

Sanna Pohlmann-Rother, Edgar Schoreit & Anja Kürzinger

Schreibkompetenzen von Erstklässlern quantitativ-empirisch erfassen – Herausforderungen und Zugewinn eines analytisch-kriterialen Vorgehens gegenüber einer holistischen Bewertung

Zusammenfassung

Im Beitrag wird der Frage nachgegangen, ob eine holistische und eine analytische Auswertungsstrategie schon bei Texten des ersten Grundschuljahres gleichermaßen geeignet sind, um Schreibkompetenz reliabel zu erfassen. Die Analysen zu den insgesamt 540 Texten stammen aus dem DFG-Projekt Narrative Schreibkompetenz in Klasse 1 (NaSch 1). Zunächst wird die Dimensionalität des Konstrukts Schreibkompetenz anhand der analytischen Auswertung untersucht. Im Anschluss werden die mit den beiden Auswertungsstrategien ermittelten Kompetenzwerte unter Berücksichtigung der Textlänge miteinander verglichen. Die Ergebnisse der Rasch-Skalierung der analytischen Kriterien weisen auf eine zweidimensionale Struktur (semantisch-pragmatisch vs. sprachsystematisch) der Schreibkompetenz hin. Dabei zeigt sich ein stärkerer Zusammenhang des holistischen Ratings mit der semantisch-pragmatischen ($r = .78$) als mit der sprachsystematischen Dimension ($r = .47$). Die Textlänge wiederum weist eine gleichermaßen hohe Übereinstimmung mit der semantisch-pragmatischen Dimension ($r = .63$) und dem holistischen Rating auf ($r = .62$), deren Zusammenhang zueinander aber auch nach Herauspartialisieren der Textlänge deutlich bleibt ($r = .64$). Zwischen Textlänge und der sprachsystematischen Schreibkompetenz lässt sich dagegen kein bedeutsamer Zusammenhang nachweisen ($r = .09$).

Prof. Dr. Sanna Pohlmann-Rother (corresponding author) · Dipl. Soz. Anja Kürzinger, Institut für Erziehungswissenschaft, Pädagogische Hochschule Ludwigsburg, Reuteallee 46, 71634 Ludwigsburg, Deutschland
E-Mail: pohlmann@ph-ludwigsburg.de
kuerzinger@ph-ludwigsburg.de

Dipl. Psych. Edgar Schoreit, Institut für Erziehungswissenschaft, Universität Kassel, Nora-Platiel-Str. 1, 34127 Kassel, Deutschland.
E-Mail: schoreit@uni-kassel.de

Schlagworte

Schreibkompetenz; Textqualität; holistisches Rating; Kriteriale Auswertung; Schreibunterricht

The quantitative empirical determination of writing skills of first graders – Challenges and gain of an approach based on analytical criteria compared to a holistic evaluation

Abstract

The article examines whether holistic and analytic evaluation-strategies are comparably suitable to reliably determine writing skills already in texts of first graders. The analyses of the 540 texts in total are sourced from the DFG project Narrative Writing Skills in Grade One (NaSch 1). First of all, the dimensionality of writing skills is being examined based on the analytic evaluation. Subsequently, the competence values determined by the two evaluation strategies are compared taking into account the length of text. The results of the Rasch scaling of the analytic criteria are indicating a two-dimensional structure (semantic-pragmatic vs. language-systematic) of writing skills where a stronger correlation of the holistic rating with the semantic-pragmatic ($r = .78$) than with the language-systematic dimension ($r = .47$) is emerging. In turn, the length of text shows a comparably high accordance with the semantic-pragmatic dimension ($r = .63$) and the holistic rating ($r = .62$), whose correlation still stays distinct when controlling the influence of the length of text using partial correlation ($r = .64$). In contrast, a correlation between the length of text and the language-systematic writing skills ($r = .09$) cannot be identified.

Keywords

Writing skills; Text quality; Holistic rating; Analytic evaluation; Writing education

1. Einleitung

Die Frage nach der Bestimmung von Textqualität hat in der gegenwärtigen Diskussion insbesondere im Kontext nationaler Schulleistungsstudien und der Überprüfung von Bildungsstandards eine herausragende Bedeutung (z. B. Böhme, Bremerich-Vos & Robitzsch, 2009). Dementsprechend wird den Instrumenten zur Erfassung der Textqualität besondere Aufmerksamkeit geschenkt. In den meisten Studien werden Kriterienkataloge eingesetzt, die nicht nur linguistischen und schreibdidaktischen Anforderungen genügen, sondern auch forschungsmethodischen Ansprüchen an die fachdidaktisch-empirische Unterrichtsforschung ge-

recht werden müssen (Blatt, Ramm & Voss, 2009; Kruse, Reichardt, Herrmann, Heinzl & Lipowsky, 2012). Je nach Interesse und Ausgangsfrage unterscheiden sich die Modelle zur Erfassung der Textqualität in ihrer Ausführlichkeit und Schwerpunktsetzung. Die meisten empirischen Studien kombinieren zur Sicherung der notwendigen Beurteilerübereinstimmung und der Reliabilität der Urteile die analytische Beurteilung auf der Grundlage getrennt erfasster Einzelkriterien mit einem Gesamturteil, bei dem mehrere Beurteiler einen Text nach ihrem „holistischen“ Eindruck einstufen (Böhme et al., 2009; Kruse et al., 2012; vgl. auch Baurmann, 2006; Grzesik & Fischer, 1984). Im Unterschied zum holistischen Urteil erlauben Kriterienkataloge differenzierte Urteile über die Qualität von Texten im Hinblick auf verschiedene Aspekte, wie z. B. die formale, sprachliche, strukturelle und inhaltliche Qualität. Vorliegende Schreibkompetenzuntersuchungen beschäftigen sich schwerpunktmäßig mit Texten aus der Sekundarstufe oder grundschulbezogen überwiegend mit Texten ab der dritten Jahrgangsstufe.

Die hier berichteten Analysen zur Textqualität beziehen sich auf Texte, die im März der ersten Jahrgangsstufe verfasst wurden. Da eine detaillierte, quantitative Analyse der Textqualität für diese Jahrgangsstufe ein Novum darstellt, soll im vorliegenden Beitrag zunächst geklärt werden, inwiefern Texte aus dieser Jahrgangsstufe überhaupt einer differenzierten Analyse unterzogen werden können und welche Herausforderungen in diesem Kontext entstehen. Konkret wird dieses Anliegen in einem ersten Schritt mit der Frage nach der Dimensionalität des Konstrukts Schreibkompetenz verfolgt. Danach werden die ermittelten Kompetenzwerte der holistischen und analytischen Auswertung miteinander verglichen und Zusammenhänge und Unterschiede zwischen beiden Ratings aufgezeigt. Dabei wird auch der Zusammenhang zwischen Textqualität und Textlänge thematisiert. Die vorliegenden Analysen stammen aus dem Projekt *Narrative Schreibkompetenz in Klasse 1* (NaSch 1)¹. Dieses Projekt greift auf ein in der Studie *Persönlichkeits- und Lernentwicklung von Grundschulkindern* (PERLE)² erhobenes Textkorpus aus den Bundesländern Sachsen, Thüringen, Berlin und Brandenburg zurück.

1 Das DFG-Projekt hat den Titel „Narrative Schreibkompetenz in Klasse 1 (NaSch 1)“. Leitung: Sanna Pohlmann-Rother & Gabriele Faust (†), Förderkennzeichen PO 1739/1-1.

2 Die PERLE-Studie untersuchte in zwei Phasen (PERLE I 2006–2008; PERLE II 2009–2011) die Persönlichkeits- und Lernentwicklung von Grundschulkindern. Leitung: Frank Lipowsky & Gabriele Faust (vgl. Lipowski, Faust & Kastens, 2013). In den PERLE-Videomodulen (PERLE I) unterrichteten die Lehrkräfte auf der Basis von Vorgaben. Im Videomodul Deutsch sollten sie den Anfangsteil des fantastischen Bilderbuchs „Lucy rettet Mama Kroko“ (Doucet & Wilsdorf, 2005) vorstellen und u. a. eine Schreibphase (Verfassen eines Briefes) durchführen. Die dort entstandenen Briefe werden in NaSch 1 ausgewertet. Das Projekt NaSch 1 dankt den Projektleitern der PERLE-Studie für die Verwendung der dort erhobenen Daten.

2. Textqualität und Schreibkompetenz

In Abgrenzung zur sprachsystematisch ausgerichteten Textlinguistik erscheint ein Text vor dem Hintergrund der kommunikationsorientierten Textlinguistik (Brinker, 2010, S. 14) nicht nur als grammatisch verknüpfte Satzfolge, sondern als (komplexe) sprachliche Handlung, mit der ein Schreiber eine bestimmte kommunikative Beziehung zum Leser oder Hörer herzustellen versucht (ebd., 2010, S. 15; Becker-Mrotzek, 1997). Dieser Auffassung muss ein Textbegriff zugrunde liegen, der es ermöglicht, den Text als eine sprachliche und zugleich kommunikative Einheit zu beschreiben: „Der Terminus ‚Text‘ bezeichnet eine begrenzte Folge von sprachlichen Zeichen, die in sich kohärent ist und die als Ganzes eine erkennbare kommunikative Funktion erfüllt“ (Brinker, 2010, S. 17). Insofern sollten Texte adressatenorientiert sein, über einen Kohärenzstiftenden Pfad verfügen, der es dem Leser erlaubt, ein mentales Modell des Textes zu (re-)konstruieren, und angemessen und domänenspezifisch im Wortschatz sein (Knopp, Jost, Nachtwei, Becker-Mrotzek & Grabowski, 2012, S. 50). In NaSch 1 wird im Sinne des Kompetenz-Performanz-Modells davon ausgegangen, dass die gezeigte Leistung der Schülerinnen und Schüler (Textqualität) ein Indikator für eine entsprechende dahinter liegende Schreibkompetenz ist (Harsch, Neumann, Lehmann & Schröder, 2007; Knopp et al., 2012, S. 51).

Schreibkompetenz kann vor dem Hintergrund des zugrunde liegenden Textbegriffs als Fähigkeit zur Lösung kommunikativer Probleme in konkreten Anwendungssituationen verstanden werden (Becker-Mrotzek & Böttcher, 2011, S. 56). Sichtet man die aktuelle sprachwissenschaftliche und sprachdidaktische Literatur, wird die Komplexität dieses Begriffs deutlich. Mehrheitlich wird Schreibkompetenz als Zusammenspiel verschiedener Teilkompetenzen, als Kombination von Wissen und Fähigkeiten und zugleich als Teil einer umfassenderen Sprachkompetenz verstanden (Becker-Mrotzek & Schindler, 2007, S. 12; Knopp et al., 2012; Becker-Mrotzek, 2014; Fix, 2008). Aufgrund der Vielfalt und Komplexität von Texten und den verschiedenen möglichen Beurteilungskriterien erscheint es sinnvoll, das jeweilige Verständnis von Schreibkompetenz an dem Schreibanlass und der darauf bezogenen Textfunktion zu orientieren, wie es beispielsweise in der Studie *Deutsch Englisch Schülerleistungen International* (DESI) realisiert wurde: „Unter Schreibkompetenz wird hier die Fähigkeit verstanden, Texte adressatengerecht zu formulieren und, je nach Zielsetzung, präzise zu informieren, überzeugend zu argumentieren oder Sprache ästhetisch ansprechend und kreativ einzusetzen“ (Harsch et al., 2007, S. 54). Da dies je nach Erwerbsstadium unterschiedlich gelingt, können Texte grundsätzlich als „Ausformung[en] einer konkreten Schreibentwicklungsphase“ (Pohl & Steinhoff, 2010, S. 12) verstanden werden. Die Beurteilung der Schreibprodukte und damit auch die Kriterienentwicklung und -skalierung zur Bestimmung der Textqualität muss rational angelegt sein, denn „Kompetenz ist ... in Relation zu einer Entwicklungsspanne bzw. der in ihr zu erwartenden Schreibleistungen zu bestimmen. *Kompetent* soll eine Lernerleistung

dann sein, wenn sie vor dem Hintergrund erwartbarer und vergleichbarer Lernleistungen aus derselben Entwicklungsphase als gelungen oder erfolgreich gelten kann“ (Pohl & Steinhoff, 2010, S. 22). In NaSch 1 wird Schreibkompetenz als altersgemäße (auf das „Schreibalter“ bezogene, vgl. Feilke, 1996, S. 1181) Fertigkeit verstanden, die mit einer Schreibaufgabe verbundenen Anforderungen zu bewältigen.

