Geschichte weniger als 40 und mehr als 270 Wörter umfasste. In der Parallelgruppe befanden sich hingegen zwei Kinder mit extrem kurzen (AW [E] < 25) und zwei mit langen (AW [E] = 312) Ausführungen. Die Darstellung eines weiteren Mädchens setzte sich sogar aus 449 Wörtern zusammen. Eliminiert man diese fünf Werte, so reduziert sich die Standardabweichung von  $SD = 92.86$  auf  $SD = 60.02$  (in der Gruppe der zuerst schreibenden Kinder liegt die Standardabweichung bei 60.18). Unter den zuerst schreibenden Kindern befand sich ein Mädchen mit einem extrem hohen FQ von 83.33. Wenn man es aus der Analyse ausklammert, so sinkt die Standardabweichung dieser Variable in ihrer Gruppe von  $SD = 18.37$  auf  $SD = 13.02$  (im Vergleich zu  $SD = 11.13$  in der Parallelgruppe).

### *Methodenkritische Reflexion*

Wie bei allen Arbeiten aus der human- und sozialwissenschaftlichen Forschung muss auch in der vorliegenden Untersuchung auf einige Einschränkungen der Aussagen hingewiesen werden. Zunächst einmal ist an dieser Stelle anzumerken, dass diese Studie letztendlich einen Beitrag zur Beantwortung der Fragestellung leisten soll, wie man das Auftreten von expressiven Schreibstörungen frühzeitig und wirksam verhindern kann. Allerdings stellt unsere Untersuchung in diesem Kontext lediglich einen Zwischenschritt dar. Es lassen sich auf der Basis der hier präsentierten Befunde lediglich Aussagen darüber treffen, welche grundlegenden Fähigkeiten die Aufsatzqualität von Grundschulkindern und Grundschüler mitbedingen und nicht darüber, inwieweit sich ihre Förderung auf die Güte von Schreibprodukten bzw. auf das Auftreten von Schreibstörungen auswirkt. Die statistischen Analysen der vorliegenden Daten erlauben es zwar, die Möglichkeit zuzulassen, dass eine Verbesserung in der Produktion längerer Texte

(Steigerung der AW [S]) und lexikalisch vielfältiger Erzählungen (Steigerung des GI [E]) sowie der Übung orthografischer Fertigkeiten (Senkung des FQ) zu einer Verbesserung in der schriftlichen lexikalischen Vielfalt führt. Jedoch kann nur ein experimentelles oder quasi-experimentelles Vorgehen sicherstellen, dass ein solches Kausalgefüge bei der Zusammensetzung schriftlicher lexikalischer Vielfalt tatsächlich vorliegt. Doch trotz aller Einwände zeigt die vorliegende Untersuchung, dass die drei unabhängigen Variablen 1) schriftliche und 2) mündliche Wortanzahl bei der Produktion von Texten sowie 3) der Fehlerquotient als erklärende Variablen im kausalen Sinne für Unterschiede in der schriftlichen lexikalischen Vielfalt als ein Aspekt der Textqualität in Frage kommen und sich zunächst in einer korrelativ angelegten Untersuchung bewährt haben.

Ferner sei auf eine weitere typische Einschränkung bei der Interpretation der Ergebnisse empirischer Untersuchungen hingewiesen. Als ein wesentlicher Aspekt der Textqualität wurde in der vorliegenden Studie die schriftliche lexikalische Vielfalt erfasst. Auch wenn die lexikalische Vielfalt sich konsistent als ein wichtiges Merkmal der Textqualität zeigt (siehe Einleitung), muss berücksichtigt werden, dass Textqualität insgesamt mehr umfasst als lexikalische Vielfalt. Aus verschiedenen (methodischen und methodologischen) Gründen ist es jedoch sinnvoll, sich auf wenige Variablen, insbesondere abhängige Variablen, zu beschränken (siehe etwa Cohen, 1990). Insgesamt halten wir daher die Betrachtung der Unterschiede in der schriftlichen lexikalischen Vielfalt als einer bedeutsamen Komponente der Textqualität in diesem Rahmen als sinnvoll und aufschlussreich.

