

Empirische Sonderpädagogik, 2018, Nr. 3, S. 294-310
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Schreibkompetenz diagnostizieren: Ein standardisiertes Testverfahren für die Klassenstufen 4-9 in der Entwicklung

*Ann-Kathrin Hennes¹, Barbara Maria Schmidt¹, Sabine Zepnik²,
Markus Linnemann³, Jörg Jost¹, Michael Becker-Mrotzek¹,
Christian Rietz⁴ & Alfred Schabmann¹*

¹Universität zu Köln, ²Pädagogische Hochschule Oberösterreich, ³Universität Koblenz-Landau, ⁴Pädagogische Hochschule Heidelberg

Zusammenfassung

Kinder und Jugendliche mit mangelnder Schreibkompetenz sind in ihrer gesellschaftlichen Teilhabe beeinträchtigt. Daher ist eine möglichst frühe Förderung dieser nötig. Eine valide und reliable Diagnostik ist hierfür unabdingbar. Zurzeit existiert jedoch im deutschen Sprachraum kein Diagnoseverfahren zur Schreibkompetenz, das theoretisch fundiert, psychometrisch abgesichert und in der schulischen Praxis anwendbar ist. Der vorliegende Beitrag will in einem ersten Ansatz zeigen, wie diese Lücke geschlossen werden könnte. Es wird die Entwicklung eines Testverfahrens vorgestellt, welches auf Basis eines Modells des Schreibprozesses entwickelt wurde und verschiedene theoretische Ansätze berücksichtigt. Neuartig ist, dass mit dem Verfahren der Versuch unternommen wird, für den Schreibprozess wichtige Einzelkompetenzen explizit zu testen. Am Beispiel einer Subskala wird die Vorgehensweise erläutert. Erste Ergebnisse zu den psychometrischen Eigenschaften dieser Skala (basierend auf einem Sample von N=541 Kindern und Jugendlichen der 4. bis 9. Klassenstufe) werden berichtet.

Schlüsselwörter: Schreibtest, Schreibkompetenz, Schülertexte, Förderdiagnostik

Measuring writing literacy. The development of a standardized test for grade 4 to 9

Abstract

Children and adolescents with poor writing skills have difficulties participating in society, which is why they need to receive support as early as possible. This requires a valid and reliable diagnostic procedure. However, no theoretically-grounded, psychometrically-sound diagnostic procedure for writing skills that can be applied in school-based practice currently exists for the German-speaking countries. The present research makes an initial attempt to close this gap. We present a testing procedure developed on the basis of a model of the writing process that takes various theoretical approaches into account. What is novel is our attempt to explicitly test individual competences that are important for the writing process. We demonstrate our approach

using one sub-scale as an example, and report initial results on the psychometric properties of this scale (based on a sample of $N = 541$ children and adolescents in Grades 4 through 9).

Keywords: Test of Written-Composition, Writing Literacy, Students' Writings

Einleitung

Die Schreibkompetenz ist eine der wichtigsten Schlüsselkompetenzen zur Teilhabe in der Gesellschaft (Becker-Mrotzek, 2014; MacArthur, Graham & Fitzgerald, 2016). Unter *Schreibkompetenz* wird die Fähigkeit zur Produktion von Texten, die auf eine entsprechende Zielgruppe (Adressat*innen) ausgerichtet sind, verstanden (Knowledge transforming; Becker-Mrotzek & Schindler, 2007; Bereiter & Scardamalia, 1987; Hayes, 2012). Je nach Textfunktion ist die Zielsetzung dabei, entweder präzise zu informieren, überzeugend zu argumentieren oder Sprache ästhetisch ansprechend und kreativ einzusetzen (Harsch, Neumann, Lehmann & Schröder, 2007). Die Schreibkompetenz stellt sowohl im schulischen als auch im beruflichen Kontext eine der wichtigsten Erfolgsdeterminanten dar (Crossley & McNamara, 2016; UNESCO, 2013). Im Schulkontext hilft das Schreiben, neue Kontexte und Lernobjekte strukturierend zu vernetzen (Linnemann & Stephany, 2014), wodurch Lerninhalte besser verinnerlicht werden (Farnan & Fearn, 2008; Pohl & Steinhoff, 2010; Shanahan, 2006a). Zudem erfordern Leistungsüberprüfungen häufig die Produktion eines Textes (MacArthur et al., 2016). Auch im Beruf erfolgt ein großer Teil der Kommunikation schriftlich (z.B. Kunden-Mails, Briefwechsel, Anweisungen in schriftlicher Form etc.; MacArthur et al., 2016). Dies gilt auch für Berufe, die keine höhere Schulausbildung erfordern (Mikulecky & Kirkley, 1998) und auch im Feld der Sonderpädagogik. So zeigten etwa Bach, Schmidt, Schabmann und van Kroll (2016), dass ein bis drei Viertel der in Ausbildung befindlichen Personen mit Lernstörungen angehalten sind, in ihrem Beruf zumindest

kurze Texte zu verfassen (siehe auch: Anforderungskatalog der Bundesagentur für Arbeit, 2006).

Aus den genannten Gründen sollte die Vermittlung und Förderung der Schreibkompetenz Basisziel aller schulischen Bildungsgänge sein. Lehrpersonen sollten Schreibkompetenzen zielgerichtet vermitteln können. Darüber hinaus sollten Schüler*innen mit Schwierigkeiten beim Schreiben durch den Einsatz von evidenzbasierten Fördermaßnahmen unterstützt werden, da nur so eine langfristige Benachteiligung verhindert werden kann. Voraussetzung dafür ist, dass Lehrpersonen in der Lage sind, die Schreibkompetenzen ihrer Schüler*innen förderorientiert und adäquat zu beurteilen. Dies wiederum setzt einheitliche Bewertungsmaßstäbe voraus, die im schulischen Kontext bislang nicht existieren (Hennes, Schmidt, Schabmann & Rietz, 2017). Zumeist wird die Schreibkompetenz in der Schule anhand eines längeren Textes erhoben und global (als Gesamturteil über den Text) bewertet. Schreibkompetenz wird also über die Messung der Textqualität bestimmt (Feenstra, 2014; Philipp, 2015).

Dieses Vorgehen ist jedoch aus zwei Gründen nicht unproblematisch. Erstens zeigen Studien, dass globale Bewertungen durch Lehrpersonen meist durch Vergleichsprozesse zwischen den zu bewertenden Texten einer Gruppe (Klasse) zustande kommen, die den Referenzrahmen für die Bewertung darstellt (Madelain & Wheldall, 2005; Pohlmann & Möller 2007; Schrader, 2006). Dies führt zu einer geringen Konsistenz, sobald dieser Bezugsrahmen experimentell aufgehoben wird und Lehrkräfte „fremde“ Texte beurteilen sollen. Urteile zu ein und demselben Text unterscheiden sich unter diesen Bedingungen stark voneinan-

der (Birkel & Birkel, 2002). Somit ist eine eindeutige und objektive Identifikation von Schüler*innen mit Förderbedarf nicht möglich. Die/der gleiche Schüler*in kann bei einer Lehrkraft als „förderbedürftig“ gelten, bei der anderen nicht.

Ein weiteres Problem der beschriebenen Vorgangsweise ist, dass die Bewertung der Schreibkompetenz auf Basis des Schreibprodukts nur bedingt Aussagen darüber erlaubt, welche speziellen Teilkompetenzen des Schreibens zu einem nicht adäquaten Text im Sinne der obigen Definition geführt haben. Zum Beispiel bleibt die Frage unbeantwortet, ob ein(e) Schreiber*in beim Verfassen eines Textes (bspw. Anleitungstext) nicht in der Lage war, wichtige Informationen als solche zu identifizieren, oder ob es ihr/ihm nicht gelungen ist, sich in die Rolle der unwissenden Lesenden hineinzuversetzen, oder ob das Problem darin besteht, dass er/sie die Aufgabenstellung nicht verstanden hat und so keine entsprechende Zielsetzung vornehmen konnte. In allen Fällen wird das Geschriebene nicht nachvollziehbar (funktionsuntüchtig) sein. Aus diesem Grund ist es für eine förderorientierte Diagnostik notwendig, neben dem „Gesamtbild“ des Textes die unterschiedlichen Komponenten der Schreibkompetenz selbst zu messen und beurteilen zu können.