Das Schreibarrangement in NaSch 1 zeichnet sich durch ein handlungs- und produktionsorientiertes Verfahren aus. Im Anschluss an das Bilderbuch „Lucy rettet Mama Krok“ (Doucet & Wilsdorf, 2005) sollten in 37 ersten Klassen Texte (Textsorte Brief) aus der Perspektive der Hauptfigur „Lucy“ an eine andere literarische Figur, „Mama Krok“, geschrieben werden. Das dieser Aufgabe zugrunde liegende Konzept des perspektivischen Schreibens (Fritzsche, 2010; Böttcher, 2010, S. 169) wird dem Anspruch gerecht, bereits im Anfangsunterricht anspruchsvolle Schreibaufgaben einzusetzen (z. B. Weinhold, 2000; Dehn, 2006; Kruse, 2006). Für die Auswertung der Texte wurde sowohl ein umfangreicher Kriterienkatalog als auch ein holistisches Rating für ein Globalurteil entwickelt (zu den deskriptiven Ergebnissen vgl. Kürzinger & Pohlmann-Rother, 2015).

Im vorliegenden Beitrag soll ein Vergleich beider Auswertungsstrategien vorgenommen und die innere Struktur der Texte analysiert werden, um Hinweise auf globale Kategorien des Schreibens zu erhalten. Die Dimensionalität von Schülertexten wurde bereits in empirischen Schreibkompetenzuntersuchungen modelliert, die im folgenden Abschnitt dargestellt werden. Die Frage nach den textimmanenten Strukturen von Schreibprodukten aus der ersten Klasse stellt allerdings ein Novum dar.

3. Textstrukturen: Dimensionalität von Schreibkompetenz

In Schreibkompetenzuntersuchungen in der empirischen Bildungsforschung spielt die Frage nach den textimmanenten Strukturen eine bedeutende Rolle. Ein Blick auf den Forschungsstand zeigt, dass bisher entweder ein- oder zweidimensionale Fähigkeitsausprägungen für das Verfassen von Texten identifiziert werden konnten. Im nationalen Teil der von der International Association for the Evaluation of Educational Achievement (IEA) durchgeführten Aufsatzstudie der 11. Jahrgangsstufe wurden die Texte der Schülerinnen und Schüler hinsichtlich des Gesamteindrucks und den analytischen Textkategorien *Inhalt*, *Aufbau*, *Stil* und *formale Korrektheit* auf einer fünfstufigen Skala erfasst. Die Ergebnisse weisen eine zweidimensionale Struktur nach. Es lässt sich ein *aufgabenspezifischer Faktor* für Inhalt, Aufbau und Stil der Texte sowie ein *mechanics-factor* für Orthographie und Grammatik identifizieren (Lehmann, 1987, 1990). Dieser Befund konnte in der DESI-Studie repliziert werden. DESI ermittelte am Anfang und am Ende der Klassenstufe 9 an einer großen bundesweiten Stichprobe die Schreibkompetenz

der Schülerinnen und Schüler in der deutschen Sprache im Briefeschreiben (Beck, Bundt & Gomolka, 2008). Diese in DESI erhobenen Daten wurden zusammen mit den Aufsatzdaten aus der Hamburger Lernausgangslagenuntersuchung in Klassenstufe 11 an allgemeinbildenden und berufsbildenden Schulen (LAU 11/ULME I) ausgewertet (Neumann, 2007). Es wurden die Kategorien *Inhaltliche und formale Merkmale*, *Textaufbau*, *Rechtschreibung*, *Grammatik*, *Stil*, *Wortwahl* und *Satzkonstruktion* auf gemeinsame dahinter liegende Dimensionen untersucht. Wieder zeigt sich eine zweidimensionale Schreibstruktur mit auf die Aufgabe bezogenen semantisch-pragmatischen Fähigkeiten (alle inhaltlich-formalen Anforderungen, Textaufbau, Stil, Wortwahl) und übergreifenden sprachsystematischen Fähigkeiten (Rechtschreibung, Grammatik, Satzkonstruktion).

Auch Böhme et al. (2009) schließen aus ihren Analysen im Rahmen der seit 2005 durchgeführten Evaluation der Bildungsstandards im Einklang mit der DESI-Studie (Neumann, 2007) auf eine „ähnliche zweidimensionale Grundstruktur der Schreibfähigkeit“ (Böhme et al., S. 319) mit einem stilistisch/inhaltlichen Bereich und einer Komponente der sprachlichen Richtigkeit (Rechtschreibung und Grammatik). Die holistische allgemeine Qualitätseinschätzung der Schülertexte lehnt sich an die sechsstufige Skala des im National Assessment of Educational Progress (NAEP) eingesetzten Holistic Scoring Guide in der Version für die vierte Jahrgangsstufe an. Die Variablen der analytischen Auswertung wurden entweder dichotom (Codes 0/1) bzw. niedrig inferent (Codes 0/1/2) kodiert. Neben dem sechsstufigen Gesamteindruck wurden drei weitere Variablen (Stil, Inhalt, sprachliche Richtigkeit) vierstufig geratet.

Es wird deutlich, dass die drei skizzierten Modelle einen ähnlichen Faktor der Sprachsystematik (Grammatik und Orthographie) beinhalten, der sich von einer aufgabenbezogenen inhaltlich-pragmatischen Dimension abgrenzen lässt. Dies deckt sich mit der von Quellmalz, Capell und Chou (1982) gefundenen Unterscheidung eines sich auf die spezifische Aufgabe beziehenden *coherence*-Faktors (Inhalt und Aufbau) und eines aufgabenübergreifenden *mechanics*-Faktors (Rechtschreibung, Grammatik) und unterstreicht, dass die Schreibkompetenz nicht losgelöst von der Aufgabenstellung betrachtet werden kann: „These data cast doubt on the assumption that ‚a good writer is a good writer‘ regardless of the assignment“ (ebd., 1982, S. 255). Mit diesen Befunden stehen auch Studien im Einklang, die sprachsystematische Aspekte unberücksichtigt lassen und bei diesem Vorgehen eine Eindimensionalität der Schreibkompetenz nachweisen. Diese bilden inhaltliche, formale und stilistische Aspekte auf einem Generalfaktor vor dem Hintergrund der Schreibaufgabe ab.

In der Internationalen Grundschul-Leseuntersuchung – Erweiterungsstudie 2006 (IGLU-E) wurde das Textschreiben (Textsorte Brief) anhand von 29 Kriterien zu den Bereichen Inhalt (Vielfalt/Qualität der Ideen, Erfassung des Themas), Aufbau (formaler Aufbau der Textsorte Brief mit Anrede, Einleitung, Abschied) und sprachliche Gestaltung (Sprachstil, Adressatenbezug, schriftsprachliche Ausdrucksmittel) – ausgenommen der *formalen Korrektheit* – explorativ untersucht (Blatt, Voss & Matthießen, 2005). Zudem wurden übergeordnete Kategorien,

wie z. B. der sich auf inhaltliche, strukturelle und sprachliche Aspekte beziehende „rote Faden“, berücksichtigt. Analysen zum Zusammenhang dieser Merkmale zeigen, dass sich wesentliche Kriterien aus Inhalt, Aufbau und Sprache durch einen Generalfaktor erklären lassen. Dieser beinhaltet die Items Adressatenbezug, Satzbau, Nebensätze, Kommentierung, Anrede, Einleitung, Abschied, neue und ungewöhnliche Ideen, Einnahme der Zeitperspektive, Veränderungen in der Zukunft (ebd., 2005, S. 127).

Im Rahmen des Länderkooperationsprojekts Lernstand in Klasse 6 wurde eine Schreibaufgabe (Textsorte Brief) anhand von 16 dichotomen Items zu Inhalt (elementarer und entfalteter Leserbezug), Aufbau (Textsorte Brief: Anrede, Hauptteil, Abschied, Unterschrift), Sprache/Stil (Erläuterung/Begründung/Kommentierung und sprachliche Komplexität) und Einhaltung der Schreibkonventionen (Beurteilung von Rechtschreibung und Grammatik im Hinblick auf die Lesbarkeit des Textes und Beurteilung des Schriftbildes im Hinblick auf Lesbarkeit) analysiert (Blatt et al., 2009). Es wurde sowohl ein eindimensionales Modell, das die Schreibkompetenz global erfasst (Aufbau, Inhalt, Sprache), sowie ein zweidimensionales Modell (Inhalt/Aufbau vs. Sprache/Stil) geprüft. Die Skalierung ergab für das *eindimensionale Modell* die beste Passung (ebd., 2009, S. 75). Für die Modellprüfung wurden auch hier nur solche Kriterien herangezogen, die Schreibkompetenz im engeren Sinne abbilden und nicht die Einhaltung der Schreibkonventionen testen (Rechtschreibung, Grammatik).

In Abgrenzung zu den bisher skizzierten Befunden weist das Modell der Kasseler Studie KoText zu kooperativen Schülerrückmeldungen bei der Textüberarbeitung (3. Jahrgangsstufe) eine gänzlich andere Struktur auf (Kruse et al., 2012). Obwohl auch in diesen Analysen Variablen zur Erfassung der sprachsystematischen und orthographischen Richtigkeit ausgeklammert wurden (ebd., 2012, S. 94), gelangen die Autoren zu einem zweidimensionalen Modell. Das in Anlehnung an das Zürcher Textanalyseraster (Nussbaumer & Sieber, 1994) entwickelte Rating umfasste neben dem holistischen Gesamturteil (vierstufig) neun vierstufige Textqualitätsitems. Es wurde ein zweifaktorielles Modell mit einem *Konventionalitätsfaktor* und einem *Unkonventionalitätsfaktor* angenommen. Nach Entfernung dreier Items ergaben sich die Faktoren *Textualität konventionell* (4 Items: Kohärenz, Expliztheit, Impliztheit, Wortschatz) und *Textualität unkonventionell* (2 Items inhaltliches und sprachliches Wagnis). Allerdings setzt diese Zweidimensionalität ohne Berücksichtigung sprachsystematischer Aspekte die Erfassung der beiden Aspekte inhaltlichen und sprachlichen Wagnisses voraus, welche möglicherweise nur bei bestimmten Schreibanlässen entfaltet werden können.

4. Texte von Erstklässlern beurteilen – Eine Herausforderung

Mittlerweile gehört es zum schulischen Alltag, dass Kinder bereits zu Beginn des Schriftspracherwerbs eigene Texte verfassen. Die empirische Analyse von Texten aus der ersten Jahrgangsstufe ist jedoch keine Selbstverständlichkeit. Die Mehrheit von Studien zur Schreibkompetenz beschränkt sich auf höhere Klassenstufen (vgl. z. B. Böhme et al., 2009 und Thonke, Groß Ophoff, Hosenfeld & Isaac, 2008 für Klasse 3 und 4; Kruse et al., 2012 für Klasse 3; Augst, Disselhoff, Henrich, Pohl & Völzing, 2007 für die Klassen 2 bis 4). Die Forschung zu *textueller Handlungskompetenz* (Feilke, 1993) von Erstklässlern steht dagegen erst am Anfang und wurde nur selten an umfangreichen Textkorpora vorgenommen (Dehn, Merklinger & Schüler, 2011; Weinhold, 2000). In der Schreibforschung wird der Beginn der Schreibentwicklung als assoziatives, wenig geplantes und kaum adressatenorientiertes Schreiben charakterisiert, was Bereiter und Scardamalia (1987) als Resultat einer Strategie des *knowledge telling* beschreiben (Fix, 2008; Bremerich-Vos & Possmayer, 2011). Damit werden Fähigkeiten vorausgesetzt, zu denen das Schreiben-Können als basale Ausführungshandlung gehört. Dies stellt aus kognitionspsychologischer Sicht eine notwendige Voraussetzung für das Textschreiben dar, weil Schreiber auf graphematische und motorische Schreibroutinen zurückgreifen müssen, um genügend freie mentale Kapazitäten für die Erfordernisse des Textschreibens zu besitzen. Über diese Routinen im Umgang mit den medialen Bedingungen von Schriftlichkeit verfügen Erstklässler Mitte der ersten Klasse nur begrenzt. Sie müssen sich mit den sprachlichen, motorischen und konzeptionellen Anforderungen des Schreibens auseinandersetzen, Sprache aus dem Zusammenhang ihres Gebrauchs herauslösen, graphematisch reflektieren und eine graphomotorische Umsetzung der Wörter leisten (Weinhold, 2002). Rechtschreibfehler stellen auf dieser Entwicklungsstufe der beginnenden orthographischen Strategie, in der noch eine lautgetreue Verschriftung vorherrscht, unterschiedliche Zugriffsweisen auf Schrift dar (Schründer-Lenzen, 2013). Die korrekte Orthografie spielt aus einer *kindzentrierten* Perspektive (Dehn, 1996) und vor dem Hintergrund der begrenzten Kapazitäten der Erstklässler beim Verfassen eigener Texte eine untergeordnete Rolle. Die Schreiber sollen vielmehr auf diesem Gebiet entlastet werden, damit sie sich auf die höherrangigen Prozesse der Ideenentwicklung und -strukturierung sowie die sprachliche Gestaltung konzentrieren können (ähnlich auch Blatt et al., 2005, S. 119 und Kruse et al., 2012, S. 93–94). Ob eine derartige Entlastung tatsächlich erfolgen kann oder ob die grundlegenden technischen Anforderungen des Schreibens diese höherrangigen Prozesse in diesem frühen Schreibalter automatisch einschränken, wirft auch die Frage auf, ob die bei älteren Schülerinnen und Schülern wiederholt vorgefundene Unterteilung in einen *mechanics-factor*, welcher die rein technischen Aspekte des Schreiben-Könnens wie die Rechtschreibung erfasst, und eine eher inhaltlich-stilistische Sprachbeherrschung (s. Kapitel 3) in Texten des ersten Schuljahrs überhaupt

schon erkennbar wird oder ob sich diese beiden Aspekte zu diesem Zeitpunkt noch nicht voneinander trennen lassen.