Ein weiteres Problem betrifft die Operationalisierung der Textqualität. Der GI (S) ist sicherlich nicht über jede Kritik erhaben. Wie oben dargestellt, hängt dieser Indikator in unserem Fall eng mit AW (S) zusammen. Zwar ist der GI (S) in aller Regel von der Länge eines Schreibprodukts relativ unabhängig



(siehe z. B. Sumner, Connelly & Barnett, 2014), allerdings besteht in diesem Kontext keine völlige Eigenständigkeit (Koizumi & In'nami, 2012). Die Korrelation dürfte in Anbetracht unserer weitergehenden Analysen nicht ausschließlich redundanter Natur sein. Dennoch lässt sich nicht leugnen, dass zwischen beiden Variablen in der vorliegenden Studie eine verhältnismäßig starke Beziehung existiert. Darüber hinaus kann man einem Index zur Erfassung der lexikalischen Vielfalt als Kriterium für die Textqualität eine gewisse Eindimensionalität unterstellen. Es lassen sich sicherlich plausible Argumente ins Feld führen, die dafür sprechen, andere Merkmale als die Breite des Wortschatzes zur Operationalisierung der Güte einer Geschichte heranzuziehen.

Anders als bei vielen anderen Kompetenzen im Schriftsprachbereich (z. B. Buchstabenkenntnisse, Leseflüssigkeit, Textverständnis, u. a.) existiert kein allgemein anerkanntes und übliches Vorgehen bei der Messung dieses Konstrukts. In den vielen Arbeiten, die in den bislang vorliegenden Sekundäranalysen über die Wirksamkeit von Schreibinterventionen Berücksichtigung finden (Cook & Bennett, 2014; Datchuk & Kubina, 2012; Gillespie & Graham, 2014; Rogers & Graham, 2008), kommen so genannte „Writing Rubrics“ zum Einsatz. Es handelt sich hierbei in aller Regel um von den jeweiligen Autorinnen und Autoren eigens konzipierte mehrstufige Skalen, mit deren Hilfe sich die Güte eines Schreibprodukts bezüglich einer Reihe von Kriterien (z. B. Struktur, Vollständigkeit, Kreativität, ...) beurteilen lässt. Auf diese Weise ist es möglich, auf viele verschiedene Facetten der Qualität einzugehen. Allerdings haftet einem solchen Vorgehen eine gewisse Beliebbarkeit an. Welche Aspekte letztendlich als maßgeblich angesehen werden, lässt sich nicht objektiv begründen. Außerdem ist es häufig schwer, eine akzeptable Übereinstimmung zwischen verschiedenen Raterinnen und Ratern zu erzielen (Spencer & Fitzgerald, 1993; Van Steendam, Tillema, Rijlaarsdam, & Van den Bergh, 2012).

Für die Verwendung des GI (S) als Indikator für die Qualität von Geschichten sprechen trotz aller berechtigter Kritik v. a. die folgenden Argumente: (1) Dieses Maß hat sich in vorangegangenen Untersuchungen als valides Kriterium für die Güte von Schreibprodukten bewährt (z. B. Mellor, 2011). (2) Obwohl der GI (S) und andere Optionen zur Messung der lexikalischen Vielfalt immer eine mehr oder minder große Abhängigkeit von der Textlänge aufweisen, werden sie im Rahmen zahlreicher standardisierter Testverfahren (z. B. IELTS, MELAB, TOWL-4) als zentrales Indiz für die Textqualität herangezogen. Eine solche Praxis stellt in Situationen, in denen man die Güte von Schreibprodukten relativ aufwändig und vereinheitlicht erfasst, einen gewissen Standard dar. (3) Im Vergleich zu den häufig verwendeten „Writing Rubrics“ ist der GI (S) äußerst objektiv. (4) Die oben zitierte Studie von Grünke et al. (2015) spricht dafür, dass der enge Zusammenhang zwischen der Länge von geschriebenen Geschichten und dem GI (S) bei Schülerinnen und Schülern während der Entwicklungsphase, in der sie erstmals Erzählungen aus der Fremdperspektive verfassen können (siehe Becker-Mrotzek & Böttcher, 2012), ein tatsächlich vorhandenes Phänomen und kein Artefakt widerspiegelt. Denn auch hier zeigte sich, dass der Anzahl der verfassten Wörter die größte Bedeutung für die Güte von Textprodukten zukommt. Und das, obwohl bei der Operationalisierung der Qualität eine Ratingskala eingesetzt wurde.

