

**Unterrichtliche Determinanten der Schreibkompetenz
in der Primarstufe**

—

**Wie wirken sich Aufgabenstellung und individuelle
Lernunterstützung auf die Textqualität im Anfangsunterricht aus?**



Inaugural-Dissertation

in der Fakultät Humanwissenschaft
der Otto-Friedrich-Universität Bamberg

**vorgelegt von
Anja Kürzinger
aus Kösching**

Bamberg, den 10.10.2016

Tag der mündlichen Prüfung: 20.01.2017

Dekan: Universitätsprofessor Dr. Stefan Hörmann

Erstgutachter: Universitätsprofessor Dr. Claus H. Carstensen

Zweitgutachterin: Universitätsprofessorin Dr. Sanna Pohlmann-Rother

URN: urn:nbn:de:bvb:473-opus4-488221

DOI: <https://doi.org/10.20378/irbo-48822>

Danksagung

Besonders möchte ich Prof. Dr. Gabriele Faust danken, ohne die diese Dissertation vermutlich nicht zustande gekommen wäre: Danke für dein Vertrauen in mich, deine Ermutigung, „fachfremd“ zu promovieren und deine Unterstützung in vielerlei Hinsicht. Dein Engagement und deine Willenskraft haben mich zutiefst beeindruckt.

Ich möchte mich vor allem bei Prof. Dr. Sanna Pohlmann-Rother als meine Zweitgutachterin und wissenschaftliche Mentorin für die intensive und inhaltlich wie persönlich wertvolle Betreuung meiner Arbeit sowie für die konstruktive, anregende Zusammenarbeit im Projekt NaSch1 herzlich bedanken.

Danke, dass du nicht nur meine Dissertation so professionell und engagiert in allen Phasen begleitet hast, sondern mich auch auf meinem wissenschaftlichen Werdegang förderst und unterstützt. Ich freue mich auf unsere weitere Zusammenarbeit!

Prof. Dr. Claus Carstensen danke ich sehr für seine Bereitschaft, das Erstgutachten nach dem Tod von Gabi Faust so unkompliziert zu übernehmen und jederzeit ein Ansprechpartner für mich zu sein. Für die Möglichkeit, im Rahmen der Bamberger BAGSS-Graduiertenschule als assoziatives Mitglied promovieren zu können, danke ich dir ebenfalls.

Ich danke auch dem PERLE-Team für die Unterstützung und für die Bereitschaft, mir die Daten aus dem PERLE-Projekt zur Verfügung zu stellen. Ein herzlicher Dank geht dabei an die Co-Autoren meiner Beiträge: Dr. Miriam Hess, Dr. Edgar Schoreit und Prof. Dr. Lipowsky. Miriam Hess hat mit ihrer hilfreichen Unterstützung entschieden zum Gelingen dieser Arbeit beigetragen. Danke für deine vielen kompetenten Ratschläge und dafür, dass du für meine Fragen immer ein offenes Ohr hattest!

Schließlich geht mein Dank auch an meine Familie und meine Freunde. Adrian Krenn hat mich vor allem in der schwierigen Anlaufphase unterstützt und mich darin bestärkt, eine wissenschaftliche Laufbahn einzuschlagen.

Die legendäre Lui-Wg in Bamberg hat sichergestellt, dass auch ein Leben neben der Dissertation existiert. Danke für die vielen schönen Ablenkungen in der WG-Küche.

Rainer Poppert danke ich aus ganzem Herzen dafür, dass er an mich glaubt, mich immer wieder motiviert und in allem unterstützt. Danke für die tiefgründigen Diskussionen, die Abende am Palast und deine mitreißende Abenteuerlust. Mit dir ist das Leben wundervoll.

Zusammenfassung

Den Ausgangspunkt der vorliegenden Dissertation bildet die Frage, welche unterrichtlichen Bedingungsfaktoren als Erklärung für die individuelle Schreibperformanz in der Primarstufe herangezogen werden können. Vor dem Hintergrund eines elementaren Forschungsdesiderats werden im Rahmen von fünf Einzelbeiträgen Schreibfähigkeiten in Klasse 1 auf breiter Basis empirisch-quantitativ analysiert und mit der Prozessqualität von Schreibunterricht in Verbindung gebracht.

Durch die Integration theoriebasierter, schreibdidaktischer Überlegungen mit der empirischen, videogestützten Unterrichtsforschung lässt sich das Forschungsvorhaben interdisziplinär verorten. So erfolgt die Modellierung der Konstrukte auf Basis fachdidaktischer Grundlagen, die mit psychometrischen Anforderungen an die Operationalisierung verbunden werden.

Dabei greift die Arbeit, die in das DFG-Projekt NaSch1 (Narrative Schreibkompetenz in Klasse 1) eingebettet ist, auf 49 Unterrichtsvideos und 540 Texte aus der ersten Jahrgangsstufe der PERLE-Studie (Persönlichkeits- und Lernentwicklung von Grundschulkindern) zurück.