Inhaltlich müssen die Erstklässler eine „zerdehnte Sprechsituation“ (Ehlich, 1984, S. 18) bewältigen, da Sender und Empfänger keinen „Raum gemeinsamer sinnlicher Gewißheit“ (Ehlich, 1984, S. 16) teilen. Anders als in einer *face-to-face*-Kommunikationssituation zeichnet sich die Kommunikation mit Texten durch einen Bruch in Zeit und Raum aus: Der schriftliche Textproduzent kann sich nicht auf Kontextinformationen stützen, sondern muss solche für das Verstehen des Textes zentralen Informationen in seinem Text erst etablieren, indem er eine aus sich heraus verständliche Welt erschafft (Fix, 2008; Baurmann & Pohl, 2009; Pohl, 2013). Vor diesem Hintergrund sind Kohärenzstiftende Kriterien, wie die Verwendung von Kohäsionsmitteln und die Entfaltung der Textidee, bedeutsam.

Angesichts der knappen textuellen Entfaltung der Schreibprodukte (im Mittel 2.1 Sätze; $SD = 1.48$; Min.–Max. = 1–10; bei durchschnittlich 17.5 Wörtern, $SD = 9.59$) war bei den vorliegenden Analysen insbesondere die Einschätzung der sprachlichen Qualitätsaspekte (z. B. Wortschatz, Entfaltung der Textidee) eine Herausforderung (vgl. Kapitel 5). Vor Beginn der Ratings wurden in einer intensiven Pilotierungsphase Kategorien mit ausführlichen Ratinganleitungen und Benchmarktexten in einem diskursiven Prozess entwickelt, wobei immer wieder Präzisierungen nötig waren, um auch geringe Varianzen in den Texten zu erfassen.

Auch der bekannte Zusammenhang zwischen Textlänge und (holistischer) Benotung durch Lehrkräfte (Birkel & Birkel, 2002, S. 221) dürfte bei den insgesamt eher kurzen Texten eine besondere Rolle spielen, zumal kürzere Texte allgemein einheitlicher benotet werden als dies bei längeren Texten der Fall ist (Baurmann, 2006, S. 144). Ob sich in dieser Einheitlichkeit nun aber primär eine unbewusste Beeinflussung der Benotung durch die Textlänge im Sinne einer Fehlerquelle äußert, wie sie als grundsätzliche Möglichkeit von Kötter und Grau (1965) zusätzlich zu einem realen Zusammenhang zwischen Textlänge und -qualität geäußert wird, darf wohl bezweifelt werden. So wird im Zusammenhang mit den Kompetenzstufen beim argumentativen Schreiben für die Sekundarstufe auch für sehr kurze Texte ein durchaus realer Zusammenhang zwischen Textlänge und Schreibkompetenz formuliert, wenn für diejenigen Schülerinnen und Schüler, welche den Mindeststandard verfehlen, konstatiert wird, dass sie „in der Regel so kurze Texte [schreiben], dass der argumentative Duktus nur ansatzweise erkennbar ist“ (KMK, 2014, S. 12). Gerade bei den besonders kurzen Texten ist also in diesem Sinne zu erwarten, dass sie tatsächlich nur eine geringe Textqualität entfalten können, welche sich wiederum speziell bei einer kriterialen Auswertung darin bemerkbar machen dürfte, dass einzelne Merkmale (z. B. Nebensätze) überhaupt erst ab einer bestimmten Textlänge auftreten können.

Da allerdings die Textqualität, welche ja faktisch als Bewertungsgrundlage dient, als Maß für die ihr zugrunde liegende Schreibkompetenz genutzt wird, muss hier sicherlich berücksichtigt werden, dass sehr kurze Texte besonders stark durch motivationale Aspekte wie Unlust oder Verweigerung bedingt sein können. Dass auch die Informationsgrundlage für eine Bewertung der jeweili-

gen Texte dabei zwangsläufig gering ausfällt, stellt darüber hinaus eine zuverlässige Differenzierbarkeit zwischen den sehr kurzen Texten, insbesondere bei der Bewertung der zugrunde liegenden Schreibkompetenz einer einzelnen Person, in Frage. Dass sie jedoch gegenüber den längeren Texten generell schwächer bewertet werden (Baurmann, 2006), erscheint gerade bei statistischen Analysen über ganze Stichproben durchaus angemessen.

Dabei dürfte es sich in der vorliegenden Studie im Hinblick auf die besonders kurzen Texte durchaus als Vorteil erweisen, dass die Bewertungsgrundlagen an eine auch durchschnittlich eher knappe Textlänge angepasst werden mussten.

5. Modellierung der analytischen Schreibkompetenz in NaSch 1

Für die Beurteilung von Schülertexten bzw. die Entwicklung geeigneter Beurteilungskriterien ist die Ausrichtung an der spezifischen Aufgabenstellung sowie an den individuellen Voraussetzungen der Schreiber notwendig (Blatt et al., 2005, S. 111). Daher wurde in NaSch 1 ein vereinfachtes Modell der Textqualität entwickelt. Diesem liegt ein Textbegriff zugrunde, der sich an linguistisch-pragmatischen Ansprüchen orientiert. Ein rein linguistisches Verständnis von Text als beispielsweise „monologische geschriebene sprachliche Äußerung von mehreren Sätzen Länge“ (Nussbaumer, 1991, S. 33) ist vor dem Hintergrund des Entwicklungsstandes der Erstklässler wenig sinnvoll. Das Verständnis von Text als abgeschlossene sprachliche Äußerung wurde daher pragmatisch erweitert, indem ein Schreibprodukt bereits dann als Text bewertet wurde, wenn eine für den Adressaten verständliche Kommunikationsabsicht bestehend aus mindestens einer Subjekt-Prädikat-Beziehung erkennbar war. Auch bei der deduktiv-induktiven Entwicklung der Auswertungskriterien nahmen die wenig ausgeprägten Schreiberfahrungen sowie das Schreibsetting einen wichtigen Bezugspunkt ein. Das Kriterium *Textidee und -entfaltung* lässt sich beispielsweise als eine Vorstufe zu den Kategorien *Gesamtidee, Thema, Absicht des Textes; Aufbau, Gliederung; thematische Entfaltung* aus dem Zürcher Textanalyseraster (Nussbaumer & Sieber 1995, S. 45–46) verstehen. Es zielt dabei auch auf die Nachvollziehbarkeit des Textsinns ab, der sich durch die logische Verknüpfung von Sätzen ergibt und als ein Aspekt der Textverständlichkeit entscheidend für die Bewertung der Kohärenz ist. Bei nur einem Satz wurde die Proposition, also die Satzaussage, als Bewertungsgrundlage herangezogen. Bei dem Item *Kohäsionsmittel* stand dagegen weniger die Nachvollziehbarkeit des Textsinns im Fokus, sondern die in einem Text eingesetzten „Sprachmittel des Textbaus“ (Nussbaumer & Sieber, 1995, S. 47), wie Verweismittel oder Verknüpfungsmittel bzw. Konnektive. Die syntaktische Komplexität wurde in Anlehnung an Steinig, Betzel, Geider und Herbold (2009) anhand der Anzahl der Nebensätze analysiert, was im vorliegenden Fall darauf eingeschränkt wurde, ob überhaupt Nebensätze verwendet wurden.

Auch die sprachsystematischen Aspekte wie Orthografie und Grammatik wurden nicht über die Anzahl identifizierbarer Fehler operationalisiert, sondern im Hinblick auf die Einschränkung der Lesbarkeit durch grammatikalische oder orthografische Fehler bewertet, wobei die lautgetreue Verschriftlichung als Bezugspunkt für die Orthografie fungierte (Kürzinger, Lotz, Gleich & Kempster, 2013, S. 284). Einzelne Aspekte ließen sich hingegen nicht zufriedenstellend an die vorliegenden Texte anpassen. So konnte das Item *Expliztheit*, das die Langatmigkeit und Redundanz eines Textes erfasst (Nussbaumer & Sieber, 1994), aufgrund des begrenzten Umfangs der Texte kaum angewendet werden und wurde deshalb auch nicht in die Skalierung des analytischen Ratings mit aufgenommen.

Insgesamt gingen 15 Items in die Textqualitätsanalyse ein. Diese lassen sich in *formal-strukturelle* (Anrede- und Schlussformel, Textende), *inhaltliche* (Perspektivenübernahme, Orientierung am Bilderbuch), *sprachliche* (Wortschatz, Angemessenheit der Sprachmittel, Kohäsionsmittel, Impliztheit, Textidee und -entfaltung, syntaktische Komplexität) und *sprachsystematische* Kriterien (Satzkonstruktion, Orthografie, Grammatik, Leserlichkeit) einteilen. Die Items wurden entweder niedrig, mittel oder hoch inferent erhoben.³ Nicht mit einbezogen wurden Items, welche sich überwiegend auf inhaltliche Aspekte (z. B. Hänseln, Andersartigkeit, Heimweh etc.) bezogen, da sie zwar interessante Informationen über die entstandenen Texte liefern, aber keine Qualitätseinschätzung zulassen. Auch Items, welche eindeutig mit anderen Items konfundiert waren, wurden bei der Skalierung nicht verwendet.

Die im vorliegenden Schreibsetting an die Schülerinnen und Schüler gestellten Anforderungen resultieren einerseits aus der konkreten Aufgabenstellung (perspektivisches Schreiben, Textsorte persönlicher Brief), die mit den Schülerinnen und Schülern in derselben Unterrichtseinheit besprochen und inhaltlich vorbereitet wurde. Andererseits besteht die Herausforderung für die Schreibanfänger in der Textproduktion selbst. Einen Text zu produzieren bedeutet für die Schülerinnen und Schüler in der für sie neuen Kommunikationssituation, in der kein Gesprächspartner mehr anwesend ist, „selbständig Gedanken aufzurufen, zu strukturieren und ihnen einen angemessenen sprachlichen Ausdruck zu geben“ (Portmann, 1996, S. 158–159). Die kommunikative Herausforderung beim Schreiben beginnt für sie also nicht erst beim Formulieren einer Schreibabsicht, beim Ideengenerieren oder beim Aufrechterhalten des Diskurses, sondern besteht bereits aufgrund des „schriftsprachliche(n) Charakteristikum(s) der Partnerferne“ (Weinhold, 2002, S. 149; vgl. auch Bereiter & Scardamalia, 1987).

Aufgrund der Herausforderung für die Schülerinnen und Schüler, einerseits die Vorgaben der Schreibaufgabe umzusetzen und andererseits einen sprachlich angemessenen und kohärenten Text zu produzieren, wurde zunächst von einer zweidimensionalen Struktur des aufgabenspezifischen Faktors sensu Lehmann (1987, 1990) ausgegangen, wobei die Idee aufgabenbezogener Anforderungen enger defi-

3 Eine ausführliche Beschreibung und Erläuterung der Items mit Indikatoren und Benchmarktexten finden sich im Kriterienkatalog zur analytischen Bewertung der Texte (Kürzinger & Pohlmann-Rother, 2014).

niert und auf inhaltliche und formale Aspekte, die im Unterricht als Vorgabe zum Schreiben formuliert worden waren, eingeschränkt wurde. Daneben wurde eine eigene sprachliche Dimension vermutet, welche insbesondere pragmatische (z. B. Wortschatz) und sprachliche Kriterien (z. B. syntaktische Komplexität) umfasste. Diese Struktur ließ sich aber faktisch kaum untersuchen, da vier der fünf aufgabenbezogenen Anforderungen Lösungshäufigkeiten über 90 % aufwiesen, wodurch die Möglichkeit, eine eigene aufgabenbezogene Dimension zu bilden, generell in Frage gestellt wurde.