Was die Prädiktoren in unserer Untersuchung angeht, so ließe sich die Schreibflüssigkeit auch unter Berücksichtigung der benötigten Zeit ermitteln (vgl. Chenoweth & Hayes, 2001), allerdings weist dieses Vorgehen die weiter oben beschriebenen Nachteile auf. Die mündliche lexikalische Vielfalt ließe sich alternativ etwa mit Hilfe der Subskalen zum Sprechen und Wortschatz des „Allgemeinen Deutschen Sprachtests“ (ADST) von Steinert (2011) erfassen. Schließlich könnte man die Rechtschreibkompetenz anstatt über den FQ unter Zuhil-

fenahme frei formulierter Texte auch mittels standardisierter Tests erfassen, z. B. durch den „Deutschen Rechtschreibtest für das dritte und vierte Schuljahr“ (DERET 3-4+) von Stock und Schneider (2008) oder den „Salzburger Lese- und Rechtschreibtest“ (SLRT-II) von Moll und Landerl (2014). Das von uns gewählte Vorgehen hat jedoch den Vorteil, Rechtschreibfertigkeiten und mündliche lexikalische Vielfalt wesentlich effizienter und für die Kinder sehr viel weniger aufwändig zu ermitteln (z. B. beträgt die Durchführungszeit für den ADST je nach Testteil 20 bis 90 Minuten, für den DERET 3-4+ etwa 30 Minuten). Da der FQ von Lehrkräften in der Praxis regelmäßig zum Einsatz kommt und ein objektives Maß zur Beurteilung von Fehlern darstellt, wurde er von uns als Operationalisierung der Rechtschreibfertigkeiten gewählt. Der große Vorteil beim Einsatz eines Sprachtests, wie des ADST oder des DERET 3-4+ besteht vor allem darin, anhand der verfügbaren Normtabellen individualstatistische Aussagen treffen zu können (etwa über den Kompetenzstand eines einzelnen Kindes relativ zu seiner Normgruppe). Da diese Studie jedoch gruppenstatistische Analysen vornimmt, scheint ein höchstens mittlerer inkrementeller Genauigkeitszuwachs den wesentlich größeren Aufwand bei der Erhebung für die Kinder nicht zu rechtfertigen. Mit dem gewählten Vorgehen ist es möglich, durch lediglich zwei Erhebungsleistungen von jeweils maximal 30 Minuten seitens der Kinder alle für die vorliegende Untersuchung relevanten Variablen mit hinreichender Genauigkeit und einer hohen Objektivität zu erfassen.

Als weitere Einschränkung ist an dieser Stelle die mangelnde Generalisierbarkeit der Befunde in Anbetracht der kleinen und speziellen Stichprobe zu nennen. Die teilnehmenden Kinder stammten allesamt aus einer speziellen Schule und damit aus einer ganz bestimmten geographischen Gegend. Somit sind Schlussfolgerungen bezüglich der Population aller Kinder aus 3. und 4. Klassen mit Vorsicht zu genießen.

In der vorliegenden Studie wurden drei unabhängige Variablen ins Blickfeld genommen, um die Textqualität vorausszusagen. Dies geschah auf der Basis einer ausführlichen Literaturrecherche bzw. zahlreicher Hinweise in einschlägigen Texten, in denen immer wieder entsprechende Zusammenhänge postuliert werden. Allerdings ist anzunehmen, dass auch noch andere Faktoren Auswirkungen auf die Qualität von Aufsätzen ausüben, die Mädchen und Jungen am Ende ihrer Grundschulzeit verfassen. Offenbar handelt es sich beim Zusammenspiel der Variablen, welche die Qualität von Geschichten bedingen, um ein komplexes Gefüge, in dem sich die einzelnen Elemente gegenseitig beeinflussen. Wie gesagt scheint die Rechtschreibkompetenz in diesem Kontext nicht allzu bedeutsam zu sein. Diese Aussage gilt allerdings nur dann, wenn man die orthographischen Fehler gemeinsam mit der Anzahl der geschriebenen Wörter betrachtet. Vor dem Hintergrund dieser Komplexität ist davon auszugehen, dass hier noch weitere (und von uns nicht berücksichtigte) Variablen eine Rolle spielen.