Die Befunde zeigen, dass sich bereits im Anfangsunterricht komplexere, funktional-pragmatische Schreibfähigkeiten entwickeln, die sich in zwei Subdimensionen (semantisch-pragmatischer und sprachsystematischer Faktor) von Schreibkompetenz differenzieren lassen. Unter Kontrolle des Geschlechts der Lernenden sowie ihrer schriftsprachbezogenen Vorläuferfähigkeit (Buchstabenkenntnis) wirken sich vor allem präzise, verständliche Aufgabenstellungen und evaluative sowie lösungsorientierte Lernunterstützungen auf die Qualität der individuellen Schreibleistungen aus. Neben diesen Wirkungsanalysen liegen auch detaillierte deskriptive Ergebnisse zur unterrichtlichen Sicht- und Tiefenstruktur vor, die Einblicke in das Lehrerhandeln und die Praxis des Schreibunterrichts in Klasse 1 ermöglichen.

Diese Befunde werden in dem vorliegenden Rahmentext unter Berücksichtigung der methodischen Vorgehensweise kritisch diskutiert und um Implikationen für die Schreibdidaktik und Schreib- bzw. Unterrichtsforschung erweitert.

Inhaltsverzeichnis

1. Einleitung	4
1.1. Status Quo der Schreibforschung und -didaktik	4
1.2. Zielsetzung und Verortung der Dissertation im Projekt NaSch1	6
1.3. Forschungsfragen der Dissertation	11
2. Synopsis	12
2.1. Skizzierung der Erhebungssituation.....	12
2.2. Überblick über die Publikationen.....	14
2.3. Zusammenfassende Darstellung der Publikationen innerhalb des Forschungsvorhabens	16
3. Gesamtdiskussion	26
3.1. Diskussion der zentralen Befunde	26
3.2. Diskussion der methodischen Vorgehensweise	34
3.3. Implikationen für die Schreibdidaktik und Schreib- bzw. Unterrichtsforschung	43
3.4. Ausblick auf zukünftige Forschung zu individuellen Schreibfähigkeiten im Grundschulalter	45
4. Literaturverzeichnis	47
5. Tabellenverzeichnis.....	60
6. Abbildungsverzeichnis.....	61
7. Anlagen.....	62

1. Einleitung

1.1. Status Quo der Schreibforschung und -didaktik

Lesen und Schreiben bilden nicht nur basale Schlüsselqualifikationen für schulischen und beruflichen Erfolg, sondern determinieren im Allgemeinen auch die gesellschaftliche und kulturelle Teilhabe. Doch trotz der Bedeutung schriftsprachlicher Kompetenzen für moderne Wissensgesellschaften wird hierzulande vor allem die individuelle Leseleistung von Schülerinnen und Schülern im forschungsbezogenen Diskurs priorisiert: „Was hingegen in der gegenwärtigen Diskussion unfreiwillig ins Hintertreffen zu geraten scheint, ist das Schreiben bzw. die Schreibkompetenz“ (Philipp, 2012, S. 59). So besteht nach wie vor ein maßgeblicher Forschungsbedarf hinsichtlich der Qualitätsbeurteilung von Texten¹ im deutschsprachigen Raum. Die Schwierigkeit manifestiert sich insbesondere darin, Beurteilungskriterien zu entwickeln, die zum einen Textqualität erfassen und zum anderen die empirischen Gütekriterien erfüllen (Blatt, Ramm & Voss, 2009, S. 54). Aufgrund der Vielfältigkeit und Komplexität von Schreibprodukten einerseits und der vergleichsweise geringen Routine der Fachdidaktik in der empirischen Bestimmung von Textqualität (Becker-Mrotzek & Schindler, 2008) andererseits ist die Konzeption derartiger Kriterien bislang nur zum Teil erfolgreich umgesetzt worden (Blatt et al., 2009). Dieser Umstand vermag erklären, weshalb sich die Schreibforschung allgemein auf vergleichsweise wenige Studien beschränkt, Untersuchungen zur Textqualität oftmals an methodischen Problemen fehlschlagen (Neumann, 2014) und vor allem auch umfangreichere, empirisch ausgerichtete Analysen rar sind. Bis auf wenige Untersuchungen wie z. B. die IEA-Aufsatzstudie (Lehmann, 1987; Lehmann, 1990), IGLU-E 2006 (Blatt, Voss & Matthießen, 2005), DESI (Neumann, 2007; Neumann & Lehmann, 2008), „Lernstand in Klasse 6“ (Blatt et al., 2009) oder die Studie zu Kodierstrategien von Schreibaufgaben (Böhme, Bremerich-Voss & Robitzsch, 2009) wurden Schreibleistungen von Schülerinnen und Schülern kaum in höheren Fallzahlen erfasst. Insbesondere für den Anfangsunterricht wird ein Forschungsdesiderat in Bezug auf die Beschreibung und empirisch-quantitative Analyse von Schreibfähigkeiten konstatiert. Dabei werden in der Primarstufe die Weichen für das Schreibenlernen und die Entwicklung von Schreibkompetenz gestellt. Auch aus der Perspektive der Didaktik sind der Prozess der schriftlichen Ontogenese und das „Erstschreiben“ von hoher Relevanz, da sie das Denken, Sprachbewusstsein und Handeln