Somit wird in der vorliegenden Untersuchung nur der Frage nachgegangen, inwiefern sich die empirisch wiederholt vorgefundene Aufteilung (vgl. Kapitel 3) in eine sprachsystematische Dimension (oder *mechanics-factor*) und eine semantisch-pragmatische (oder *aufgabenspezifische*) Dimension auch im ersten Schuljahr identifizieren lässt. Gerade die in Kapitel 4 dargestellten Ausführungen Weinholds (2002) lassen es zumindest als möglich erscheinen, dass die noch nicht vorhandenen Routinen der eher technischen Aspekte des Schreiben-Könnens eine von ihnen unabhängige Entfaltung sprachlicher Textqualität in diesem Schreibalter gar nicht unbedingt zulassen. Dies führt bei der Skalierung des analytischen Ratings zu einer Gegenüberstellung einer semantisch-pragmatischen Dimension, welche die formal-strukturellen, inhaltlichen und sprachlichen Kriterien umfasst, und einer sprachsystematischen Dimension, welche sich konkret aus Orthographie, Grammatik, Satzkonstruktion und Leserlichkeit zusammensetzt. Daraus ergibt sich die erste Fragestellung:

Lässt sich die Schreibkompetenz im ersten Schuljahr in die zwei Dimensionen semantisch-pragmatisch und sprachsystematisch unterteilen oder stellt sie sich eher als eine globale Kompetenz dar?

Auch für die holistische Bewertung der Texte (Gesamteindruck) waren Kriterien formuliert worden, welche formal-strukturelle, inhaltliche, sprachliche sowie sprachsystematische Aspekte explizit thematisierten (vgl. Kapitel 6). Somit sollte das holistische Urteil inhaltlich einer globalen Dimension des analytischen Ratings entsprechen bzw. sowohl die Aspekte einer semantisch-pragmatischen Dimension als auch die einer sprachsystematischen Dimension umfassen. Dies führt zur zweiten Fragestellung:

Welche Zusammenhänge zeigen sich zwischen dem holistischen Urteil (Gesamteindruck) und den latenten Teilfähigkeiten bzw. einem Generalfaktor der analytisch ermittelten Schreibkompetenz?

Schließlich soll auch der Frage nachgegangen werden, welche Rolle die Textlänge dabei spielt. Da diesbezüglich neben einem realen Zusammenhang zwischen Textlänge und Textqualität auch die Möglichkeit der Textlänge als Fehlerquelle bei der (holistischen) Benotung genannt wurde (Kötter & Grau, 1965), stellt sich zunächst die Frage, ob sich ein derartiger Einfluss im Sinne eines besonders starken Zusammenhangs zwischen Textlänge und holistischem Rating auffinden lässt. Bei den sehr kurzen Texten aber ist eher zu erwarten, dass gerade der Zusammenhang zwischen Textlänge und kriterialer Erfassung der Textqualität besonders deutlich ausgeprägt ist, da die Möglichkeit zur Erfüllung der einzelnen

Kriterien auch von einer jeweiligen Mindestlänge der Texte abhängt. Gleichzeitig soll allgemein der Frage nachgegangen werden, ob sich Hinweise dafür finden lassen, dass die Einbeziehung der besonders kurzen Sätze in die Analysen zu einer Verzerrung der allgemeinen Zusammenhänge führt. Daraus resultieren die letzten Fragestellungen:

- Zeigt sich eine Übereinstimmung zwischen holistischem Rating und analytischer Auswertung auch jenseits der (gemeinsamen) Textlänge?
- Wie stark schlägt sich die Textlänge allgemein in der Bewertung der Schreibkompetenz nieder?
- Zeigt sich dabei ein größerer Zusammenhang zum holistischen Rating als zur analytischen Auswertung? Dreht sich dieser Zusammenhang bei den besonders kurzen Texten um?
- Zeigen sich substantiell andere Zusammenhänge, wenn die besonders kurzen Texte in den Analysen nicht mitberücksichtigt werden?

6. Methoden – Daten und Analysen

6.1 Stichprobe

Die untersuchten Texte entstanden im Rahmen des PERLE-Videomoduls Deutsch im März 2007 in der ersten Klassenstufe (Schuljahr 2006/2007). Von der ursprünglichen Stichprobe von 618 Texten mussten zunächst 71 entfallen, da sie aufgrund ihrer Beschaffenheit (Text besteht z. B. lediglich aus Anrede oder ist nicht lesbar) anhand der vorgestellten Instrumente nicht weiter analysiert werden konnten. Sieben weitere Texte wurden ebenfalls aus der Bewertung ausgeschlossen, da sie auf Anordnung der Lehrkraft in Teamarbeit entstanden und sich somit nicht einer einzelnen Schülerin/einem einzelnen Schüler zuordnen lassen. Die endgültige Stichprobe von 540 Schreibprodukten enthält dabei Texte von 239 Schülern und 301 Schülerinnen.

6.2 Erhebungssituation

Das dem Schreibarrangement zugrunde liegende handlungs- und produktionsorientierte Verfahren bestand in dem Verfassen eines Textes (Textsorte persönlicher Brief) im Anschluss an die Rezeption eines Bilderbuchs. In diesem 90-minütigen videografierten unterrichtlichen Setting wurde u. a. das Bilderbuch vorgestellt und der Arbeitsauftrag direkt vor der Schreibphase erläutert, wobei teilweise Schreibideen gesammelt oder Briefmerkmale thematisiert wurden.

6.3 Auswerter

Für alle Ratings und Kodierungen wurden jeweils zwei studentische Hilfskräfte geschult, die die Auswertung der Texte zusammen mit dem Entwickler der Instrumente vornahmen. Bei den mehrtägigen Trainings wurde unter Einsatz einer hohen Anzahl von Übungstexten (ca. 100 Texte) auf ein gemeinsames theoretisches Verständnis der Kriterien fokussiert, was die Basis für eine objektive, reliable und valide Auswertung darstellt. Für die verschiedenen Instrumente wurden jeweils andere Hilfskräfte geschult und eingesetzt, um Konsistenztendenzen zu vermeiden (vgl. Böhme et al., 2009) und faktische Abweichungen zwischen den Instrumenten identifizieren zu können.

6.4 Holistisches Rating

Das holistische Rating wurde anhand der Originaltexte der Erstklässler vorgenommen. Es wurde in Anlehnung an den NAEP Holistic Scoring Guide (Persky, Daane & Jin, 2003) sowie das holistische Analyseinstrument von Böhme et al. (2009) entwickelt und entsprechend modifiziert. Aufgrund der geringen Varianz wurde statt der sechsstufigen eine vierstufige Skala verwendet. Als Bewertungsgrundlage für das holistische Rating wurden folgende Teilaspekte bzw. Indikatoren herangezogen (Kürzinger et al., 2013; vgl. auch Böhme et al., 2009):

- Erfüllung der Aufgabenstellung (Schreiben eines Briefs aus der Perspektive von Lucy an Mama Kroko);
- Textaufbau und Kohärenz;
- Lautgetreue Verschriftlichung und Leserlichkeit;
- Funktional angemessener und korrekter Sprachgebrauch;
- Thematisierung des Verlassens von Mama Kroko;
- Detailreichtum.

Der relative Generalisierbarkeitskoeffizient (g_{relativ}) des holistischen Ratings (Gesamteindruck) ($N = 500$) kann mit .91 eine hohe Übereinstimmung zwischen beiden Ratern nachweisen.⁴ Die Intraraterreliabilität (Messzeitpunkt nach zwei Monaten; $N = 100$) weist mit $g_{\text{relativ}} = .91$ beziehungsweise .96 für beide Rater ebenfalls hohe Generalisierbarkeitskoeffizienten auf (Kürzinger & Pohlmann-Rother, 2015). Das Vorgehen kann aufgrund des Generierens einer einzigen Gesamtbewertung anhand der Originaltexte unter bewusster Berücksichtigung definierter Teilaspekte einschließlich der sprachsystematischen Aspekte wohl als ähnlich zu einer Textbewertung durch Lehrkräfte gelten.

4 Die Kontrolle der Objektivität und Reliabilität orientiert sich an den in der PERLE-Studie festgelegten Mindestkennwerten, die bei der prozentualen Übereinstimmung (PÜ) $\geq 85.00\%$, bei Cohens Kappa (k) $\geq .70$ und beim relativen Generalisierbarkeitskoeffizienten (g) $\geq .70$ betragen (Lotz et al., 2013).

6.5 Analytische Auswertung

In die analytische Auswertung fließen einerseits die Bewertungen der Lesbarkeit im Hinblick auf Orthografie und Grammatik sowie die Leserlichkeit hinsichtlich des Schriftbildes und andererseits die Ratings und Kodierungen der übrigen zwölf analytischen Kriterien ein, welche anhand zweier unterschiedlicher Textvorlagen und unterschiedlicher Personen erstellt wurden. Die Bewertung der jeweiligen Lesbarkeit und der Leserlichkeit erfolgte wie das holistische Rating auf Grundlage der Originaltexte anhand vierstufiger Items (Kürzinger et al., 2013, S. 283–286).

Für die übrigen zwölf analytischen Kriterien sowie die meisten bei der Skalierung nicht berücksichtigten analytischen Kriterien wurden Reinschriften verwendet. Für diese wurden die 540 Texte in einem ersten Schritt buchstabengetreu transkribiert, wobei es entscheidend war, die Original-Schreibweise der Schülerinnen und Schüler so präzise wie möglich wiederzugeben. Für die anschließende Reinschrift wurden sodann Interpunktions- sowie Orthografiefehler korrigiert (ebd., 2013, S. 258 ff.). Diese Modifikation war der zum Teil schwer zu erlesenden Schrift und Schreibweise der Erstklässler geschuldet, um eine möglichst einheitliche und eindeutige Erfassung der einzelnen analytischen Merkmale zu gewährleisten und dabei mögliche Halo-Effekte zu vermeiden.

Die Kennwerte der fünf vierstufigen semantisch-pragmatischen Items weisen hohe Interraterreliabilitäten auf ($g_{\text{relativ}} = .97 - .99$) und bei den sieben dichotomen Items der analytischen Textauswertung zeigen sich ebenfalls zufriedenstellende Werte. So wurden bei der prozentualen Übereinstimmung Werte von 90,5 % – 99,4 % und $k = .75 - .97$ bei Cohens Kappa erreicht, die damit ebenfalls deutlich über den Mindestkennwerten liegen (Kürzinger & Pohlmann-Rother, 2015). Für die Intraraterreliabilität, die nach rund zwei Monaten berechnet wurde, ergeben sich bei den niedrig inferenten Items ($N = 67$) für Cohens Kappa Werte zwischen $k = .83 - 1.0$, bei der prozentualen Übereinstimmung Werte zwischen 90,2 % – 100 %. Für die hoch inferenten Items wurden Generalisierbarkeitskoeffizienten von $g_{\text{relativ}} = .95 - 1.0$ ermittelt.

Für die Analyse der jeweiligen Lesbarkeit/Leserlichkeit wurden ebenfalls zufriedenstellende Reliabilitätskennwerte von $g_{\text{relativ}} = .86 - .92$ erreicht. Ergebnisse zur Intraraterreliabilität liegen hier nicht vor.

6.6 Analysen

Zur Beantwortung der Frage, ob sich anhand der Kriterien die zwei Dimensionen *semantisch-pragmatischer Faktor* (bzw. *aufgabenspezifisch*) und *sprachsystematischer Faktor* (bzw. *mechanics*) ermitteln lassen oder sich eher eine eindimensionale Struktur der Schreibkompetenz abzeichnet, wurden entsprechende ein- und zweidimensionale Partial-Credit-Modelle mit ConQuest2 (Wu, Adams, Wilson & Haldane, 2007) berechnet. Diejenigen Kriterien, welche beim Rating selbst vierstufig erfasst worden waren, wurden auch vierstufig verwendet. Dies betrifft

fünf Merkmale des semantisch-pragmatischen Faktors (vgl. Kapitel 5) und drei Kriterien des sprachsystematischen Faktors. Die übrigen sieben Merkmale flossen als dichotome Items in die Skalierung ein.