### *Praktische Implikationen*

Schülerinnen und Schüler darin zu fördern, qualitativ hochwertige Kommentare, Berichte, Reportagen, Erzählungen und andere Schreibergebnisse verfassen zu können, wird im Unterricht oftmals vernachlässigt, da sich Lehrkräfte mit der Vermittlung dieser komplexen Kompetenz überfordert fühlen (Grünke & Leonard-Zabel, 2015; Troia & Graham, 2003). Dieser Eindruck der Überforderung resultiert oftmals aus der fälschlichen Annahme, dass entsprechende Ansätze stets den gesamten Prozess der Textproduktion im Auge behalten müssten. Allerdings sollten Interventionen, die zur Verbesserung der Schreibfähigkeit dienen, immer an Teilfertigkeiten des Gesamten ansetzen und nicht an allen Grundfertigkeiten zeitgleich arbeiten (Ise, Engel & Schulte-Körne, 2012).

Es gibt, wie beschrieben, eine Vielzahl von Faktoren bzw. von Vorläuferkompetenzen, welche die Fähigkeit des Verfassens von qualitativ hochwertigen Geschichten beeinflussen. Zu den in dieser Studie untersuchten Variablen (Schreibflüssigkeit, Breite des Wortschatzes und orthographische Richtigkeit), kommt höchstwahrscheinlich noch eine Vielzahl weiterer Merkmale hinzu, die hier keine Berücksichtigung finden konnten. Es ist anzunehmen, dass letztendlich das Zusammenspiel vieler verschiedener Facetten die Qualität der Geschichten von Schülerinnen und Schülern bestimmt. Nicht jede Variable nimmt dabei aber in gleichem Maße Einfluss auf die Güte der Texte. So konnte gezeigt werden, dass vor allem die Länge zu einer höheren Qualität des Geschriebenen beizutragen scheint. Laut Grabowski, Becker-Mrotzek, Knopp, Jost und Weinzierl (2014) trifft dies insbesondere bei Fünft- bis Neuntklässlerinnen und -klässlern zu. Konsequenterweise muss ein wichtiges Ziel von Schreibförderung während dieser Entwicklungsphase darin bestehen, die Produktivität der Mädchen und Jungen zu steigern.

Gemäß der Metaanalyse von Rogers und Graham (2008) sind die Ansätze des *Goal Settings for Productivity*, des *Reinforcements* und der *Prewriting Activities* besonders dazu geeignet, Schülerinnen und Schüler zum Verfassen relativ langer Texte zu animieren. Durch eine Zielformulierung vor dem Schreiben (*Goal Setting for Productivity*), die durch die Lehrkraft oder die Kinder bzw. Jugendlichen selbst vorgenommen wird, fühlen sich die Mädchen und Jungen meist dazu motiviert, vergleichsweise ausführliche Geschichten zu produzieren. Zielsetzungen beziehen sich hierbei entweder auf eine bestimmte Anzahl an Wörtern oder auf die Vorgabe, beim nächsten Mal mehr zu schreiben als beim letzten. Die Wirksamkeit dieser Ansätze wird u. a. dadurch erhöht, indem die Kinder oder Jugendlichen ein unmittelbares Feedback darüber erhalten, ob das Ziel erreicht wurde (ebd.).

Eine zusätzliche Steigerung der Effektivität lässt sich über *Reinforcements* (also über Belohnungen) erwirken, die Schülerinnen und Schüler nach dem Erreichen einer Etappe erhalten. Die Palette an verbalen und materiellen Verstärkern ist hierbei ausgesprochen breit (Grünke & Coeppicus, 2017; Linderkamp, 2014). Eine zusätzliche Maßnahme zur Steigerung der Schreibmotivati-on stellt das Finden von geeigneten Möglichkeiten dar, die jeweils beste Leistung eines Kindes oder Jugendlichen zu operationalisieren (etwa über die Anzahl der korrekten Wortsequenzen) und an irgendeiner prominenten Stelle zur Schau zu stellen (z. B. auf einer Karteikarte in einem Sichtfenster einer häufig verwendeten Arbeitsmappe [Grünke, Sperling & Burke, 2017]).