Erste Ansatzpunkte zur adäquaten Erfassung der Schreibkompetenz bieten standardisierte Testverfahren. Aus der internationalen Literatur sind verschiedene Verfahren bekannt. Häufig sind sie in globalere Schulleistungstests eingebettet (wie z.B. im Kaufmann Test of Educational Achievement; Kaufman & Kaufman, 2005) und erheben Schreibkompetenz zusätzlich zu weiteren Fähigkeiten wie Lesen und Rechnen. Tatsächlich erfasst werden im Rahmen dieser Verfahren jedoch lediglich die Rechtschreibung und die Interpunktion, zum Teil auch graphomotorische Fähigkeiten. Die Textproduktion spielt im Rahmen dieser Verfahren keine Rolle. Einen Schritt in die Richtung der tatsächlichen Beurteilung einzelner Teilkomponenten der Schreibkompetenz stellt

der Test of Written Language (TOWL; Hamill & Larsen, 2009) dar. Im TOWL werden Aufgaben zu Wortschatz, Rechtschreibung, Interpunktion sowie zur Herstellung von Kohäsion/lokaler Kohärenz als Ergänzung zu einer Langtextaufgabe vorgegeben. Der TOWL misst somit zum Großteil nur sehr basale Fähigkeiten des Schreibens, die eher bei Schüler*innen im Grundschulalter relevant sind, die über eine geringe Schreiberefahrung verfügen (z.B. Bereiter & Scardamalia, 1987; Graham & Harris, 2010; Sturm & Weder, 2016). Außerdem bleiben die Subtests zu den einzelnen Dimensionen der Schreibkompetenz im TOWL alle auf Wort- oder Satzebene. Hierarchiehöhere Fähigkeiten, die zur Etablierung globaler Textkohärenz nötig sind, werden nur marginal bei der Langtextaufgabe anhand von textmusertypischen Vorgaben erfasst bzw. bewertet.

Das Forschungsvorhaben

In diesem Aufsatz wird ein neues Verfahren zur Messung der Schreibkompetenz zur Diskussion gestellt. Präsentiert werden grundlegende Ansätze zur Entwicklung eines (förderdiagnostischen) Verfahrens, das den genannten Bedarfen zur Diagnostik der Schreibkompetenz gerecht werden soll. Grundidee des Instruments ist es, modellorientiert Teilkomponenten zur Messung der Schreibkompetenz heranzuziehen. Das Verfahren bezieht sich schwerpunktmäßig auf informierende Texte, da diese entsprechend den Lehrplänen in der adressierten Altersgruppe (4. – 9. Jahrgangsstufe) bekannte Textsorten beinhalten, die über die Spanne der Jahrgangsstufen hinweg durchgängig relevant sind (Becker-Mrotzek & Böttcher, 2012). Die Altersgruppe wurde gewählt, da nach den Modellen zur Entwicklung der Schreibkompetenz ab dem 4. Schuljahr davon auszugehen ist, dass basale Fähigkeiten grundlegend erworben sind und Schreiber*innen die kommunikative

Funktion der Textproduktion funktional nutzen können (Bereiter, 1980; Bereiter & Scardamalia, 1987). Zusätzlich stellt der Wechsel von der Grundschule in die weiterführenden Schulen einen wichtigen Übergang in der schulischen Entwicklung dar. In weiterführenden Schulen werden schriftliche Prüfungsformate zunehmend relevant, und die Schreibkompetenz spielt in verschiedenen Schulfächern eine entscheidende Rolle. Zudem stellt das Ende der Sekundarstufe I für einen Teil der Schüler*innen das Ende der schulischen Laufbahn dar. Speziell im sonderpädagogischen Feld könnte eine Erhebung der Textkompetenz Hinweise für eine nötige (zusätzliche) Förderung geben.

Das Basismodell

In der Forschung wurden verschiedene Modellvorstellungen zum Schreiben entwickelt. Grob voneinander unterscheiden kann man Entwicklungsmodelle, in denen Phasen der Schreibentwicklung postuliert werden (z.B. Bereiter & Scardamalia, 1987), Prozessmodelle und kognitive Modelle (Alamargot & Chanquoy, 2001; Becker-Mrotzek & Böttcher, 2006). Prozessmodelle beschreiben den Prozess der Textentstehung selber, während kognitive Modelle den/die Schreiber*in und deren/dessen Kognitionen in den Mittelpunkt stellen (Jakobs & Perrin, 2014). Die beiden letzten sind für die vorliegende Arbeit von besonderer Relevanz, da die (kognitiven) Komponenten des Schreibprozesses möglichst differenziert erfasst werden sollen.

Das hier vorgestellte Basismodell für die Textentwicklung beruht auf dem Prozessmodell von Hayes und Flower (1980) und dem darauf aufbauenden kognitiven Modell von Hayes (2012). Das hieraus entstandene Basismodell für die Textkonstruktion haben wir mit dem Begriff *Konvergenzmodell* (KM) bezeichnet, da es die Über-/Bearbeitung von vorläufigen Texten oder Schreibideen hin zu einem „optimalen“ Text in den Mittelpunkt rückt.

Im KM wird der Schreibprozess als eine iterative Abfolge von Planung und Übersetzung (das eigentliche Schreiben) gesehen. „Triebfeder“ dieses Prozesses ist die ständige Über-/Bearbeitung des Plans und/oder der Übersetzung auf Basis des (noch nicht perfekten) bisherigen Textes (Dam-Jensen & Heine, 2013). Planung, Übersetzung und Über-/Bearbeitung werden von drei kognitiven Instanzen auf einer Metaebene kontrolliert und koordiniert. Die Benennungen und die Funktionsweisen dieser Instanzen wurden aus dem Modell von Hayes (1980, 2012) übernommen.

Die oberste Instanz ist der *Evaluator*, der zwei ihm untergeordnete Instanzen (*Proposer* und *Translator*) überwacht und diese hinsichtlich der Gesamtzielsetzung des Textes steuert. Der *Proposer* ist die primär kreative Instanz im Schreibprozess. Er liefert einen im Idealfall auf Ziel und Adressat abgestimmten „Vorschlag“ für den Inhalt und die Grobstruktur des Textes (als vorläufiger Schreibplan). Hierzu berücksichtigt er Informationen (z.B. über den Rezipienten) aus dem Langzeitgedächtnis der Schreibenden/des Schreibenden und aus der Schreibumgebung (Hayes & Flower, 1980). Der entstandene Schreibplan kann in dieser ersten Arbeitsschleife durchaus noch unscharf („fuzzy“; vgl. Dam-Jensen & Heine, 2013) sein (z.B. erste unklare Ideen, Halbsätze etc.). Beim Schreiben kann dann immer wieder auf den *Proposer* zurückgegriffen werden, was eine ständige Konkretisierung des vorläufigen Schreibplans nach sich zieht.

Der *Translator* steuert den Übersetzungsprozess. Er übersetzt die Elemente des Plans in konventionelle (Schrift-)Sprache. So entsteht ein (vorläufiger) Text, der sichtbares Ergebnis der Planung und konkrete Basis für eine (erste oder weitere) Überarbeitungsschleife(n) ist.