Den Fragen, welche Zusammenhänge sich zwischen holistischem und analytischem Urteil (auch jenseits der Textlänge) zeigen, wird zunächst anhand bivariater und partieller Korrelationen (mit SPSS) nachgegangen. Da die Skalierung des analytischen Ratings tatsächlich zwei Subdimensionen nahelegt, welche inhaltlich beide im holistischen Rating vereint sein sollten, wurden darüber hinaus multiple Regressionen mit den beiden analytischen Skalen als Prädiktoren und der holistischen Skala als Kriterium berechnet. Um auch hier die Einflüsse der Textlänge eliminieren zu können (analog zu den partiellen Korrelationen), wurde diese Regression auch über die Residuen der drei Schreibkompetenzwerte berechnet, welche zuvor in einfachen Regressionen mit der Textlänge als Prädiktor und der jeweiligen Schreibkompetenz als Kriterium ermittelt worden waren. Die multiplen Regressionen wurden mit Mplus (Muthén & Muthén, 1998–2010) berechnet, um die Clusterung anhand der Klassenzugehörigkeit bei der Berechnung der Schätzfehler und Signifikanzangaben berücksichtigen zu können (type = complex).

Sowohl die Skalierung selbst als auch die Berechnung der Korrelationen und multiplen Regressionen wurden sowohl für den gesamten Datensatz als auch für einen reduzierten Datensatz, in welchem nur noch die Texte mit mindestens zehn Wörtern enthalten sind, durchgeführt, um zu ermitteln, ob sich substantiell abweichende Ergebnisse zwischen den unterschiedlichen Stichproben zeigen. Die Grenze von zehn Wörtern ergibt sich dabei aus dem insgesamt eher geringen durchschnittlichen Umfang der Texte und führt zum Ausschluss von genau einem Sechstel der Texte. Eine stärkere Reduzierung des Datensatzes aufgrund einer höheren Mindest-Textlänge hätte erwartungsgemäß auch zu einer stärkeren Reduzierung der tatsächlichen Streuung der erfassten Schreibkompetenzen geführt. Da dies bei Vorliegen linearer Zusammenhänge automatisch zu einer geringeren Ausprägung der untersuchten Zusammenhänge führen sollte, würde es auch die Interpretation der jeweiligen Ergebnisse erschweren.

7. Ergebnisse

7.1 Dimensionalität des Konstrukts Schreibkompetenz

Beide Partial-Credit-Modelle (eindimensional und zweidimensional) führen zu zufriedenstellenden oder sogar guten EAP/PV-Reliabilitäten (Tabelle 1). Diese betragen bei dem eindimensionalen Modell .838, bei dem zweidimensionalen Modell .857 (semantisch-pragmatisch) und .665 (sprachsystematisch). Auch bei einer getrennten Skalierung der beiden Subdimensionen als eindimensionale Modelle bleiben die jeweiligen EAP/PV-Reliabilitäten annähernd gleich (semantisch-pragmatisch: .862; sprachsystematisch: .640). Angesichts ei-

Tabelle 1: Kennwerte der Partial-Credit-Modelle für eine oder zwei Dimensionen, $N = 540$ (450)^a

Variable	Skalenkennwerte								
	Häufig in % ^b	eindimensional				zweidimensional			
		$\sigma_i^{a,c,d}$	Infit ^{a,e}	$\sigma_i^{a,c,f}$	Infit ^{a,g}				
Anrede	90.5	-2.67 (-2.83)	1.04 (1.05)	-3.13 (-3.15)	1.22 (1.20)				
Schlussformel	66.4	-0.82 (-1.14)	1.09 (1.13)	-0.99 (-1.29)	1.24 (1.28)				
Aufgabenstell.	96.5	-3.80 (-4.19)	0.97 (0.99)	-4.38 (-4.62)	0.96 (0.98)				
Perspektiven- übernahme	92.3	-2.90 (-2.93)	1.00 (1.01)	-3.39 (-3.26)	1.16 (1.10)				
Textende	92.9	-2.99 (-3.33)	0.99 (1.00)	-3.49 (-3.70)	1.12 (1.11)				
Wortschatz	11 / 31 / 46 / 13	-0.15 (-0.69)	0.74 (0.75)	-0.22 (-0.83)	0.77 (0.82)				
Sprachmittel	12 / 32 / 47 / 8	0.13 (-0.19)	0.64 (0.68)	0.13 (-0.26)	0.71 (0.77)				
Kohäsionsmittel	27 / 37 / 20 / 16	0.35 (0.11)	1.05 (1.13)	0.48 (0.14)	1.17 (1.14)				
Implizitheit	10 / 25 / 33 / 31	-0.61 (-0.94)	1.02 (1.10)	-0.81 (-1.14)	1.20 (1.20)				
Textidee	16 / 30 / 33 / 21	-0.14 (-0.59)	0.69 (0.69)	-0.19 (-0.72)	0.67 (0.68)				
Syntaktische Komplexität	37.5	0.64 (0.36)	0.96 (0.99)	0.76 (0.40)	1.03 (1.01)				
Satzkonstruktion	69.8	-1.01 (-1.00)	0.93 (0.90)	-1.00 (-1.01)	0.90 (0.91)				
Leserlichkeit	6 / 21 / 45 / 28	-0.85 (-0.81)	1.43 (1.40)	-0.85 (-0.84)	1.13 (1.17)				
Orthografie	18 / 25 / 37 / 19	-0.08 (-0.07)	1.34 (1.22)	-0.07 (-0.08)	1.03 (1.00)				
Grammatik	4 / 11 / 32 / 54	-1.43 (-1.44)	1.08 (0.96)	-1.44 (-1.49)	0.92 (0.90)				

Reliabilität (EAP/PV): *eindimensional*: .838 (.834).

zweidimensional: sprachlich-semantisch: .857 (.839); sprachsystematisch: .665 (.686).

Anmerkungen. N = Anzahl. σ_i = Itemschwierigkeit. Infit = WMNSQ.

^a Angaben in Klammern = Stichprobe umfasst nur Texte mit mind. 10 Wörtern. ^b Bei dichotomen Items Lösungshäufigkeit, bei vierstufigen Items Belegung jeder Stufe (0 bis 3). ^c Mittlere Personenfähigkeit = 0. ^d Für vierstufige Items nur Angaben zur mittleren Aufgabenschwierigkeit, Streuung der Schwierigkeiten der einzelnen Thresholds über alle acht Items: -1.83 bis 2.34 (bzw. -2.23 bis 2.55). ^e Für vierstufige Items nur Item-WMNSQ, WMNSQ-Streuung für einzelne Stufen: 0.74–1.24 (bzw. 0.77–1.22). ^f Für vierstufige Items nur Angaben zur mittleren Aufgabenschwierigkeit, Streuung der Schwierigkeiten der einzelnen Thresholds: über alle fünf semantisch-pragmatischen Items: -2.28 bis 2.99 (bzw. -2.68 bis 3.02), über alle drei sprachsystematischen Items -1.47 bis 1.72 (bzw. -1.36 bis 1.68). ^g Für vierstufige Items nur Item-WMNSQ; WMNSQ-Streuung für einzelne Stufen: *sprachlich semantisch*: 0.80–1.14 (bzw. 0.77–1.15), *sprachsystematisch*: 0.92–1.11 (bzw. 0.94–1.10).

ner vorgegebenen mittleren Personenfähigkeit = 0 kann bei einem Blick auf die mittleren Aufgabenschwierigkeiten in Tabelle 1 der Eindruck entstehen, dass der obere Fähigkeitsbereich über alle Items nur marginal erfasst wird. Dabei ist aber zu berücksichtigen, dass es sich bei insgesamt acht Kriterien um vierstufige Items handelt, bei denen die Schwierigkeiten der einzelnen Thresholds zum Zuge kommen (aber aus Platzgründen nicht im Einzelnen aufgeführt werden): Somit erstrecken sich die tatsächlich erfassten Aufgabenschwierigkeiten beim eindimensionalen Modell von $\sigma_i = -3.80$ bis $\sigma_i = 2.34$, bei der semantisch-pragmatischen Subdimension von $\sigma_i = -4.38$ bis $\sigma_i = 2.99$ und bei der sprachsystemati-

schen Dimension von $\sigma_i = -1.47$ bis $\sigma_i = 1.72$, wobei bei der sprachsystematischen Dimension nicht nur der Range der Aufgabenschwierigkeiten, sondern auch der der Personenfähigkeiten geringer ausfällt als bei den beiden anderen Skalen.

Bei der Entscheidung zwischen beiden Modellen sprechen zunächst einmal die vergleichsweise niedrige latente Korrelation ($r = .49$) zwischen den beiden Subdimensionen des zweidimensionalen Modells und dessen niedrigere Informationskriterien (BIC = 12648; AIC = 12288; CAIC = 12468 versus BIC = 13049; AIC = 12710; CAIC = 12879 beim eindimensionalen Modell) für eine Zweidimensionalität. Auch der Umstand, dass bei dem eindimensionalen Modell zwei Infit-Werte (Leserlichkeit: WMNSQ = 1.43; Orthografie: WMNSQ = 1.34) oberhalb des angestrebten Bereichs von 0.75 bis 1.33 (Adams & Khoo, 1996) liegen, während sie im zweidimensionalen Modell vergleichsweise geringfügig von Eins abweichen (Leserlichkeit: WMNSQ = 1.13; Orthografie: WMNSQ = 1.03), unterstützt diese Einschätzung. Dass die beiden Merkmale Sprachmittel und Textidee in beiden Modellen unterhalb dieses Bereichs liegen (Tabelle 1), erscheint hingegen noch als akzeptabel, da es sich hier um einen Hinweis auf eine besonders hohe Trennschärfe dieser Merkmale handelt. Auch der Umstand, dass im eindimensionalen Modell zusätzlich der WMNSQ des Kriteriums Wortschatz etwas zu niedrig ausfällt, kann kaum als substantieller Unterschied zwischen beiden Modellen gewertet werden.

Werden die Skalierungen nur über die 450 Texte mit mindestens zehn Wörtern ausgeführt (Tabelle 1, Angaben in Klammern), so lassen sich die Unterschiede zur Gesamtstichprobe als insgesamt eher geringfügig ansehen. Die deutlichsten Abweichungen stellen eine höhere latente Korrelation zwischen den beiden Subdimensionen ($r = .58$) und der Umstand dar, dass der Infit-Wert der Orthografie (WMNSQ = 1.22) im eindimensionalen und der der Sprachmittel (WMNSQ = 0.77) im zweidimensionalen Modell diesmal innerhalb des angestrebten Bereichs liegen. Insgesamt aber sprechen auch hier die Ergebnisse für die zweidimensionale Lösung.

Für alle weiteren Analysen werden die WLE-Personenschätzer verwendet, welche über die zwei getrennten, eindimensionalen Skalierungen der Subdimensionen über die Gesamtstichprobe ermittelt wurden.

7.2 Vergleich analytische vs. holistische Auswertung und die Rolle der Textlänge

Da auch im Weiteren die Frage, wie die Bewertung sehr kurzer Texte (mit weniger als 10 Wörtern) einzuschätzen ist, eine Rolle spielen wird, erscheint es angebracht, zunächst einige grundlegende deskriptive Angaben zu betrachten (Tabelle 2), welche für eine Einordnung der weiteren Ergebnisse hilfreich erscheinen. Für eine bessere Übersichtlichkeit und Vergleichbarkeit werden hierfür z-standardisierte Werte ($M = 0$, $SD = 1$) genutzt.

Tabelle 2: Deskriptive Angaben (z-standardisiert, $M = 0$, $SD = 1$)

Textlänge	holistisch		semantisch-pragmatisch		sprach-systematisch		Textlänge	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
< 10 Wörter ($N = 90$)	-0.91	0.55	-1.10	0.59	0.02	1.03	-1.02	0.14
≥ 10 Wörter ($N = 450$)	0.18	0.97	0.22	0.92	-0.00	0.99	0.20	0.97

Anmerkungen. *M* = Mittelwert. *SD* = Standardabweichung. Semantisch-pragmatisch = semantisch-pragmatische Dimension. Sprachsystematisch = sprachsystematische Dimension

Hier sind vor allem zwei Aspekte auffällig: So liegt die durchschnittliche holistische und semantisch-pragmatische Bewertung der Texte mit weniger als 10 Wörtern ungefähr eine Standardabweichung unter der durchschnittlichen Bewertung aller Texte, was mit der durchschnittlichen Textlänge dieser Texte übereinstimmt. Während allerdings die Standardabweichung der Textlängen in diesem Bereich nur $SD = 0.14$ beträgt, beträgt sie bei den beiden Textbewertungen mehr als die Hälfte der Standardabweichung aller Texte ($SD = 0.55$ bzw. 0.59), die Streuung der Textbewertungen in diesem Bereich fällt also deutlich stärker aus als die der Textlängen. Bei der sprachsystematischen Dimension hingegen ergibt sich gar kein nennenswerter Unterschied zwischen den Mittelwerten und Standardabweichungen der Texte mit weniger als oder mindestens 10 Wörtern (Tabelle 2), sodass beide auch den Werten für die gesamte Stichprobe entsprechen.