Schließlich können auch *Prewriting Activities* dazu beitragen, die Textlänge zu erhöhen. Hierzu gehören alle Maßnahmen, die zur Planung einer Geschichte eingesetzt werden, bevor der Prozess des Translatings (also der Überführung von Gedanken in einen Text) geschieht. Als wichtigste Beispiele lassen sich in diesem Kontext alle Arten von Graphic Organizern (wie zum Beispiel Geschichtenkarten) nennen, die dabei helfen, die Ideen vor dem Schreiben zu strukturieren, um letztendlich besser geplante und zumeist auch längere Texte zu verfassen (Grünke, Wilbert, Tsirotakis & Agirregoikoa, 2017).

Als weniger wichtig für die Verbesserung der Textqualität sind die Variablen der lexikalischen Vielfalt und des Fehleranteils einzuordnen. Aber auch sie sollten als zu fördernde Elemente nicht außer Acht gelassen werden. Die Breite des Wortschatzes lässt sich über kurze, aber regelmäßige Vokabeltrainings erweitern (Solís, Sammacca, Barth & Roberts, 2017). Dazu bietet sich laut Green, Stockhom, Cearley und Sheffield-Anderson (2015) vor allem für schwächere Lernende die direkte Vokabelinstruktion an. Hierbei werden den Schülerinnen und Schülern unbekannte Vokabeln präsentiert und durch ein von der Lehrkraft gesteuertes Vorgehen mit Hilfe von Modellierung

und Einübung, deren Definitionen, Bedeutungen im Kontext und sinnvollen Anwendungen vermittelt (ebd.; Lebens & Lauth, 2014). Zusätzlich zur Erweiterung des Wortschatzes kann der Fehleranteil in Schülertexten durch spezielle Rechtschreibtrainings minimiert werden. Die Ergebnisse der Metaanalyse von Ise, Engel und Schulte-Körne (2012) weisen darauf hin, dass aus der Vielzahl bereits bestehender Programme für Schülerinnen und Schüler mit Rechtschreibschwierigkeiten speziell solche Trainings geeignet sind, die konkret an den vorhandenen Schwierigkeiten der Schülerinnen und Schüler ansetzen (also symptomorientiert arbeiten) und die einen Zeitraum von mindestens 20 Wochen umfassen.

### Ausblick

In weiterführenden Forschungsarbeiten sollte versucht werden, die Befunde aus der vorliegenden Studie auf verschiedene Weise zu untermauern. Wie bereits angesprochen gibt es diverse Möglichkeiten, die von uns ins Auge gefassten Variablen (Textqualität, Schreibflüssigkeit, mündliche lexikalische Vielfalt und Rechtschreibkompetenz) zu operationalisieren bzw. diagnostisch zu erfassen. Die hier zu Tage geförderten Befunde sollten sich auch bei einer Umsetzung des Prinzips der multimodalen Konvergenz (siehe Fäh & Fischer, 1998) immer wieder beobachten lassen. Außerdem ist die Auswahl unserer unabhängigen Variablen zu überdenken. Textqualität ist ganz offenbar das Resultat eines sehr komplexen Zusammenwirkens verschiedener Faktoren. Es wäre vermessen, davon auszugehen, mit der Schreibflüssigkeit, der Breite des mündlichen Wortschatzes und der orthographischen Kompetenz alle Facetten berücksichtigt zu haben, die in diesem interaktionalen Prozess eine nennenswerte Rolle spielen. Schließlich gilt es, die hier präsentierten Befunde nicht nur mit anderen Formen der Operationalisierung der unabhängigen Variablen (bzw. anderen unabhängigen Variablen) zu replizieren, sondern ihre Validität

auch mit verschiedenen Stichproben, verschiedenen Untersuchungsdesigns und verschiedenen Textgattungen zu überprüfen.