Der *Evaluator* überprüft die vorläufigen Schreibprozesse und initiiert gegebenenfalls die Überarbeitung (des Plans oder des Textes). Änderungen stehen in Wechselwirkung: Veränderte Textteile können neue Pla-

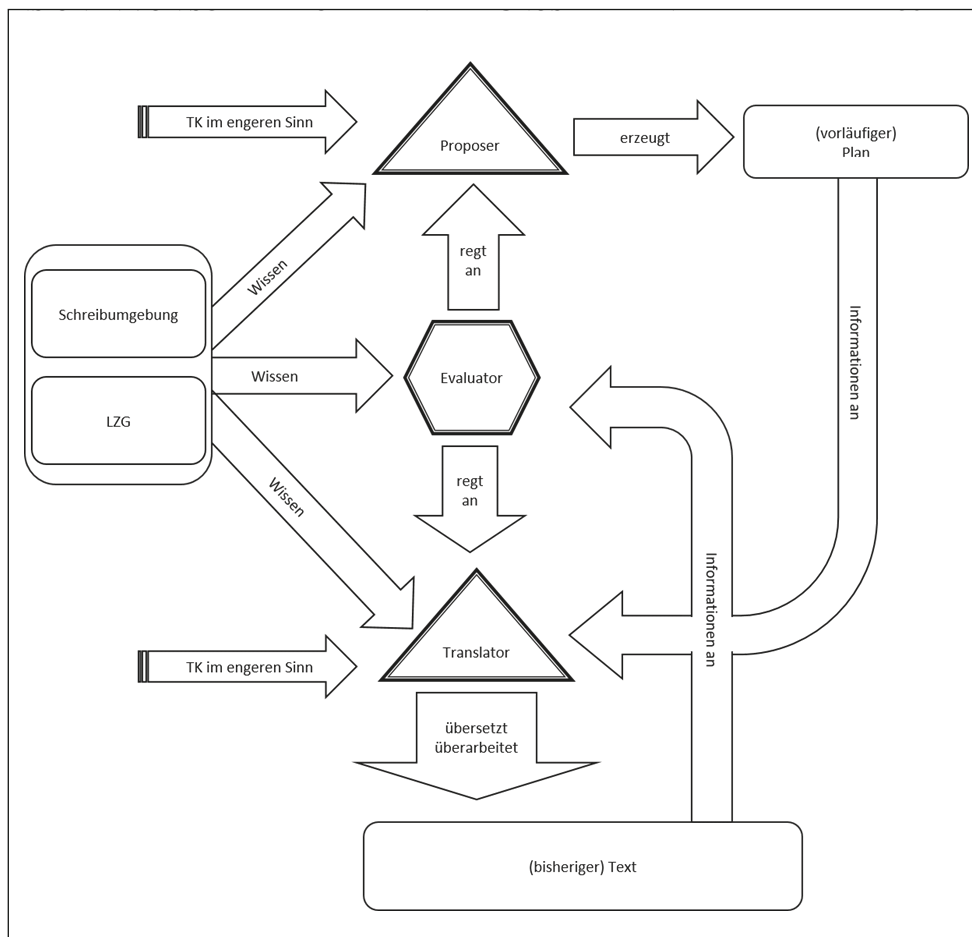


Abbildung 1: Konvergenzmodell: Theoretisches Basismodell der Textkonstruktion

nungsprozesse nach sich ziehen und umgekehrt (MacArthur & Graham, 2016). Abgeschlossen ist der Iterationsprozess dann, wenn die/der Schreiber*in keine Veränderungsnotwendigkeit mehr sieht (wobei sich der/die Schreiber*in auch mit einer vorläufigen Fassung zufriedengeben kann).

Textkompetenz im engeren Sinn: Die Umwandlung von Ideen in Text erfolgt auf Basis einzelner (kleinteiliger) Arbeitsschritte. Ergänzend zu den Komponenten des Schreibprozessmodells und des kognitiven Modells finden sich im KM daher zusätzlich konkrete Kompetenzen, auf die die/der Schreiber*in zurückgreift, um einen funktionsstüchtigen Text zu produzieren. Wir be-

zeichnen sie vorläufig als Textkompetenz im engeren Sinn (TK), im Gegensatz zur Schreibkompetenz, die die abstrakten Prozessvariablen umfasst (s. Abb. 1). Auf der Ebene der TK erhebt das Modell (noch) keinen Anspruch auf Vollständigkeit. Es stellt jedoch einen ersten Versuch dar, in der Literatur (Becker-Mrotzek, 2014; Becker-Mrotzek, Grabowski, Joost, Knopp & Linnemann, 2014; Becker-Mrotzek & Schindler, 2007; Knopp, Joost, Nachtwei, Becker-Mrotzek & Grabowski, 2012; Knopp, Becker-Mrotzek & Grabowski, 2013; Knopp, Jost, Linnemann & Becker-Mrotzek 2014; Kruse, Reichardt, Herrmann, Heinzel & Lipowsky, 2012; Nussbaumer, 1991) häufig diskutierte

Komponenten der Schreibkompetenz auf Textebene in das Modell einzuordnen. In diesem Artikel beschränken wir uns bei der Darstellung darauf, die Teilkomponenten im Kontext informierender Texte zu definieren. Diese Einschränkung liegt darin begründet, dass sich das Forschungsvorhaben der Testentwicklung schwerpunktmäßig auf diese Textgattung bezieht. Im Modell werden unter TK folgende Kompetenzen verstanden:

- *Adressat*innenorientierung*: Aufgrund der räumlichen und zeitlichen Trennung von Produzent und Rezipient (zerdehnte Kommunikation; Ehlich, 1983) müssen Texte optimal auf die Bedürfnisse der/ des Rezipient*innen ausgerichtet sein (Becker-Mrotzek & Schindler, 2007). Kompetente Schreiber*innen müssen in der Lage sein, den Text so zu gestalten, dass er dem Informationsstand, den kognitiven sowie den sprachlichen Fähigkeiten der/des Adressat*innen entspricht (Becker-Mrotzek et al., 2014). Der Adressat*innenorientierung kommt im Modell eine übergeordnete Funktion als Grundlage der weiteren Komponenten zu (Becker-Mrotzek & Schindler, 2007).
- *Informationslogistik*: Hierunter wird die Fähigkeit zur Auswahl und Ordnung von Informationen verstanden¹. Nur ein Teil aller möglichen im Proposer generierten Informationen ist für den jeweiligen Rezipienten relevant (Hayes & Flower, 1980), es muss also ein Auswahlprozess stattfinden, und die Informationen müssen in einer sinnvollen Reihenfolge angeführt werden (z.B. Schnotz, 2006).
- *Textmusterorientierung*: Übersetzungsprozesse, die entlang festgelegter Konventionen (bspw. typische Gliederungen oder Formulierungen) verlaufen, führen i.A. zu leichter verständlichen Texten als solche, die diese sog. Textmuster nicht berücksichtigen (Heinemann, 2000; Knopp et al., 2014; Sandig, 1997). Das

Wissen um solche Textmuster sowie deren Ausnutzung beim Schreiben ist daher als Teilkomponente in das KM aufgenommen.

- *Kohärenzstiftung*: Textübergreifend zeigen sich gelungene Planungs- und Übersetzungsprozesse darin, dass Schreibende Kohärenz etablieren, d.h. einen schlüssigen Text produzieren. Hierzu können Schreiber*innen verschiedene kohärenzstiftende Mittel einsetzen. Man unterscheidet lokale Kohärenz auf Satzebene (Verwendung von passenden Konnektoren, Rekurrenz oder Referenzen) von globaler Kohärenz auf Ebene des Gesamttextes im Sinne semantischer Verknüpfungen oder globaler Strukturierungsmerkmale (Überschriften, Einrückungen etc.; z.B. Averintseva-Klisch, 2013).

Umsetzung im Test

Aus dem KM wird klar, dass Textproduktion auf verschiedenen Ebenen erfasst werden kann, nämlich auf der Ebene der Schreibkompetenz im Sinne von Hayes und Flower (1980) bzw. Hayes (2012), sowie auf der Ebene der im Modell ausgewiesenen TK im engeren Sinn. Im angestrebten Verfahren steht der Zugang über die TK im engeren Sinne im Fokus. Dieser Ansatz erscheint deshalb geeignet, weil über die TK ein klarer Rahmen für die Diagnostik und Förderung gegeben ist. So können förderdiagnostische Fragestellungen präzise und Interventionen konkret abgestimmt werden (z.B. etwa das Einüben von Konnektoren, die Berücksichtigung des Rezipienten etc.). Bislang wurden folgende Aufgaben zur Messung der TK entwickelt (Tabelle 1).

Wie in Tabelle 1 erkennbar, werden verschiedene erhebungsmethodische Zugänge genutzt, um die unterschiedlichen Facetten der Teildimensionen der TK zu erfassen. Das Verfahren erhebt vorderhand nicht den

¹ Der Begriff "Information" wird hier synonym mit dem ursprünglichen Begriff der "Idee" (Hayes & Flower, 1980) verwendet. Ideen lassen sich jedoch letztlich als eine geordnete (nicht notwendigerweise lineare) Abfolge von Informationen beschreiben.