Was nun die Korrelationen zwischen den verschiedenen Textbewertungen betrifft (Tabelle 3), so zeigt sich zunächst über alle Texte mit einem $r = .78$ ein sehr enger bivariater Zusammenhang zwischen der semantisch-pragmatischen Subdimension und dem holistischen Rating, welcher bei den Texten mit mind. 10 Wörtern nur geringfügig schwächer ausfällt ($r = .73$) und selbst bei den Texten mit weniger als 10 Wörtern immer noch als stark zu bezeichnen ist ($r = .63$). Auch der bivariate Zusammenhang zwischen der sprachsystematischen Subdimension und dem holistischen Rating erweist sich über alle Texte ($r = .47$) und bei Texten mit mind. 10 Wörtern ($r = .54$) als stark und bei weniger als zehn Wörtern ($r = .40$) als mittel bis stark.

Insgesamt deutet sich hier zunächst einmal eine etwas stärkere Übereinstimmung des holistischen Ratings mit der semantisch-pragmatischen als mit der sprachsystematischen Subdimension des analytischen Ratings an. Auffällig sind dabei die erstaunlich hohen Übereinstimmungen bei den sehr kurzen Texten mit weniger als 10 Wörtern.

Was nun die Textlänge betrifft, so zeigen sich über alle Texte starke Zusammenhänge zum holistischen Rating ($r = .62$) und zur semantisch-pragmatischen Dimension ($r = .63$), welche bei den Texten mit mindestens 10 Wörtern numerisch etwas geringer, aber immer noch stark ausfallen ($r = .55$ und $r = .54$). Ein stärkerer Zusammenhang mit dem holistischen Rating ist hier offensichtlich nicht gegeben. Vielmehr zeigt sich bei den Texten mit weniger als 10 Wörtern ein immer

noch mittlerer bis starker Zusammenhang der Textlänge mit dem semantisch-pragmatischen Rating ($r = .41$), während zum holistischen Rating nur noch ein schwacher Zusammenhang ($r = .16$) besteht, welcher auch nicht signifikant wird. Hierin könnte sich der Umstand bemerkbar machen, dass bei der isolierten Erfassung einzelner Merkmale gerade im Bereich der sehr kurzen Texte die Textlänge für das Auftreten einzelner Merkmale eine entscheidende Rolle spielt. Demgegenüber fällt der Zusammenhang zwischen Textlänge und sprachsystematischer Dimension grundsätzlich schwach aus ($r = .09$, $r = .12$ und $r = -.14$) und wird bei den Texten mit weniger als 10 Wörtern auch nicht signifikant.

Tabelle 3: Korrelationen (bivariat unterhalb Diagonale / partiell – Prädiktor Textlänge – oberhalb Diagonale)

	alle Texte, N = 540			mind. 10 Wörter, N = 450			< 10 Wörter, N = 90		
	1	2	3	1	2	3	1	2	3
1 holistisch	–	.64***	.52***	–	.62***	.56***	–	.63***	.43***
2 semantisch-pragmatisch	.78***	–	.36***	.73***	–	.41***	.63***	–	.45***
3 sprachsystematisch	.47***	.34***	–	.54***	.41***	–	.40***	.35**	–
4 Textlänge	.62***	.63***	.09*	.55***	.54***	.12**	.16	.41***	-.14

Anmerkungen. Semantisch-pragmatisch = semantisch-pragmatische Dimension.
Sprachsystematisch = sprachsystematische Dimension.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Nach einem Herauspartialisieren der Textlänge (Tabelle 3, oberhalb der jeweiligen Diagonale) zeigt sich zwischen dem holistischen Rating und der semantisch-pragmatischen Dimension unabhängig davon, welche Texte berücksichtigt werden, ein nahezu gleichbleibender starker Zusammenhang ($r_{\text{part}} = .62$ – $.64$), während der Zusammenhang zwischen sprachsystematischer Dimension und holistischem Rating bei den Texten mit weniger als 10 Wörtern numerisch etwas schwächer ausfällt ($r_{\text{part}} = .43$) als bei den Texten mit mindestens 10 Wörtern ($r_{\text{part}} = .56$).

Insgesamt lässt sich festhalten, dass auch jenseits einer gemeinsamen Beeinflussung durch die Textlänge beide Dimensionen der analytischen Auswertung deutliche Zusammenhänge mit dem holistischen Rating aufweisen, welche bei den Texten mit mindestens 10 Wörtern annähernd gleich stark ausfallen und für beide Dimensionen bei den Texten mit weniger als 10 Wörtern als erstaunlich hoch ausgeprägt gelten können.

Da das holistische Rating inhaltlich sowohl die Aspekte der semantisch-pragmatischen Dimension als auch die der sprachsystematischen Dimension umfassen sollte, werden zuletzt noch die Ergebnisse der multiplen Regressionen dargestellt, bei denen die beiden analytischen Dimensionen zur Vorhersage des holistischen Ratings herangezogen werden.

Wie aus den Korrelationen schon vorhersehbar, erweist sich dabei die semantisch-pragmatische Schreibkompetenz als stärkerer Prädiktor für die holistisch erfasste Schreibkompetenz als die sprachsystematische Schreibkompetenz (Tabelle 4). Dieser Unterschied fällt bei dem ersten Modell (über *alle Texte*) besonders deutlich aus ($\beta = .70$ vs. $\beta = .23$). Auch das Inkrement der sprachsystematischen Schreibkompetenz gegenüber der alleinigen Vorhersage der holistischen Werte durch die semantisch-pragmatische Schreibkompetenz wird zwar signifikant, leistet aber mit 4.6 % zusätzlich erklärter Kriteriums-Varianz einen eher geringen zusätzlichen Beitrag zu der gemeinsam erklärten Kriteriums-Varianz von 65.2 %.

Tabelle 4: Regressionen zur Vorhersage der holistisch ermittelten Schreibkompetenz durch die beiden analytisch ermittelten Dimensionen der Schreibkompetenz

Prädiktoren	alle Texte		≥ 10 Wörter		< 10 Wörter		Residuen	
	β	SE	β	SE	β	SE	β	SE
Semantisch-pragmatisch	.70***	.026	.62***	.032	.56***	.078	.52***	.038
Sprachsystematisch	.23***	.035	.29***	.040	.21*	.094	.33***	.040
R^2 in %	65.2***		60.7***		43.4***		50.2***	
ΔR^2 in % ^a	4.6***		6.8***		3.7*		9.6***	
	N = 540		N = 450		N = 90		N = 540	

Anmerkungen. β = standardisierte Regressionskoeffizienten. SE = Schätzfehler. semantisch-pragmatisch = semantisch-pragmatische Dimension. sprachsystematisch = sprachsystematische Dimension.

^a Inkrement der sprachsystematischen Schreibkompetenz gegenüber dem (alleinigen) Prädiktor semantisch-pragmatische Schreibkompetenz. Signifikanz-Angaben für Inkremente aus hierarchischer Regression mit SPSS.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Erwartungsgemäß fällt die gemeinsame Vorhersage der holistisch erfassten Schreibkompetenz bei den Texten mit weniger als zehn Wörtern (< 10 Wörter) mit $R^2 = 43.4$ % am niedrigsten aus. Als interessant erweisen sich die Ergebnisse des letzten Modells (*Residuen*), bei dem für beide Prädiktoren und das Kriterium die Residuen aus einer jeweils einfachen Regression mit dem Prädiktor Textlänge genutzt wurden, sodass die jeweiligen Übereinstimmungen mit der Textlänge aus allen drei Variablen vollständig herausgerechnet worden waren. Bei Verwendung dieser Residuen ergibt sich der geringste Unterschied zwischen dem semantisch-pragmatischen Vorhersageanteil ($\beta = .52$) und dem sprachsystematischen Vorhersageanteil ($\beta = .33$), was sich auch im höchsten Inkrement der sprachsystematischen Schreibkompetenz niederschlägt, welche hier immerhin einen zusätzlichen Beitrag von 9.6 % zu den insgesamt 50.2 % an Varianzaufklärung leistet. Da die inhaltliche Bedeutung der Übereinstimmungen der jeweiligen Schreibkompetenzbeurteilungen mit der Textlänge durchaus zwiespältig er-

scheint, wird durch ihr vollständiges Herausrechnen aber möglicherweise auch ein Teil des tatsächlichen Zusammenhangs zwischen den verschiedenen Werten für Schreibkompetenz eliminiert. Im Sinne dieser Überlegung stellen die Ergebnisse des Regressionsmodells über die Texte mit mindestens zehn Wörtern (≥ 10 Wörter) die vielleicht beste Annäherung an die tatsächlichen Zusammenhänge zwischen den beiden analytischen Subdimensionen und der holistischen Bewertung dar ($\beta = .62$ bzw. $\beta = .29$ sowie $R^2 = 60.7\%$), da hier die Übereinstimmungen mit der Textlänge zwar eingeschränkt (siehe auch Tabelle 3), aber nicht vollständig gelöscht sind.

8. Diskussion und Ausblick

Im vorliegenden Beitrag wurde zunächst der Frage nachgegangen, ob für Schreibkompetenz schon im ersten Schuljahr eine zweidimensionale Struktur ermittelt werden kann, bei der sich die technische Komponente des Schreiben-Könnens (sprachsystematisch) von den sprachlich-inhaltlichen Aspekten (semantisch-pragmatisch) trennen lässt. Während eine derartige Struktur für ältere Schülerinnen und Schüler in verschiedenen Studien nachgewiesen werden konnte (vgl. Kapitel 3), werfen insbesondere die Annahmen Weinholds (2002), dass die technischen Anforderungen des Schreibens in diesem frühen Stadium des Schriftspracherwerbs eine besonders große Herausforderung darstellen, die Frage auf, ob eine derartige Differenzierung der beiden Komponenten zu diesem Zeitpunkt schon möglich ist. Tatsächlich deuten die Ergebnisse der Rasch-Skalierung der analytischen Kriterien auf das Vorliegen einer derartigen Zweidimensionalität. Diese kann im Hinblick auf weitergehende Analysen als klarer Zugewinn gegenüber einer eindimensionalen holistischen Bewertung der Textqualität angesehen werden. Auch wird hier mit Blick auf die Förderung der Schreibentwicklung deutlich, dass Schülerinnen und Schüler im Bereich des Textverfassens verschiedene Teilkompetenzen erwerben müssen, die nur durch die Fokussierung der bewussten Auseinandersetzung mit den Anforderungen auf semantisch-pragmatischer und sprachsystematischer Ebene erreicht werden können (vgl. auch Neumann, 2007). Insofern haben isolierte Übungen zu Wortfeldern oder zum Erwerb orthographischer Kompetenzen ebenso wie ganzheitlich funktional-integrative Konzepte zur Schreibförderung ihre Berechtigung.

Als möglicher Kritikpunkt bei den Analysen ist der Umstand anzusehen, dass drei der vier Kriterien der sprachsystematischen Dimension auf einer anderen Textgrundlage (ursprünglicher Text statt Reinschrift) basieren und von anderen Auswertern erfasst wurden als die übrigen analytischen Kriterien. Dieses Vorgehen war zwar bewusst gewählt worden, da die meisten Kriterien der analytischen Auswertung anhand der Reinschrift erhoben werden sollten, um eine Beeinträchtigung und unbewusste Beeinflussung dieser Auswertung durch die jeweilige Lesbarkeit der Texte zu vermeiden. Gleichzeitig aber wurden die Kriterien

Leserlichkeit, Orthografie und Grammatik, welche zwangsläufig anhand der ursprünglichen Texte erfasst werden mussten, zu diesem Zeitpunkt nicht als integraler Bestandteil des analytischen Ratings der Textqualität bzw. Schreibkompetenz angesehen, sondern zunächst als ergänzende Informationen zusätzlich zur analytischen Erfassung der Schreibkompetenz verstanden. Dass sie angesichts der Ergebnisse früherer Studien (vgl. Kapitel 3) tatsächlich notwendig waren, um eine sprachsystematische Dimension abbilden zu können, wurde hingegen erst im Nachhinein realisiert. Dadurch wurde ein systematischer Vergleich, welche Rolle insbesondere der Umstand unterschiedlicher Auswerter bei der jeweiligen Auswertung spielen könnte, nicht mit eingeplant. So kann an dieser Stelle nur darauf verwiesen werden, dass das vierte Kriterium (Satzkonstruktion) der sprachsystematischen Dimension anhand der Reinschriften und von den entsprechenden Auswertern erhoben worden war, sodass sich hier zumindest ein Hinweis darauf ergibt, dass die ermittelte Zweidimensionalität der formulierten Definition entspricht.