In Anbetracht der hohen Anzahl an Schülerinnen und Schülern mit expressiven Schreibstörungen steht die Sonderpädagogik als empirische Disziplin in der Verantwortung, Ansatzpunkte für die Verbesserung der Inklusionschancen der Betroffenen in einer Welt zu finden, in der schriftliche Kommunikation allgegenwärtig ist. Die vorliegende Studie liefert hierzu einen bescheidenen Beitrag. Allerdings bleibt die Notwendigkeit für mehr Forschungsarbeiten zu diesem Themenkomplex hiervon unberührt.

### Literatur

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington, DC: APA.
- Alves, R. A.; Limpo, T., Fidalgo, R., Carvalhais, L., Pereira, L. Á. & Castro, S. L. (2015). The impact of promoting transcription on early text production: Effects on bursts and pauses, levels of written language, and writing performance. *Journal of Educational Psychology, 108*, 665-679.
- Becker-Mrotzek, M. (2014). Schreibkompetenz. In J. Grabowski (Hrsg.). *Sinn und Unsinn von Kompetenzen: Fähigkeitskonzepte im Bereich Sprache, Medien und Kultur* (S. 51-71). Opladen: Budrich.
- Becker-Mrotzek, M. & Böttcher, I. (2006). *Schreibkompetenz entwickeln und beurteilen: Praxishandbuch für die Sekundarstufe I und II*. Berlin: Cornelsen.
- Becker-Mrotzek, M. & Böttcher, I. (2012). *Schreibkompetenz entwickeln und beurteilen*: Berlin: Cornelsen.
- Berninger, V. W. & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. MacArthur, S. Graham & J.

- Fitzgerald (Eds.), *Handbook of writing research* (pp. 96-114). New York, NY: Guilford.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Cambridge Michigan Language Assessments (2015). *Michigan English Language Assessment Battery (MELAB)*. Ann Arbor, MI: English Language Institute (ELI) at the University of Michigan.
- Canz, T. (2015). *Validitätsaspekte bei der Messung von Schreibkompetenzen*. <http://edoc.hu-berlin.de/dissertationen/canz-thomas-2015-10-19/PDF/canz.pdf> (Zugriff am 01.09.2017).
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*, 80-98.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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*, 278-302.
- Cook, K. B. & Bennett, K. E. (2014). Writing interventions for high school students with disabilities: A review of single-case design studies. *Remedial and Special Education, 35*, 344-355.
- Daller, M. (2010). *Guirauds index of lexical richness*. [http://eprints.uwe.ac.uk/11902/1/11902\\_Guiraud%27s\\_index\\_of\\_lexical\\_richness\\_%28MD\\_BAAL\\_2010%29.pdf](http://eprints.uwe.ac.uk/11902/1/11902_Guiraud%27s_index_of_lexical_richness_%28MD_BAAL_2010%29.pdf) (Zugriff am 01.09.2017).
- Daller, H., van Hout, R. & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics, 24*, 197-222.
- Datchuk, S. M. & Kubina, R. M. (2012). A review of teaching sentence-level writing skills to students with writing difficulties and learning disabilities. *Remedial and Special Education, 34*, 180-192.
- Dilling, H., Mombour, W., Schmidt, M. H. & Coltart, I. (2015). *Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V*. Göttingen: Hogrefe.
- Fäh, M. & Fischer, G. (1998). *Sinn und Unsinn in der Psychotherapieforschung*. Gießen: Psychosozial Verlag.
- Fayol, M. L., Alamargot, D. & Berninger, V. W. (2012). From cave writers to elite scribes to professional writers to universal writers, translating is fundamental to writing. In M. L. Fayol, D. Alamargot & V. W. Berninger (Eds.), *Translation of thought to written text while composing* (p. 3-14). New York, NY: Taylor & Francis.
- Gillespie, A. & Graham, S. (2014). A meta-analysis of writing interventions for students with learning disabilities. *Exceptional Children, 80*, 454-473.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.
- 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). Frankfurt/M.: Lang.
- Graham, S. (2010). Want to improve children's writing? Don't neglect their handwriting. *The Education Digest, 76*, 49-55.
- Graham, S., Collins, A. A. & Rigby-Wills, H. (2007). Writing characteristics of students with learning disabilities and typically achieving peers: A meta-analysis. *Exceptional Children, 83*, 199-218.
- Graham, S, Harris, K. & Chorzempa, B.F. (2002). Contribution of spelling instruction to the spelling, writing and reading of poor spellers. *Journal of Educational Psychology, 94*, 669-686.
- Green, L., Stockholm, M., Cearley, J. & Sheffield-Anderson, L. (2015). Direct vocabulary instruction with two 5th-grade English-language learners with language-learning disabilities: A treatment study. *Contemporary Issues in Communication Science and Disorders, 42*, 191-201.