Tabelle 1: Bisher entwickelte Aufgaben des Schreibtests

Teildimension	Aufgabe(n)	Gemessene Teilkompetenz
Adressatenorientierung	Bauanleitung für eine Person mit verbundenen Augen schreiben	Berücksichtigung der Situation des Adressat*innen (z.B. keine Farben benennen)
Informationslogistik	Aus einem Set an vorgegebenen Informationen die jeweils wichtigste Information identifizieren	Hierarchisierung der Information, Auswahl anhand der Relevanz
	Bauanleitung schreiben	Strukturierung von Informationen (unter Berücksichtigung eines unwissenden Adressaten).
	Unwichtige Hypertextverknüpfungen streichen, wichtige erkennen	Unterscheidung von wichtigen und unwichtigen Informationen
Kohärenzstiftung	Auswahl von passenden Konnektoren (Lückensatz)	Einsatz von Konnektoren auf Satzebene (rezeptiv)
	Freies Einsetzen von möglichen Konnektoren in den gleichen Lückensatz	Analog oben (produktiv)
	Texte mit mangelnder Kohärenz erkennen	Herstellung von Kohärenz bzw. Erkennen von Inkohärenzen (trotz augenscheinlicher Kohäsion auf Textoberfläche)
	Lücke in einem Text auf Basis vorgegebener Möglichkeiten kohärent schließen	Semantische Kohärenz auf Textebene produktiv herstellen
	Lücke in einem Text ohne vorgegebene Möglichkeiten kohärent zu schließen	Semantische Kohärenz auf Textebene produktiv herstellen
	„Stilblüten“ erkennen und durch die Neuordnung der vorliegenden Sätze auflösen	Kompetenz im Erkennen von (falschen) Bezügen
	Falsche Bezüge erkennen und durch am Text vorgenommene Umformulierungen auflösen	Analog oben (stärker produktiv)
	In einem längeren Fließtext Anaphern/Referenzen zu einem zu Beginn benannten Referenten zu identifizieren	Kompetenz im Umgang mit Anaphern/Referenzen
Textmusterorientierung	Textbausteine dem passenden Textmuster zuordnen	Kompetenz im Umgang mit testmusterspezifischen Formulierungen
	Sprachliche Formulierungen entsprechend Textmuster	Analog oben (produktiv)

Anspruch, die einzelnen Dimensionen völlig überschneidungsfrei und vollständig zu erfassen. Komponenten, die entwicklungsbedingt als Voraussetzung für die Schreibkompetenz angesehen werden können (z.B. Rechtschreibung oder kognitive Faktoren), werden nicht erfasst. Für sie existieren bereits entsprechende psychometrische Verfahren.

Erste Analysen zu den psychometrischen Eigenschaften einer Skala zur Dimension Informationslogistik (Subkategorie adäquate Informationsdichte)

Im Folgenden wird beispielhaft an einer Skala zur Informationslogistik (adäquate Informationsdichte) dargestellt, wie die Einzelaufgaben grundsätzlich konstruiert sind.

Die Skala soll messen, inwiefern die Testperson (TP) in der Lage ist, vollständig Informationen zu geben. Hierbei besteht die Aufgabe darin, für unwissende Adressat*innen eine Bauanleitung zu einem Gebäude aus handelsüblichen Bauklötzen zu schreiben. Insgesamt gibt es fünf Einzelaufgaben, wobei sich die Anzahl der Bausteine und die Komplexität der Gebäude steigern (für eine genaue Beschreibung siehe Abschnitt „Methode“). Es werden erste Ergebnisse zur internen Konsistenz der Skala, zur Konstruktvalidität sowie zur Entwicklung des erreichten Testscores über verschiedene Klassenstufen berichtet. Es bestehen folgende Erwartungen:

- 1) Interne Konsistenz und faktorielle Validität: Angesichts der Kürze der Skalen wird ein (zufriedenstellendes) Cronbachs α von $\sim .80$ erwartet. Die ansteigende Komplexität der Bausteingebäude sollte sich im Testscore monoton abbilden. Zudem wird erwartet, dass sich die Aufgaben durch eine gemeinsame latente Dimension („Informationsdichte“) erklären lassen.
- 2) Konvergente und divergente Validität: Wichtig war es, die Aufgaben so zu konstruieren, dass die Raumvorstellung der TP bei der Bearbeitung keine Rolle spielt. Es wurde darauf geachtet, dass die rechts-links Orientierung der Steine irrelevant war. Zudem waren alle Steine während der Bearbeitung der Schreibaufgabe verfügbar. Entsprechend wird eine Nullkorrelation zwischen dem Testscore der Aufgabe und der Raumvorstellung erwartet. Hingegen sind Korrelationen von Schreibaufgaben mit dem basalen Lesen (Ahmed, Wagner & Lopez, 2014), dem Wortschatz (Castillo & Tolchinsky, 2018) dem Leseverständnis (Ahmed, Wagner & Lopez, 2014; Berninger et al., 2002; Shanahan, 2006b), grammatischen Fähigkeiten (Drijbooms, Groen & Verhoeven, 2015), dem Arbeitsgedächtnis (Drijbooms, Groen & Verhoeven, 2017; Kellog, 1996; Olive, Kellog & Piolat, 2008) und der Exekutiv-

funktion (Drijbooms et al., 2015; Drijbooms et al., 2017; Olive et al., 2008) aus der Literatur bekannt. Allerdings werden angesichts der Spezifität der Aufgabe zur Informationsdichte nur maximal mittlere Korrelationen zwischen dem Testscore und den genannten Variablen erwartet. Die Korrelation zur Intelligenz kann aufgrund der im theoretischen Teil beschriebenen Anforderungen (z.B. Planungsprozesse) an Schreiber*innen nicht ausgeschlossen werden. Es werden daher mittlere Korrelationen zwischen dem Testscore und der Intelligenz erwartet.

- 3) Abhängigkeit des Testscores von der Klassenstufe: Zu erwarten ist ein Anstieg des Testscores in allen Schwierigkeitsstufen über die Klassenstufen (entsprechend Bereiter & Scardamalia, 1987).

Methode

Stichprobe

Die Stichprobe besteht aus insgesamt 541 Kindern und Jugendlichen (49.4 % männlich) der 4. bis 9. Klassenstufe aus Grund- ($N = 41$) und Gesamtschulen in Köln und dem Umland von Köln. In der Gesamtgruppe aller 541 Schüler*innen sprechen 75.5% Deutsch als Muttersprache. Die ausgewählten Schulen waren bezüglich sozio-ökonomischer Bedingungen möglichst homogen; es handelte sich um Schulen in innerstädtischen und ländlichen Randregionen von Köln. Eine Kontrolle der sozio-ökonomischen Bedingungen war aus Datenschutzgründen in dieser ersten Untersuchung nicht möglich. Die Erhebung fand von April 2016 bis Juni 2017 in ruhigen Räumen an den Schulen in Gruppentestungen statt.

Instrumente

Die Aufgabe „Verrückte Gebäude“ (*Informationsdichte*): Zwei Aspekte waren bei der Konstruktion der Aufgabe wesentlich. Erstens sollten Unterschiede im Vorwissen der Testpersonen (TP) ausgeglichen werden. Dies ist nötig, da entsprechend dem KM der Proposer auf Vorwissen aus dem Langzeitgedächtnis (LZG) des Schreibers/der Schreiberin zugreift und so das Vorwissen die Messung der Schreibkompetenz beeinflusst. Das sachliche Vorwissen wurde daher a priori im Test vorgegebenen (s. Testaufgaben). Außerdem wurde die Aufgabe so konstruiert, dass der Adressatenkreis des zu verfassenden Textes eindeutig benannt wurde und das Wissen der adressierten Person explizit beschrieben wurde. Dieser Schritt war notwendig, da die Informationen, die Schreibende geben müssen, stark vom Adressatenkreis des Textes abhängen.

Testaufgaben. Konkret sollte die TP einen instruktiven Text zum Bau von „Gebäuden“ aus handelsüblichen Bausteinen verfassen (Abb. 2). Der TP lagen hierzu eine Abbildung des Bausteingebäudes in Form eines Fotos sowie ein Set der für das jeweilige Gebäude verwendeten Bausteine vor. Folgende Instruktion wurde gegeben:

„Du siehst auf der Abbildung ein Gebäude aus Bausteinen.