Was den Abgleich der analytisch erfassten Schreibkompetenz mit der holistisch erfassten Schreibkompetenz betrifft, zeigt sich zunächst eine besonders hohe Übereinstimmung zwischen der semantisch-pragmatischen Dimension und den holistischen Werten. Auch die sprachsystematische Dimension weist dabei einen mittleren bis hohen Zusammenhang mit der holistischen Bewertung auf, in der sie ja inhaltlich auch mit enthalten sein soll. Tatsächlich nähert sich das Ausmaß der Übereinstimmung beider analytischer Dimensionen mit der holistischen Bewertung einander an, sobald die Textlänge vollständig herauspartialisiert wird. Diese wiederum weist eine gleichermaßen hohe Übereinstimmung mit der semantisch-pragmatischen Schreibkompetenz und dem holistischen Rating auf, sodass sie nicht als eine für das holistische Rating spezifische Fehlerquelle angesehen werden kann.

Auffallend ist an dieser Stelle allerdings, dass sich zwischen der Textlänge und der sprachsystematischen Schreibkompetenz kaum ein Zusammenhang auffinden lässt. Ob sich daraus tatsächlich der Schluss ziehen lässt, dass die primär technischen Schreibkompetenzen keinen Einfluss auf die Textlänge haben, kann aber zumindest in einer Hinsicht in Frage gestellt werden: So ist es durchaus vorstellbar, dass manch ein kurzer, weitgehend „fehlerfreier“ Text nur wegen seiner Kürze fehlerfrei blieb, dass also die geringe Textlänge das Ergebnis einer erfolgreichen Fehlervermeidung darstellt. Für eine Überprüfung dieser Annahme wäre allerdings ein Abgleich mit schreibtechnischen Fertigkeiten, welche über einen standardisierten Test erhoben wurden und somit keine individuelle Fehlervermeidung im zuvor beschriebenen Sinne ermöglichen, notwendig. Anhand der vorliegenden Ergebnisse hingegen muss konstatiert werden, dass ein konkreter Hinweis auf einen Einfluss der sprachsystematischen Schreibkompetenz auf die Textlänge nicht vorliegt.

Bei der Frage, wie die Zusammenhänge zwischen der Textlänge und den beiden Skalen semantisch-pragmatischer und holistisch erfasster Schreibkompetenz zu bewerten sind, müssen verschiedene Aspekte berücksichtigt werden. So lässt sich die Textlänge einerseits als Fehlerquelle ansehen, welche zu einer verzerrten Bewertung des jeweiligen Textes beitragen kann. Andererseits legt gerade der

Umstand, dass die vorliegenden Texte allgemein einen eher geringen Umfang aufweisen, den Gedanken nahe, dass hier auch eine inhaltliche Konfundierung vorliegen kann, bei der eine zunehmende Textlänge eine stärkere Entfaltung der Textqualität überhaupt erst ermöglicht und umgekehrt durch eine entsprechende Schreibkompetenz bedingt, also auch eingeschränkt werden kann. Wie weiter oben schon kurz angerissen wurde, zeigen sich über alle Texte hinweg sowie bei den Texten mit mindestens zehn Wörtern gleichermaßen hohe Zusammenhänge der Textlänge mit diesen beiden Skalen der Schreibkompetenz (holistisch und semantisch-pragmatisch), welche wiederum auch nach Herauspartialisieren der Textlänge eine sehr hohe Übereinstimmung miteinander aufweisen. Ein Hinweis darauf, dass die jeweilige Art der Bewertung (holistisch oder analytisch) eine unterschiedliche Anfälligkeit für eine Beeinflussung durch die Textlänge aufweist, ist also hier nicht gegeben. Gleichzeitig kommen beide Bewertungsarten offensichtlich auch jenseits der Textlänge zu sehr ähnlichen Urteilen, sodass der Anteil der Textlänge an der Gemeinsamkeit beider Skalen als eher sekundär angesehen werden kann. Hinweise auf spezifische Verzerrungen der Beurteilung der Textqualität durch die Textlänge sind hier zumindest nicht erkennbar.

Schwieriger gestaltet sich eine Einschätzung der sehr kurzen Texte mit weniger als zehn Wörtern, bei welchen sich generell die Frage stellt, ob sie valide und reliable Rückschlüsse auf die Schreibkompetenz zulassen oder aber gar nicht mit berücksichtigt werden sollten. Hier zeigt sich ein bedeutender Zusammenhang zwischen Textlänge und Schreibkompetenz nur noch bei der semantisch-pragmatischen Dimension, was sich wohl plausibel auf das isolierte Erfassen der einzelnen Kriterien zurückführen lässt. Da einige dieser Kriterien tatsächlich erst ab einer bestimmten Textlänge auftreten können (beispielsweise die Verwendung von Nebensätzen), könnten sich in diesem Bereich der Textlängen kritische Grenzen für einige der Kriterien befinden. Dies könnte den Unterschied zum holistischen Rating erklären, bei dem entsprechende Kriterien eher implizit berücksichtigt werden und welches bei den sehr kurzen Texten eine nur noch schwache und nicht signifikante Übereinstimmung mit der Textlänge aufweist. Dennoch fällt die partielle Korrelation zwischen beiden Bewertungen genauso stark wie bei den Texten mit mindestens zehn Wörtern und damit unerwartet hoch aus. Auch bewegen sich die Beurteilungen der Texte bei beiden Bewertungsarten auf einem ähnlich niedrigen Niveau und weisen in gleicher Weise eine wiederum erstaunlich hohe Streuung, verglichen mit der Streuung der Textlängen selbst, auf. Insgesamt deuten diese Ergebnisse darauf hin, dass sich auch bei diesen sehr kurzen Texten spürbare Unterschiede in der Textqualität jenseits der unterschiedlichen Textlängen ermitteln lassen, welche von beiden Arten der Bewertung mit hoher Übereinstimmung erfasst werden. Diese übereinstimmende Erfassung einer unerwartet großen Streuung von Ausprägungen der Textqualität bei den sehr kurzen Texten scheint auch der bisher schon beobachteten größeren Einheitlichkeit bei der Benotung kürzerer Texte (Baurmann, 2006, S. 144) zu entsprechen.

Dabei könnte sich tatsächlich der Umstand bemerkbar machen, dass die jeweiligen Beurteilungskriterien sowieso schon auf Texte geringen Umfangs zugeschnit-

ten waren. Gerade bei der analytischen Auswertung dürfte sich eine komplexere Operationalisierung der einzelnen Kriterien für die Bewertung deutlich längerer Texte wohl als anfälliger für sehr kurze Texte erweisen. Die ursprünglich geäußerte Annahme, dass zumindest die Erfassung der vorliegenden Textqualität auch bei den sehr kurzen Texten im vorliegenden Fall als möglich erachtet wird, scheint durch die vorliegenden Ergebnisse durchaus bestätigt zu werden. Ob allerdings eine Übertragung dieser Bewertung der Textqualität auf die ihr zugrunde liegende Schreibkompetenz gleichermaßen gerechtfertigt ist, lässt sich sicherlich stärker hinterfragen. Hier stellen Aspekte wie die geringe Informationsbasis und das relative Ausmaß motivationaler Einflüsse als mögliche Ursache für geringe Textlängen sicher eine gewichtigere Einschränkung der Zuverlässigkeit der Bewertungen dar.

Da die sehr kurzen Texte zu diesem sehr frühen Schreibalter allerdings auch das Ergebnis einer noch kaum entwickelten Schreibkompetenz darstellen können, stellt dieser Aspekt eine durchaus interessante Frage für zukünftige Untersuchungen mit Hilfe von Drittvariablen dar. Auch die Frage, inwiefern die Textlänge allgemein als Ergebnis der jeweiligen Schreibkompetenz angesehen werden kann oder aber als eher eigenständige Einschränkung der zustande gekommenen Textqualität fungiert, kann bei Berücksichtigung weiterer Informationen möglicherweise mehr Klärung erfahren, als es in dem vorliegenden Kontext möglich ist.

Dass es in der vorliegenden Untersuchung gelungen ist, bei der analytischen Auswertung eine zweidimensionale Struktur der Schreibkompetenz (semantisch-pragmatisch vs. sprachsystematisch) zu ermitteln, wie sie für ältere Schülerinnen und Schüler schon in verschiedenen Studien vorgefunden wurde, stellt aber schon an sich einen interessanten Befund dar. Diese zweidimensionale Erfassung der Schreibkompetenz ermöglicht darüber hinaus auch differenziertere weitergehende Untersuchungen unter Berücksichtigung der Ergebnisse aus Rechtschreib- und Leseverständnistests oder beispielsweise motivationaler Aspekte, als dies bei der Verwendung einer einzigen Gesamtbewertung im Sinne des holistischen Ratings möglich wäre. Dies gilt auch für weitere Analysen zum Zusammenhang zwischen Unterricht und Textqualität: Da die Texte nicht in einer Testsituation, sondern im Schreibunterricht entstanden sind (vgl. Kapitel 6), lassen sich Einflüsse des videografierten unterrichtlichen Settings (z. B. Qualität der Aufgabenstellung, Art und Inhalte der Lernunterstützung und des Feedbacks während der Schülerarbeitsphase) nicht nur auf die global erfasste Textqualität, sondern auch auf die beiden Teilkompetenzen der analytisch erfassten Schreibkompetenz aufdecken.

Mit der vorgenommenen Klassifikation der Briefe als Textsorte fungieren sie als „Repräsentationen des ihnen zugrunde liegenden Musters“ (Heinemann, 2000, S. 517). In einem Textmuster gehen eine bestimmte Textfunktion und „spezifische Ausprägungen von Textmerkmalen“ (Sandig, 2006, S. 487) eine Verbindung ein. Die Realisierung eines Textmusters (z. B. Brief) als „idealtypische[s] Muster“ (Heinemann, 2000, S. 518) impliziert die Verwendung „üblicher“ Elemente und bestimmter Formulierungen (z. B. Anrede, Verabschiedung). Inwieweit vor diesem Hintergrund die formal-strukturellen und sprachlich-inhaltlichen Merkmale, die

in der vorliegenden Analyse in die semantisch-pragmatische Dimension eingeflossen sind, jeweils zwei verschiedene Fähigkeitsdimensionen abbilden und dies möglicherweise eine begriffliche Differenzierung zwischen Textsorte und Textmuster erfordert, kann hier nicht geklärt werden. Dies stellt aber eine interessante Perspektive für weiterführende Untersuchungen dar, in denen im Vorhinein spezifische Teiltextmuster detaillierter zu identifizieren und operationalisieren sind.

Bezüglich der Schreibentwicklung der Schülerinnen und Schüler und vor dem Hintergrund vorliegender Schreibentwicklungsmodelle, die entweder von einem linearen (vgl. Feilke & Augst, 1989; Augst et al., 2007) oder parallelen Verlaufs der Schreibentwicklung (z. B. Becker-Mrotzek, 1997) ausgehen, dürfte es sich als interessant erweisen, ob und wie diese beiden Aspekte der Schreibkompetenz im Verlauf der weiteren Entwicklung miteinander interagieren, was im Rahmen der angestrebten Auswertung der narrativen Texte des zweiten und vierten Schuljahres im Folgeprojekt *Narrative Schreibkompetenzentwicklung in der Grundschule* (NaSch GS) längsschnittlich untersucht werden kann.

Literatur

- Adams, R. J. & Khoo, S. T. (1996). *Quest*. Melbourne: Australian Council for Educational Research.
- Augst, G., Disselhoff, K., Henrich, A., Pohl, T. & Völzing, P.-L. (2007). *Text – Sorten – Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter*. Frankfurt a. M.: Lang.
- Baurmann, J. (2006). *Schreiben – Überarbeiten – Beurteilen. Ein Arbeitsbuch zur Schreibdidaktik* (2. Aufl.). Seelze: Kallmeyer.
- Baurmann, J. & Pohl, T. (2009). Schreiben – Texte verfassen. In A. Bremerich-Vos, D. Granzer, U. Behrens & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule: Deutsch konkret* (S. 75–103). Berlin: Cornelsen.
- Beck, B., Bundt, S. & Gomolka, J. (2008). Ziele und Anlage der DESI-Studie. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch, Ergebnisse der DESI-Studie* (S. 11–25). Weinheim: Beltz.
- Becker-Mrotzek, M. (1997). *Schreibentwicklung und Textproduktion*. Opladen: Westdeutscher Verlag.
- 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: Barbara Budrich.
- Becker-Mrotzek, M. & Böttcher, I. (2011). *Schreibkompetenz entwickeln und beurteilen. Praxishandbuch für die Sekundarstufe I und II* (3. Aufl.). Berlin: Cornelsen.
- Becker-Mrotzek, M. & Schindler, K. (2007). Schreibkompetenz modellieren. In M. Becker-Mrotzek & K. Schindler (Hrsg.), *Texte schreiben* (Kölner Beiträge zur Sprachdidaktik 5, Reihe A, S. 7–26). Duisburg: Gilles & Franck.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.
- Birkel, P. & Birkel, C. (2002). Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 49(3), 219–224.
- Blatt, I., Ramm, G. & Voss, A. (2009). Modellierung und Messung der Textkompetenz im Rahmen einer Lernstandserhebung in Klasse 6. *Didaktik Deutsch*, 15(26), 54–81.