- Grünke, M., Büyüknarci, Ö., Wilbert, J. & Breuer, E. (2015). To what extent do certain characteristics of a child's written story influence the way it is rated? Insights into features necessary for supporting struggling writers. *Insights into Learning Disabilities, 12*, 163-177.
- Grünke, M. & Coeppicus, C. (2017). Contingency contracting and its influence on the punctuation skills of fifth graders with learning disabilities. *Insights into Learning Disabilities, 14*, 125-134.
- Grünke, M. & Leonard Zabel, A. M. (2015). How to support struggling writers. *International Journal of Special Education, 30*, 137-150.
- Grünke, M., Sperling, M. & Burke, M. D. (2017). The effects of immediate feedback, explicit timing, and positive reinforcement on the writing performance of struggling 5th grade students in Germany. *Insights into Learning Disabilities, 14*, 135-153.
- Grünke, M. & Strathmann, A. (2007). Förderung des Lesens. In F. Linderkamp & M. Grünke (Hrsg.), *Lern- und Verhaltensstörungen* (S. 198-209). Weinheim: Beltz.
- Grünke, M., Wilbert, J., Tsiriatakis, I. & Agirre-goikoa, A. L. (2017). Improving the length and quality of narratives written by fourth-graders with learning disabilities through a peer tutoring graphic organizing strategy. *Insights into Learning Disabilities, 14*, 167-188.
- Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique*. Paris, France: Presses Universitaires de France.
- Hammill, D. D. & Larson, S. C. (2009). *Test of Written Language* (TOWL-4). Austin, TX: Pro-Ed.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum.
- Hayes, J.R., & Flower, L.S. (1980). Identifying the organisation of writing processes. In L.W. Gregg & E.R.; Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NY: Lawrence Erlbaum.
- Hirmer, M. & Hirmer, E. (2007). *Erfolgreich und kreativ durch alle Aufsatzarten. 5./6. Klasse. Band 1*. München: PB.
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: Guilford.
- Ise, E., Engel, R. R., Schulte-Körne, G. (2012). Was hilft bei der Lese-Rechtschreibstörung? Ergebnisse einer Metaanalyse zur Wirksamkeit deutschsprachiger Förderansätze. *Kindheit und Entwicklung, 21*, 122-136.
- Karadag, R. & Kayabasi, B. (2013). Future scenarios regarding tablet computer usage in education and writing. *Asian Social Science, 9*, 105-110.
- Katusic, S. K., Colligan, R. C., Weaver, A. L., & Barbaresi, W. J. (2009). The forgotten learning disability: Epidemiology of written-language disorder in a population-based birth cohort (1976-1982). *Pediatrics, 123*, 1306-1313.
- Kellog, R. T. & Raulerson, B. A. (2007). Improving the writing skills of college students. *Psychonomic Bulletin & Review, 14*, 237-242.
- Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 Tokens? *Vocabulary Learning and Instruction, 1*, 60-69.
- Koizumi, R & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System, 40*, 554-564.
- Kultusministerium Hessen (2016). *Oberstufen- und Abiturverordnung (OAVO)*. [https://kultusministerium.hessen.de/sites/default/files/media/2016-08-09\\_textfassung\\_oavo\\_zuletzt\\_geaendert\\_durch\\_verordnung\\_vom\\_13\\_juli\\_2016.pdf](https://kultusministerium.hessen.de/sites/default/files/media/2016-08-09_textfassung_oavo_zuletzt_geaendert_durch_verordnung_vom_13_juli_2016.pdf)
- Latif, M. M. M. A. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics, 34*, 99-105.