- 1) Schau dir das Bild genau an. Baue das Gebäude nach.
- 2) Überprüfe, ob du alles richtig gebaut hast. Achte besonders auf diese Stelle: (Markierung der explizit zu beschreibenden Stelle im Gebäude; siehe unten)
- 3) Ein Freund soll das Gebäude nachbauen. Er hat das Bild nicht. Schreibe ihm eine genaue Anleitung dazu, wie er das Gebäude auf dem Bild nachbauen kann!“

Für die Aufgabe „verrückte Gebäude“ wurde ein Set von fünf Bausteingebäuden mit unterschiedlicher Komplexität konstruiert. Die Konstruktion der Gebäude erfolgte anhand folgender Prinzipien:

- Jedes Gebäude besteht aus Steinen, die hinsichtlich Farbe und Form nur gemeinsam variieren. So besteht für die Schreibenden die Möglichkeit, die Steine anhand ihrer Farbe zu benennen. Geometrische Bezeichnungen müssen nicht bekannt sein. Die Relevanz des (mathematischen) Vorwissens wird so reduziert.
- Die Komplexität der Bausteingebäude nimmt mit jeder Schwierigkeitsstufe zu. Pro Stufe wird ein Baustein mehr verbaut. Dieser zusätzliche Stein erforderte beim Schreiben eine oder mehrere neue Instruktionen in Bezug auf seine Lage und Position. Mit zunehmender Komplexität erhöht sich demnach die adäquate Informationsdichte.
- Daten aus Pilotversuchen in verschiedenen Jahrgangsstufen (4-9) zeigten, dass die TP sehr viel implizites Wissen des Adressaten voraussetzten und nur sehr grobe Informationen verschriftlichten. Aus diesem Grund enthält in der vorliegenden Version jedes Gebäude mindestens ein besonderes Element (kritisches Element), dessen Position oder die Ausrichtung im Text detaillierter als bei anderen Elementen beschrieben werden müssen (z.B. der hochkant gestellte Stein in Abbildung 2).
- Für die Fertigstellung der Aufgaben gab es keine Zeitbeschränkung.

Scoring: Für die Auswertung ist die Frage leitend, ob die Informationen, die für den Nachbau notwendig sind, von der TP vollständig gegeben wurden. Allerdings ist es schwer, die notwendige Informationsdichte a priori zu bestimmen, zumal manche Informationen durchaus verzichtbar sind (z.B. „Stelle den Stein hochkant auf die Tischplatte“ – „hochkant“ und „Tischplatte“ möglicherweise verzichtbar), andere müssen unabdingbar genannt werden. Aus diesem Grund wurde empirisch ein Kriterienkatalog entwickelt, der für jede Schwierigkeitsstufe die unbedingt notwendigen Informationen vorgibt: Versuchspersonen (VP)

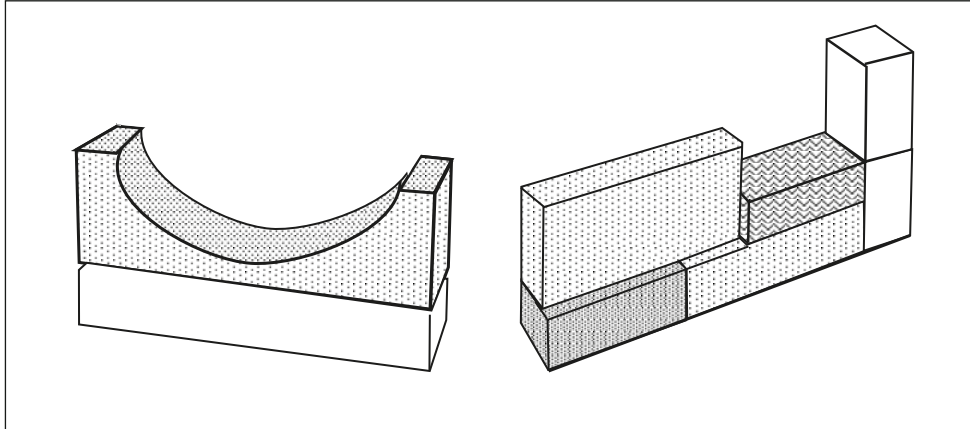


Abbildung. 2 Beispiel einer Figur der niedrigsten (links) und der höchsten Komplexitätsstufe (Abbildungen im Original farbig). Die linke Figur wird entsprechend ihrem „Weltwissen“ häufig als „Brücke“ bezeichnet. Beide Figuren enthalten explizit zu beschreibende Elemente (links: „Brücke“ mit den Pfeilern nach oben; rechts hochkant stehender Stein).

(Studierende der Universität zu Köln, $N=193$, durchschnittliches Alter 22 Jahre, 85.5% weiblich) wurden Bauanleitungen vorgegeben, wobei die hierin enthaltene Informationsdichte durch Weglassen von Informationen variiert wurde. Es wurde dann jeweils geprüft, ob der Nachbau des Bausteingebäudes trotz geringerer Informationsdichte noch möglich war. Kriterium war, dass 80 Prozent der VP das Gebäude anhand der gegebenen Informationen nachbauen konnten. Die Anzahl der als notwendig identifizierten Informationen variierte zwischen den Komplexitätsstufen. Sie nahm, wie erwartet, mit steigender Komplexität zu. Auf Basis dieses Kriterienkatalogs wurde bei der späteren Auswertung der von den Schüler*innen verfassten Texte ein Punkt für jede korrekt gegebene Information vergeben und der Quotient aus von der TP gegebenen und allen nötigen Informationen als Maß für die Informationsdichte bestimmt.

Zusätzliche Instrumente: Die *sprachfreie Intelligenz* wurde mit dem Grundintelligenztest (CFT 20-R; Weiß, 2006) erhoben; die *Raumvorstellung* mit dem Mental Rotation Test (MRT; Peters, Laeng, Latham, Jackson, Zaiyouna & Richardson, 1995). Das

basale Lesen wurde mit dem Salzburger Lesescreening (SLS 2-9; Wimmer & Mayringer, 2014) erfasst, der *Wortschatz* mit dem Wortschatz-Ergänzungstest aus dem CFT 20-R (WS-R) und über drei Subtests (Synonyme, Antonyme und Wortendungen) des Allgemeinen Deutschen Sprachtests (ADST; Steinert, 2011). Das *Leseverständnis* mit dem Lesegeschwindigkeits- und -verständnistest (LGVT 5-12+; Schneider, Schlagmüller & Ennemoser, 2017). Zur Messung des (*phonologischen*) *Arbeitsgedächtnisses* wurde ein Test in Anlehnung an den Reading Span Test (RST; Carroll et al., 2015) konstruiert. Hier hatten die TP die Aufgabe, eine Serie von Sätzen nacheinander zu lesen. Unmittelbar nach jedem Satz sollten die TP beurteilen, ob der jeweilige Satz eine richtige oder eine falsche Aussage trifft. Zusätzlich war es Aufgabe der TP, sich das jeweils letzte Wort des Satzes zu merken. Unmittelbar im Anschluss an eine präsentierte Satzserie (umfasst zwischen 2 bis 5 Sätze) sollten die TP dann die jeweils letzten Wörter der präsentierten Sätze entsprechend der vorgegebenen Reihenfolge schriftlich reproduzieren. Zur Messung des (*räumlich-visuellen*) *Arbeitsgedächtnisses* wurden zwei Aufgaben in Anlehnung an Hasselhorn et

al. (2012) eingesetzt: Im *Matrizentest* wurden auf einem 4 x 4 Raster ein schwarz-weißes Muster mit ansteigender Komplexität für 2-7 Sekunden (je nach Schwierigkeit) präsentiert. Die TP mussten dieses Muster nachzeichnen. Bei der *Corsiblock-Aufgabe* wurden rasch nacheinander Smileys in einer unsystematischen Reihenfolge auf einer 9 x 9 Matrix präsentiert, die von den TP reproduziert werden musste. Zur Erfassung der zentralen Exekutive wurde der Star Counting Test (SCT; De Jong & Das-Smaal, 1990) durchgeführt.