- Blatt, I., Voss, A. & Matthießen, I. (2005). Kinder schreiben Briefe aus der Zukunft. Qualitative Analysen von Kindertexten und fachdidaktische Diskussion. In W. Bos (Hrsg.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 109–158). Münster: Waxmann.
- Böhme, K., Bremerich-Vos, A. & Robitzsch, A. (2009). Aspekte der Kodierung von Schreibaufgaben. In D. Granzer, O. Köller, A. Bremerich-Vos, M. v. d. Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik* (S. 290–329). Weinheim: Beltz.
- Böttcher, I. (Hrsg.). (2010). *Kreatives Schreiben* (6., überarb. Aufl.). Berlin: Cornelsen.
- Bremerich-Vos, A. & Possmayer, M. (2011). Zur Reliabilität eines Modells der Entwicklung von Textkompetenz im Grundschulalter. *Didaktik Deutsch*, 17(31), 30–49.
- Brinker, K. (2010). *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden* (7. Aufl.). Berlin: Erich Schmidt.
- Dehn, M. (1996). Zur Entwicklung der Textkompetenz in der Schule. In P. R. Portmann & H. Feilke (Hrsg.), *Schreiben im Umbruch. Schreibforschung und schulisches Schreiben* (S. 172–185). Stuttgart: Klett.
- Dehn, M. (2000). Der andere Blick auf das Schreiben von Texten am Schulanfang. *Grundschule*, 32(7/8), 50–52.
- Dehn, M. (2006). Anforderungen an Aufgaben zum Textschreiben. *Grundschule*, 38(7/8), 39–42.
- Dehn, M., Merklinger, D. & Schüler, L. (2011). *Texte und Kontexte. Schreiben als kulturelle Tätigkeit in der Grundschule*. Seelze: Klett.
- Doucet, S. A. & Wilsdorf, A. (2005). *Lucy rettet Mama Krokodil*. Hamburg: Oetinger.
- Ehlich, K. (1984). Zum Textbegriff. In A. Rothkegel & B. Sandig (Hrsg.), *Text – Textsorten – Semantik. Linguistische Modelle und maschinelle Verfahren* (S. 9–25). Hamburg: Busche.
- Feilke, H. (1993). Schreibentwicklungsforschung. Ein kurzer Überblick unter besonderer Berücksichtigung der Entwicklung prozessorientierter Schreibfähigkeiten. *Diskussion Deutsch*, 24(129), 17–34.
- Feilke, H. (1996). Die Entwicklung der Schreibfähigkeiten. In H. Günther & O. Ludwig (Hrsg.), *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung* (S. 1178–1191). Berlin: Walter de Gruyter.
- Feilke, H. & Augst, G. (1989). Zur Ontogenese der Schreibkompetenz. In G. Antos & H. P. Krings (Hrsg.), *Textproduktion. Ein interdisziplinärer Forschungsüberblick* (S. 297–327). Tübingen: Niemeyer.
- Fix, M. (2008). *Texte schreiben. Schreibprozesse im Deutschunterricht* (2. Aufl.). Paderborn: Schöningh.
- Fritzsche, J. (2010). *Schreibwerkstatt. Geschichten und Gedichte: Schreibaufgaben, -übungen, -spiele* (2. Aufl.). Stuttgart: Klett.
- Grzesik, J. & Fischer, M. (1984). *Was leisten Kriterien für die Aufsatzbeurteilung? Theoretische, empirische und praktische Aspekte des Gebrauchs von Kriterien und der Mehrfachbeurteilung nach globalem Ersteindruck*. Opladen: Westdeutscher Verlag.
- Harsch, C., Neumann, A., Lehmann, R. H. & Schröder, K. (2007). Schreibfähigkeit. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie* (S. 42–62). Weinheim: Beltz.
- Heinemann, W. (2000). Textsorte – Textmuster – Texttyp. In K. Brinker, G. Antos, W. Heinemann & S. V. Sager (Hrsg.), *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung* (1. Halbband, S. 507–523). Berlin: de Gruyter.
- Hugener, I. (2006). Überblick über die Beobachtungsinstrumente. In I. Hugener, C. Pauli & K. Reusser (Hrsg.), *Videoanalysen. Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unter-*

- richtsqualität, *Lernverhalten und mathematisches Verständnis*“ (Materialien zur Bildungsforschung, Band 15, S. 45–54). Frankfurt a. M.: Gesellschaft zur Förderung Pädagogischer Forschung.
- KMK – Kultusministerkonferenz. (2014). *Kompetenzstufenmodelle zu den Bildungsstandards im Kompetenzbereich Schreiben, Teilbereich freies Schreiben für den Mittleren Schulabschluss*. Beschluss der Kultusministerkonferenz (KMK) vom 13./14.03.2014. Zugriff am 04.06.2015 unter: https://www.iqb.hu-berlin.de/bista/ksm/KSM_Schreiben_MS_1.pdf
- Knopp, M., Jost, J., Nachtwei, N., Becker-Mrotzek, M., Grabowski, J. (2012). Teilkomponenten von Schreibkompetenz untersuchen: Bericht aus einem interdisziplinären empirischen Projekt. In H. Bayrhuber, U. Harms, B. Muszynski, B. Ralle, M. Rothgangel, L.-H. Schön, H. J. Vollmer & H.-G. Weigand (Hrsg.), *Formate fachdidaktischer Forschung. Empirische Projekte – historische Analysen – theoretische Grundlegungen* (S. 47–65). Münster: Waxmann.
- Kötter, L. & Grau, U. (1965). Zur Bedingtheit der uneinheitlichen Benotung von Schüleraufsätzen (Nacherzählungen). *Zeitschrift für Experimentelle und Angewandte Psychologie*, 12, 278–301.
- Kruse, N. (2006). Rückmeldungen beim Textschreiben. Oder: Wie kann eine gute Rückmeldekultur zu Texten in der Klasse entstehen? *Grundschulunterricht*, 53(10), 14–17.
- Kruse, N., Reichardt, A., Herrmann, M., Heinzl, F. & Lipowsky, F. (2012). Zur Qualität von Kindertexten, Entwicklung eines Bewertungsinstruments in der Grundschule. *Didaktik Deutsch*, 18(32), 87–110.
- Kürzinger, A., Lotz, M., Gleich, A.-K. & Kempfer, I. (2013). Auswertung der Lucybriefe: Perspektivenübernahme und Schreibkompetenz. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien* (S. 255–296). Frankfurt a. M.: Deutsches Institut für Internationale Pädagogische Forschung.
- Kürzinger, A. & Pohlmann-Rother, S. (2014). *Kriterienkatalog Textkorpus 1 – Ein Instrument zur Bestimmung von Textqualität in Klasse 1*. Bamberg: University Press.
- Kürzinger, A. & Pohlmann-Rother, S. (2015). Möglichkeiten einer objektiven und reliablen Bestimmung von Textqualität im Anfangsunterricht – Methodisches Vorgehen und deskriptive Befunde aus dem Projekt NaSch 1. *Didaktik Deutsch*, 20(38), 60–79.
- Lehmann, R. H. (1987). The Scoring of student compositions. In W. Hartmann & R. H. Lehmann (Hrsg.), *The Hamburg Study of Achievement in Written Composition. Part I: Method and findings* (S. 1–89). Hamburg: University.
- Lehmann, R. H. (1990). Aufsatzbeurteilung – Forschungsstand und empirische Daten. In K. Ingenkamp & R. S. Jäger (Hrsg.), *Tests und Trends 8. Jahrbuch der pädagogischen Diagnostik* (S. 64–94). Weinheim: Beltz.
- Lipowsky, F., Faust, G. & Kastens, C. (Hrsg.). (2013). *Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren*. Münster: Waxmann.
- Lotz, M., Lipowsky, F. & Faust, G. (2013). Technischer Bericht zu den PERLE-Videostudien. In F. Lipowsky & G. Faust (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 3* (Materialien zur Bildungsforschung, Band 23/3, S. 15–26). Frankfurt a. M.: Deutsches Institut für Internationale Pädagogische Forschung.
- Muthén, L. & Muthén, B. (1998–2010). *Mplus user's guide* (6. Aufl.). Los Angeles, CA: Muthén & Muthén.
- Neumann, A. (2007). *Briefe schreiben in Klasse 9 und 11. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen*. Münster: Waxmann.
- Nussbaumer, M. (1991). *Was Texte sind und wie sie sein sollen*. Tübingen: Niemeyer.

- Nussbaumer, M. & Sieber, P. (1994). Texte analysieren mit dem Zürcher Textanalyseraster. In P. Sieber (Hrsg.), *Sprachfähigkeiten – besser als ihr Ruf und nötiger denn je!* (Reihe Sprachlandschaft, Bd. 12, S. 141–186). Aarau: Sauerländer.
- Nussbaumer, M. & Sieber, P. (1995). Über Textqualitäten reden lernen – z. B. anhand des „Zürcher Textanalyserasters“. *Diskussion Deutsch*, 26(141), 36–52.
- Persky, H. R., Daane, M. C. & Jin, Y. (2003). *The nation's report card: Writing 2002* (NCES 2003–529). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Zugriff am 04.04.2014 unter <http://nces.ed.gov/nationsreportcard/pdf/main2002/2003529.pdf>
- Pohl, T. (2013). Texte schreiben in der Grundschule. In S. Gailberger & F. Wietzke (Hrsg.), *Handbuch Kompetenzorientierter Deutschunterricht* (S. 212–231). Weinheim: Beltz.
- Pohl, T. & Steinhoff, T. (2010). Textformen als Lernformen. In T. Pohl & T. Steinhoff (Hrsg.), *Textformen als Lernformen* (Kölner Beiträge zur Sprachdidaktik, Reihe A, Bd. 7, S. 5–26). Duisburg: Gilles & Francke.
- Portmann, P. R. (1996). Arbeit am Text. In H. Feilke & P. R. Portmann (Hrsg.), *Schreiben im Umbruch. Schreibforschung und schulisches Schreiben* (S. 158–171). Stuttgart: Klett.
- Quellmalz, E. S., Capell, F. J. & Chou, C.-P. (1982). Effects of discourse and response mode on the measurement of writing competence. *Journal of Educational Measurement*, 19(4), 241–258.
- Sandig, B. (2006). *Textstilistik des Deutschen* (2. vollst. neu bearb. u. erw. Aufl.). Berlin: de Gruyter.
- Schründer-Lenzen, A. (2013). *Schriftspracherwerb und Unterricht. Bausteine professionellen Handlungswissens* (4., völlig überarb. Aufl.). Wiesbaden: VS.
- Seidel, T., Prenzel, M., Duit, R. & Lehrke, M. (Hrsg.). (2003). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN).
- Steinig, W., Betzel, D., Geider, F. J. & Herbold, A. (2009). *Schreiben von Kindern im diachronen Vergleich. Texte von Viertklässlern aus den Jahren 1972 und 2002*. Münster: Waxmann.
- Thonke, F., Groß Ophoff, J., Hosenfeld, I. & Isaac, K. (2008). Kriteriengestützte Erfassung von Schreibleistungen im Projekt VERA. In B. Hofmann & R. Valtin (Hrsg.), *Checkpoint Literacy. Tagungsband 2 zum 15. Europäischen Lesekongress in Berlin* (S. 28–35). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- Weinhold, S. (2000). *Text als Herausforderung. Zur Textkompetenz am Schulanfang. Mit 296 Schülertexten aus Klasse 1*. Freiburg i. B.: Fillibach.
- Weinhold, S. (2002) Textkompetenz am Schulanfang. Ergebnisse einer Untersuchung von Texten aus dem ersten Schuljahr. In P. R. Portmann-Tselikas (Hrsg.), *Textkompetenz. Neue Perspektiven für das Lernen und Lehren* (S. 147–162). Innsbruck: Studienverlag.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest Version 2.0. Generalised item response modelling software*. Melbourne: ACER.