- Laufer, B., & Nation, P. (1995). Vocabulary size and use lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Lebens, M. & Lauth, G. W. (2014). Direkte Instruktion. In G. W. Lauth, M. Grünke & J. C. Brunstein (Hrsg.) *Interventionen bei Lernstörungen. Förderung, Training und Therapie in der Praxis* (S. 418-428). Göttingen: Hogrefe.
- Lenhard, W. & Lenhard, A. (2014). *Signifikanztests bei Korrelationen*. <https://www.psychometrica.de/korrelation.html> (Zugriff am 01.09.2017).
- Lougheed, L. (2007). *International English Language Testing System (IELTS)*. Hauptpaue, NY: Barron's Educational Series
- McCarthy, P. & Jarvis, S. (2010). MTL, vocD-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381-392.
- Mellor, A. (2011). Essay length, lexical diversity and automatic essay scoring. *Memoirs of the Osaka Institute of Technology*, 55, 1-14.
- Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth-and eighth-grade students. *Reading and Writing*, 21, 131-151.
- Moll, K. & Landerl, K. (2014). *Salzburger Lese- und Rechtschreibtest (SLRT-II)*. Göttingen: Hogrefe.
- Morrow, L. M. (2014). *Literacy development in the early years: Helping children read and write*. London, UK: Pearson.
- Nussbaumer, M. (2010). *Was Texte sind und wie sie sein sollen*. Berlin: De Gruyter
- Polio, C. & Friedman, D. A. (2017). *Understanding, evaluating, and conducting second language writing research*. New York, NY: Routledge.
- Rodríguez, C., Grünke, M., González-Castro, P., García, T., & Álvarez-García, D. (2015). How do students with Attention Deficit/Hyperactivity Disorders and Writing Learning Disabilities differ from their non-labeled peers in the ability to compose texts? *Learning Disabilities: A Contemporary Journal*, 13, 157-175.
- Rogers, L. A., & Graham, S. (2008). A meta-analysis of single-subject design writing intervention research. *Journal of Educational Psychology*, 100, 879-906.
- Saddler, B. (2013). Best practices in sentence construction skills. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 238-256). New York, NY: Guilford Press.
- Solís, M., Sammacca, N., Barth, A. E. & Roberts, G. J. (2017). Text-based vocabulary intervention training study: Supporting fourth graders with low reading comprehension and learning disabilities. *Learning Disabilities: A Contemporary Journal*, 15, 103-115.
- Spencer, S. L. & Fitzgerald, J. (1993). Validity and structure, coherence, and quality measures in writing. *Journal of Reading Behavior*, 25, 209-231.
- Steinert, J. (2011). *Allgemeiner Deutscher Sprachtest (ADST)*. Göttingen: Hogrefe.
- Stock, C. & Schneider, W. (2008). *Deutscher Rechtschreibtest für das dritte und vierte Schuljahr (DERET 3-4+)*. Göttingen: Hogrefe.
- Sumner, E., Connelly, V. & Barnett, A. L. (2014). The Influence of spelling ability on handwriting production: Children with and without dyslexia. *Journal of Learning Disabilities*, 49, 1-12.
- Thomé, G. & Thomé, D. (2016). *Oldenburger Fehleranalyse für die Klassen 3-9 (OLFA 3-9)*. Oldenburg: Institut für sprachliche Bildung.
- Troia, G. A. (2010). *Instruction and assessment for struggling writers*. New York, NY: Guilford.
- Troia, G. A., & Graham, S. (2003). Effective writing instruction across the grades: What every educational consultant should know. *Journal of Educational and Psychological Consultation*, 14, 75-89.
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17, 65-83.

- Van Steendam, E., Tillema, M., Rijlaarsdam, G., & Van den Bergh, H. (Eds.) (2012). *Measuring writing: Recent insights into theory, methodology and practices*. Leiden, NL: Brill.
- Zareva, A., Schwanenflugel, P. & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency-variable sensitivity. *Studies in Second Language Acquisition*, 27, 567-595.

**Matthias Grünke**

Universität zu Köln  
Department Heilpädagogik &  
Rehabilitation  
Klosterstr. 79b  
matthias.gruenke@uni-koeln.de

Erstmalig eingereicht: 30.08.2017

Überarbeitung eingereicht: 09.11.2017

Angenommen: 13.11.2017