Statistische Analyse: Zur Bestimmung der internen Konsistenz wurde Cronbachs' α berechnet, zur Bestimmung der konvergenten und divergenten Validität Korrelationen mit den zusätzlich erhobenen Variablen, für die Entwicklung über die Zeit wurde eine Varianzanalyse mit Messwiederholung durchgeführt. Die faktorielle Validität wurde mittels einer konfirmatorischen Faktorenanalyse überprüft, wobei neben ML auch Bayes-Schätzverfahren zum Einsatz kamen.

Ergebnisse

Interne Konsistenz: Die relative Informationsdichte der Texte zu den fünf vorgegebenen Gebäuden ist sehr gut durch eine Dimension beschreibbar. Das entsprechen-

de Modell ist gut angepasst ($\chi^2_{(5)} = 5.35$; $p = .08$; $RMSEA = .04$; $CFI = .99$; $NFI = .98$; $PPP = .49$; Tab. 2). Cronbachs Alpha beträgt $\alpha = .79$. Die Aufgabenschwierigkeit steigt monoton mit den a priori festgelegten Schwierigkeitsstufen ($F_{(1, 505)} = 1250.8$; $p < .001$, $\eta_p^2 = .71$).

Konstruktvalidität / divergente und konvergente Validität: Wie erwartet korreliert der Gesamtscore nicht mit der Raumvorstellung (MRT; $r = .07$; $p > .05$). Die Korrelationen mit basalem Lesen (SLS; $r = .40$; $p < .01$), Wortschatz (ADST Subskalen + WS-R; $r = .50$; $p < .01$), Grammatik (ADST Subskala; $r = .45$; $p < .01$), Leseverständnis (LGVT; $r = .41$; $p < .01$), Arbeitsgedächtnis (Faktorenscore aus Corsiblock, Matrix und RST; $r = .03$; $p > .05$) und zentraler Exekutive (SCT; $r = .16$; $p < .01$) sind von maximal mittlerer Größe. Gleiches gilt für die Intelligenz (CFT; $r = .42$; $p < .01$).

Abhängigkeit der gegebenen Informationsdichte von der Klassenstufe: Erwartet wurde ein Ansteigen der Testscores über die Klassenstufen. Eine ANOVA mit dem Messwiederholungsfaktor Klassenstufe ergibt einen signifikanten Effekt ($F_{(5, 536)} = 25.5$; $p < .001$, $\eta_p^2 = .19$). Es kommt zu einem Anstieg des Informationsgehalts der Instruktionen in den höheren Klassenstufen. Post-hoc Analysen zeigen allerdings, dass sich nur die Klassenstufen 4 von 5-7 und 8-9, sowie 5-7 von 8-9 signifikant unterscheiden (Tab 3).

Tabelle 2: Ergebnisse einer konfirmatorischen Faktorenanalyse zur Dimensionalität der Aufgabe „Verrückte Gebäude“

	Schätzung (S.E.)	95% CI	Schiefe	Exzess	stand. ML-Schätzung	rel. ID
Gebäude 1	1				.53	.82
Gebäude 2	1.43 (0.006)	[1.20, 1.75]	0.63	0.88	.73	.59
Gebäude 3	1.03 (0.004)	[0.84, 1.28]	0.72	1.47	.60	.57
Gebäude 4	1.30 (0.006)	[1.07, 1.56]	0.73	1.06	.72	.50
Gebäude 5	1.16 (0.005)	[0.95, 1.42]	0.55	0.61	.68	.43

Verwendung von Bayes und ML-Schätzer. Letzte Spalte: Mittelwert der relativen Informationsdichte (rel. ID) für die jeweilige Aufgabe

Tabelle 3: Unterschiede in der Aufgabe „Verrückte Gebäude“ in Abhängigkeit von der Klassenstufe

Jahrgang	N	M	SE	95%-Konfidenzintervall	
				Untergrenze	Obergrenze
4 ^a	41	.36	.02	.31	.41
5 ^b	130	.51	.01	.49	.54
6 ^b	153	.53	.01	.50	.55
7 ^b	117	.56	.01	.53	.59
8 ^c	66	.63	.02	.60	.67
9 ^c	35	.71	.03	.65	.76

gleiche Indizi bezeichnen homogene Subgruppen bei $\alpha = 5\%$

Diskussion

Die vorliegende Arbeit beschreibt die Entwicklung eines förderdiagnostisch orientierten Verfahrens zur Erfassung der Text(produktions)kompetenz für den deutschen Sprachraum. Der Ansatz verbindet zwei wichtige theoretische Modelle der Schreibforschung (Hayes & Flower, 1980; Hayes 2012) und ergänzt sie um Aspekte der Textkompetenz (im engeren Sinne). Hierbei handelt es sich um Fähigkeiten, die für konkrete Schreibaufgaben relevant sind.

Der Ansatz löst sich von dem Konzept, Schreibkompetenz indirekt über den Umweg globale oder kriteriale Einschätzungen von längeren Schreibleistungen zu erheben, die in der Literatur als wenig reliabel bekannt sind (Birkel & Birkel, 2002; Hennes et al., 2017). Stattdessen stellt der Ansatz – wie am Beispiel des Aufgabensets „Verrückte Gebäude“ gezeigt – den Versuch dar, entlang klarer und empirisch gesicherter Bewertungsrichtlinien Subkomponenten des Schreibprozesses möglichst eindeutig zu erfassen. Der Blick dieses in erster Linie förderdiagnostischen Ansatzes ist dabei stark auf die Intervention gerichtet. Angestrebtes Ziel ist es, die Möglichkeit zu schaffen, auf die einzelnen Komponenten der Textkompetenz einzugehen und zu bestimmen, in welchen dieser Bereiche Förderbe-

darf besteht. Gelingt dies, können Kinder/Jugendliche gezielt trainiert werden. Die beispielhaft präsentierte Aufgabe zeigt, dass dies prinzipiell möglich ist, wengleich sich daraus natürlich vorderhand keine Rückschlüsse auf die anderen angeführten Aufgaben ziehen lassen². Die Skala „verrückte Gebäude“ weist eine gute Konstruktvalidität auf. Das Konstrukt ist eindimensional und theoretische Vorannahmen über divergente und konvergente Validität wurden bestätigt.

Es ist selbstverständlich, dass die Beschreibung eines in Entwicklung befindlichen Instruments zahlreiche Limitationen aufweist. Derzeit befindet sich das Projekt noch in einer Phase, in der bislang lediglich Aussagen über die Eignung einer Subskala des Testverfahrens getroffen werden können. Nun gilt es, auch die übrigen Skalen (weiter) zu entwickeln und psychometrisch zu prüfen. Zusätzlich könnte es nötig sein, für einzelne Aspekte (z.B. Textmusterorientierung) noch weitere Aufgaben zu entwickeln³. Ein wichtiger Punkt dabei wird die Frage nach der Abgrenzung der Skalen (bzw. der einzelnen Dimensionen) voneinander sein. Schreibkompetenz ist ein, wie das Modell zeigt, komplexes Konstrukt mit mannigfaltigen Wechselwirkungen. So können z.B. Komponenten des Textmusters (Gliederungselemente, Formulierungen) gleichzeitig kohärenzstiftende Maßnahmen

² Erste Analysen zeigen allerdings, dass auch für diese Skalen ähnlich positive Ergebnisse zu erwarten sind.

³ Dies ist auch einer der Gründe, schon das noch unfertige Produkt vorzustellen. Wir erhoffen uns hier durchaus fachlichen Austausch.

sein, die dazu beitragen, dass die Bedürfnisse der adressierten Personen berücksichtigt und Erwartungen der Lesenden bedient werden. Dennoch bleibt es Ziel des vorgestellten Ansatzes, die einzelnen Testaufgaben/-skalen so zu gestalten, dass die Komponenten der Textkompetenz möglichst unabhängig erfasst/bewertet werden können. Mitgedacht wird dabei die Möglichkeit, im Rahmen einer Aufgabe unterschiedliche Dinge zu erheben/bewerten. Derzeit wird beispielsweise geprüft, ob sich die vorgestellte Aufgabe „verrückte Gebäude“ ebenfalls dazu eignet, anhand der produzierten Anleitungstexte auch Elemente der Textmusterrealisierung zu erheben bzw. zu beurteilen. Denkbar wäre eine Bewertung der Bauanleitung daraufhin, ob es sich hierbei lediglich um eine reine – wenn auch vollständige – Aneinanderreihung von Informationen handelt (z.B. „rot liegt verkehrt herum auf blau“) oder um einen instruktiven Text, der sich gezielt an die Adressat*innen wendet und der Aufgabenstellung „Bauanleitung“ entspricht (z.B. „Lege den roten Stein verkehrt herum auf den blauen Stein“). Es bleibt jedoch zu untersuchen, inwieweit auf diese Weise (relativ) unabhängige Konstrukte gemessen werden können.

Ein weiterer wichtiger ausstehender Punkt im Kontext der Validierung ist die Bestimmung des Zusammenhangs der entwickelten Einzelskalen mit einem globalen Maß der Schreibkompetenz. Die Erfassung einer solchen Leistung soll über einen längeren Text erfolgen. Hierzu wurde bereits eine Schreibaufgabe entwickelt, die sich momentan noch im Teststadium befindet. Ergebnis der Aufgabe soll ein Text sein, in dem die Funktionsweise eines vorab per Video erklärten technischen Apparates erklärt wird. Über eine globale Bewertung soll dann ein Maß für Schreibkompetenz generiert werden. Wichtig ist jedoch, mit Blick auf das bereits benannte Problem der fehlenden Reliabilität bei der globalen Textbewertung, dass das generierte Maß für die Schreibkompetenz gewissen statistischen Reliabilitätsanforderungen genügt. Nur

wenn dies der Fall ist, kann ein solches Maß zur Bestimmung von Zusammenhängen im Kontext von Validierungsuntersuchungen genutzt werden. Zur Bestimmung eines reliablen Maßes für die Schreibkompetenz soll die Methode des *Comparative Judgment* (Thurstone, 1927) genutzt werden. Diese Methode arbeitet mit Paarvergleichen: Mehrere Beurteiler*innen vergleichen unabhängig voneinander eine Serie von Textpaaren und entscheiden in jedem Vergleich neu, welcher der beiden Texte, bezogen auf die globale Textqualität, der bessere ist. Basierend auf der Anzahl von „gewonnenen“ oder „verlorenen“ Vergleichen pro Text kann eine Rangordnung der Texte im Hinblick auf deren globale Qualität gebildet werden (Lesterhuis, Verhavert, Coertjens, Donche & De Maeyer, 2017). Zur Herstellung dieser Rangfolge wird mit Hilfe eines logistischen Modells ein Logit-Score pro Text bestimmt. Dieser gibt die Wahrscheinlichkeit an, einen Vergleich zu einem Referenztext zu gewinnen (Politt, 2012). Die interne Konsistenz (Reliabilität) der geschätzten Logit-Werte kann über die sogenannte Scale-Separation-Reliabilität (analog zu Cronbach’s Alpha) ermittelt werden (Jones & Karadeniz, 2016). Die Logit-Werte können demnach als Qualitätsparameter der Texte, also als Maß der „globalen“ Schreibkompetenz, genutzt werden (Hennes et al., 2017). Die Umsetzung des Comparative Judgment erfolgt mit Hilfe der Software d-pac (www.d-pac.be), welche von einer belgischen Forschungsgruppe der Universität Antwerpen entwickelt wurde (Lesterhuis et al., 2017). Bei Einsatz der Software kann vorab bestimmt werden, welchen Reliabilitätswert die Rangfolge der Texte mindestens erreichen soll. Der auf diese Weise ermittelte Wert für die globale Schreibkompetenz (Logit-Score pro Text) einer TP ist dann mit den Ergebnissen der einzelnen Subtests zu korrelieren. Fragen zum Zusammenhang der entwickelten Einzelskalen und der Schreibkompetenz können auf diese Weise zuverlässig beantwortet werden.

Literatur

- Ahmed, Y., Wagner, R. K., & Lopez, D. (2014). Developmental relations between reading and writing at the word, sentence, and text levels: A latent change score analysis. *Journal of Educational Psychology, 106*, 419–434.
- Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. Dordrecht: Springer.
- Averintseva-Klisch, M. (2013). *Textkohärenz*. Heidelberg: Universitätsverlag Winter.
- Bach, R., Schmidt, B.M., Schabmann, A., & van Koll, S. (2016). Braucht mein Friseur wirklich Zirkel und Lineal? – Schulisches Basiswissen im Kontext der Ausbildungsreife. *Heilpädagogische Forschung, 42*, 61–72.
- Becker-Mrotzek, M., & Böttcher, I. (2006). *Schreibkompetenz entwickeln und beurteilen: Praxishandbuch für die Sekundarstufe I und II*. Berlin: Cornelsen Scriptor.
- Becker-Mrotzek, M., & Böttcher, I. (2012). *Schreibkompetenz entwickeln und beurteilen: Buch mit Kopiervorlagen* (4., überarbeitete Auflage). Berlin: Cornelsen Scriptor.
- Becker-Mrotzek, M. (2014). Schreibkompetenz. In J. Grabowski (Hrsg.), *Sinn und Unsinn von Kompetenzen: Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur* (S. 51–72). Opladen: Verlag Barbara Budrich.
- Becker-Mrotzek, M., Grabowski, J., Jost, J., Knopp, M., & Linnemann, M. (2014). Adressatenorientierung und Kohärenzherstellung im Text. Zum Zusammenhang kognitiver und sprachlich realisierter Teilkomponenten von Schreibkompetenz. *Didaktik Deutsch, 37*, 20–43.
- Becker-Mrotzek, M., & Schindler, K. (2007). Schreibkompetenz modellieren. In M. Becker-Mrotzek & K. Schindler (Hrsg.), *Texte schreiben* (S. 7–26). Köln: Gilles & Francke.
- Bereiter, C. (1980). Development in writing. In L. W. Gregg & E. R. Steinberg. (Eds.), *Cognitive processes in writing* (pp. 73–93). Hillsdale, NJ: Erlbaum.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written communication*. New York, NY: Routledge.
- Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities, 35*, 39–56.
- Birkel, P., & Birkel, C. (2002). Wie enig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht, 49*, 219–224.
- Bundesagentur für Arbeit (2006). Kriterienkatalog zur Ausbildungsreife. Nürnberg.
- Carroll, R., Meis, M., Schulte, M., Vormann, M., Kießling, J., & Meister H. (2015). Development of a German reading span test with dual task design for application in cognitive hearing research. *International Journal of Audiology, 54*, 136–141.
- Castillo, C., & Tolchinsky, L. (2018). The contribution of vocabulary knowledge and semantic orthographic fluency to text quality through elementary school in Catalan. *Reading and Writing, 31*, 293–323.
- Crossley, S. A., & McNamara, D. S. (2016). Say More and Be More Coherent: How Text Elaboration and Cohesion Can Increase Writing Quality. *Journal of writing research, 7*, 351–370.
- Dam-Jensen, H., & Heine, C. (2013). Writing and translation process research: Bridging the gap. *Journal of Writing Research, 5*, 89–101.
- De Jong, P. F., & Das-Smaal, E. A. (1990): The Star Counting Test: An attention test for children. *Personality and Individual Differences, 11*, 597–604.
- Drijbooms, E., Groen, M. A., & Verhoeven, L. (2015). The contribution of executive functions to narrative writing in fourth grade children. *Reading and Writing, 28*, 989–1011.
- Drijbooms, E., Groen, M. A., & Verhoeven, L. (2017). How executive functions predict development in syntactic complexity of

- narrative writing in the upper elementary grades. *Reading and Writing*, 30, 209–231.
- Ehlich, K. (1983). Text und sprachliches Handeln. Die Entstehung von Texten aus dem Bedürfnis nach Überlieferung. In A. Assmann, J. Assmann & C. Hardmeier (Hrsg.), *Schrift und Gedächtnis* (S. 24–43). München: Fink.
- Farnan, N., & Fearn, L. (2008). Writing in the disciplines: More than just writing across the curriculum. In D. Lapp, J. Foold & N. Farnan (Eds.), *Content area reading and learning* (pp 403–424). New York, NY: Erlbaum.
- Feenstra, H. (2014). *Assessing writing ability in primary education: on the evaluation of text quality and text complexity*. Enschede: Universiteit Twente.
- Graham, S., & Harris, K. R. (2010) The role of self-regulation and transcription skills in writing and writing development, *Educational Psychologist*, 35:1, 3–12.
- Hamill, D.D., & Larsen, S.C. (2009). *Test of Written Language - TOWL* (Fourth Edition). Austin, TX: Pro-ed.
- Harsch, C., Neumann, A., Lehmann, R., & Schröder, K. (2007). Schreibfähigkeiten. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung: DESI-Studie* (S. 38–58). Weinheim: Beltz.
- Hasselhorn, M., Schumann-Hengsteler, R., Gronauer, J., Grube, D., Mähler, C., Schmid, I., Seitz-Stein, K., & Zoelch, C. (2012). *AGTB 5-12. Arbeitsgedächtnisbatterie für Kinder von 5 bis 12*. Göttingen: Hogrefe.
- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29, 369–388.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the Organization of Writing Processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Heinemann, W. (2000). Textsorten: Zur Diskussion um Basisklassen des Kommunizierens. Rückschau und Ausblick. In K. Adamzik (Hrsg.), *Textsorten. Reflexionen und Analysen* (S. 9–29). Tübingen: Stauffenburg.
- Hennes, A.-K., Schmidt, B.M., Schabmann, A., & Rietz, C. (2017). *Teacher's expertise in assessing written composition*. Spoken paper presented at Twenty-fourth Annual Meeting of the Society for the Scientific Study of Reading, Halifax, Canada July, 12.-15. 2017.
- Jones, I., & Karadeniz, I., 2016. *An alternative approach to assessing achievement*. In: Proceedings of the 2016 40th Conference of the International Group for the Psychology of Mathematics Education, Szeged, Hungary, 3-7 August 2016.
- Kaufman, A.S., & Kaufman, N.L. (2005). *Kaufman Test of Educational Achievement* (Second Edition). London, UK: Pearson.
- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. E. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57–71). Mahwah, NJ: Erlbaum.
- Jakobs, E. M., & Perrin, D. (Eds.). (2014). *Handbook of writing and text production* (Vol. 10). Berlin: Walter de Gruyter.
- Knopp, M., Becker-Mrotzek, M., & Grabowski, J. (2013). Diagnose und Förderung von Teilkomponenten der Schreibkompetenz. In: A. Redder & Weinen, S. (Hrsg.), *Sprachförderung und Sprachdiagnostik: Interdisziplinäre Perspektiven aus Psychologie, Sprachwissenschaft und empirische Bildungsforschung* (S. 296–315). Münster Waxmann.
- Knopp, M., Jost, J., Linnemann, M., & Becker-Mrotzek, M. (2014). Textprozeduren als Indikatoren von Schreibkompetenz: Ein empirischer Zugriff. In T. Bachmann & H. Feilke (Hrsg.), *Werkzeuge des Schreibens: Beiträge zu einer Didaktik der Textprozeduren* (S. 111–128). Stuttgart: Klett.
- Knopp, M., Jost, J., Nachtwei, N., Becker-Mrotzek, M., & Grabowski, M. (2012). Teilkomponenten von Schreibkompetenz untersuchen: Bericht aus einem interdisziplinären empirischen Projekt. In: Bayrhuber, H. et al. (Hrsg.), *Forschung: Empirische Projekte*

- historische Analysen – theoretische Grundlagen* (S. 47–65). Münster: Waxmann.
- Kruse, N., Reichardt, A., Herrmann, M., Heinzel, F., & Lipowsky, F. (2012). Zur Qualität von Kindertexten, Entwicklung eines Bewertungsinstruments in der Grundschule. *Didaktik Deutsch*, 18, 87–110.
- Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2017). Comparative judgement as a promising alternative to score competences. In E. Cano & G. Ion (Eds.), *Innovative practices for higher education assessment and measurement* (pp. 119–138). Hershey, PA: IGI Global Publications.
- Linnemann, M., & Stephany, S. (2014). Supportive writing assignments for less skilled writers in mathematic classroom. In P. Klein, P., Boscolo, L. Kirkpatrick, & C. Gelati, C. (Eds.), *Studies in writing: Writing as a learning activity* (pp. 6–93). Leiden, NL: Koninklijke Brill NV.
- McArthur, C. A., Graham, S., & Fitzgerald, J. (2016). *Handbook of writing research*. New York, NY: Guilford Press
- MacArthur, C. A., & Graham, S. (2016). Writing research from a cognitive perspective. In: C. A., MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*, (pp. 24–40). New York, NY: Guilford Press.
- Madelaine, A., & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52, 33–43.
- Mikulecky, L., & Kirkley, J. R. (1998). Literacy instruction for the 21st-century workplace. *Peabody Journal of Education*, 73, 290–316.
- Nussbaumer, M. (1991). *Was Texte sind und wie sie sein sollen. Ansätze zu einer sprachwissenschaftlichen Begründung eines Kriterienrasters zur Beurteilung von schriftlichen Schülertexten*. Tübingen: Niemeyer.
- Olive, T., Kellogg, R. T., & Piolat, A. (2008). Verbal, visual, and spatial working memory demands during text composition. *Applied Psycholinguistics*, 29, 669–687.
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R. and Richardson, C. (1995). A Redrawn Vandenberg & Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition*, 28, 39–58.
- Philipp, M. (2015). *Schreibkompetenz: Komponenten, Sozialisation und Förderung* (Vol. 4457). Tübingen: A. Francke Verlag.
- Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education*, 22, 157–170.
- Pohl, T., & Steinhoff, T. (2010). *Textformen als Lernformen*. Duisburg: Gilles & Francke.
- Pohlmann, B. & Möller, J. (2007). Assimilations- und Kontrasteffekte bei der Bewertung von Texten. *Zeitschrift für Pädagogische Psychologie*, 21, 297–303.
- Sandig, B. (1997). Formulieren und Textmuster am Beispiel von Wissenschaftstexten. In E.-M. Jakobs & D. Knorr (Hrsg.), *Schreiben in den Wissenschaften* (S. 25–44). Frankfurt/M.: Peter Lang.
- Schneider, W., Schlagmüller, M., & Ennemoser, M. (in press). *Lesegeschwindigkeits- und -verständnistest für die Klassenstufen 5-13 (LGVT 5-12+)*. Göttingen: Hogrefe.
- 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, & U. H. Waßner (Hrsg.), *Text - Verstehen. Grammatik und darüber hinaus* (S. 222–238). Berlin: de Gruyter.
- Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 95–100). Weinheim: Beltz.
- Shanahan, T. (2006a). Where does writing fit in Reading First? In C. Cummins (Eds.), *Understanding and implementing Reading First initiatives* (S. 106–115). Newark, DE: International Reading Association.
- Shanahan, T. (2006b). Relations among oral language, reading, and writing develop-

- ment. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 171–183). New York, NY: Guilford Press.
- Steinert, J. (2011). *Allgemeiner Deutscher Sprachtest (ADST)*. Göttingen: Hogrefe.
- Sturm, A., & Weber, M. (2016). *Schreibkompetenz, Schreibmotivation, Schreibförderung. Grundlagen und Modelle zum Schreiben als soziale Praxis*. Seelze: Kallmeier.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286.
- UNESCO (2013). *Implementation of the international plan of action for the United Nations literacy decade: Report of the director-general of the United Nations Educational, Scientific and Cultural Organization*. Verfügbar unter http://www.unesco.de/fileadmin/medien/Dokumente/Bildung/Alphabetisierung_Literacy_decade_Abschlussbericht.pdf (Accessed 28.02.2018.)
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2-revision (CFT 20-R)*. Göttingen: Hogrefe.
- Wimmer, H., & Mayringer, H. (2014). *Salzburger Lese-Screening für die 2.-9. Klasse (SLS 2-9)*. Bern: Huber.

Ann-Kathrin Hennes

Humanwissenschaftliche Fakultät
Universität zu Köln
Klosterstraße 79b
50931 Köln
Deutschland

Telefon: 0221/470-5536

E-Mail: ann-kathrin.hennes@uni-koeln.de

Erstmalig eingereicht: 12.08.2017

Überarbeitung eingereicht: 05.03.2018

Angenommen: 30.04.2018