¹ Das Verhältnis zwischen Schreibkompetenz und Textqualität wird anhand des Kompetenz-Performanz-Modells expliziert. Dieses Konzept beruht auf der Annahme, dass eine gezeigte Leistung als Indikator für eine entsprechende dahinterliegende Kompetenz betrachtet werden kann (Harsch, Lehmann, Neumann & Schröder, 2007; Knopp, Jost, Nachtwei, Becker-Mrotzek & Grabowski, 2012; Neumann, 2007). Eine ausführlichere Darstellung erfolgt im Rahmen der Zusammenfassung des ersten Beitrags auf S. 16 und der Diskussion auf S. 26-27 sowie S. 32.

der Kinder maßgeblich beeinflussen (Weinhold, 2014). Pohl (2014) nimmt an, „(...) dass ganz entscheidende Erwerbsphasen auf dem Weg zu einer entwickelten Textproduktionskompetenz bereits in den Grundschuljahren erfolgen, die als zentrales Fundament für die sich dann weiter ausdifferenzierenden Schreibkompetenzen der Mittel- und Oberstufe gelten müssen“ (S. 212). Entsprechend wird auch die Notwendigkeit einer frühen Förderung der Schreibfähigkeiten als unstrittig postuliert (Bachmann, Ospelt-Geiger, Ospelt & Vital, 2007). „Schulanfänger sollen das Schreiben von Anfang an als sprachlich-konzeptionellen, kommunikativen und kooperativen Prozess erlernen und als Ausdrucksmittel ihrer Gedanken in verschiedenen Funktionen nutzen“ (Weinhold, 2014, S. 144). So ist auch im Kompetenzbereich „Schreiben – Texte verfassen“ der Bildungsstandards für die Grundschule im Fach Deutsch vorgesehen, dass Schülerinnen und Schüler am Ende der Jahrgangsstufe 4 „verständlich, strukturiert, adressaten- und funktionsgerecht“ schreiben können (Bildungsstandards für die Grundschule: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005, S. 11)². Mit diesen normativen Standards werden erwartete Kompetenzen der Lernenden benannt und überprüfbar gemacht, die Gestaltung der Unterrichtspraxis bleibt zunächst allerdings offen (Klieme, 2005, S. 7).

Befasst man sich mit der Frage nach der lernförderlichen Gestaltung ebendieses Schreibunterrichts, wird jedoch eine weitere Problematik offenkundig: Denn nicht nur die Schreibfähigkeiten, auch die Praxis des Schreibunterrichts – generell und verstärkt im Anfangsunterricht – ist bislang weitgehend unerforscht. Die schreibdidaktische Lehr-Lern-Forschung konzentriert sich hauptsächlich auf konzeptionelle Empfehlungen zur Umsetzung des Unterrichts, die jedoch nicht anhand eines quantitativen Zugangs wissenschaftlich überprüft wurden (vgl. auch Klieme et al., 2008). „Wenn man genau hinsieht, gibt es fast keine gängige Praxis im Deutschunterricht, die durch empirische Forschung wirklich abgesichert wäre“ (Kammler & Knapp, 2002, S. 6).

Aber auch innerhalb der Unterrichtsforschung wurde im deutschsprachigen Raum bisher weitgehend versäumt, domänenspezifische Qualitätsaspekte von Schreibunterricht zu erfassen. Während allgemeine, fachübergreifende Qualitätskriterien von Unterricht national wie international empirisch gut dokumentiert sind, fokussieren Studien zur fachdidaktischen Unterrichtsqualität überwiegend mathematische und naturwissenschaftliche Fächer (Gräsel & Göbel, 2015). „Vor der DESI-Studie wurden so gut wie keine breiten, repräsentativen Erhebungen zur Praxis des Deutschunterrichts durchgeführt, ganz zu schweigen von systematischen, experimentellen oder quasi-experimentellen Wirkungsanalysen, mit denen

² Die Erwartungen an die Kompetenzen der Lernenden im Rahmen der Bildungsstandards in der Primarstufe werden beispielsweise von Becker-Mrotzek (2014) als „deutlich zu hoch“ kritisiert. In diesem Kontext könnten quantitative Analysen zur Entwicklung der Schreibfähigkeiten von Grundschulkindern mit höheren Stichprobenumfängen auch zur Klärung der Frage beitragen, welche Kompetenzen sich während der Grundschulzeit erwerben lassen.

Effekte bestimmter Unterrichtsmerkmale hätten nachgewiesen werden können“ (Klieme et al., 2008). Dieses Defizit erscheint nicht nur im Hinblick auf die Auseinandersetzung mit der domänenspezifischen Unterrichtsqualität bedenklich. Vielmehr ist zu kritisieren, dass sich mit dem Mangel an erforschten Prozessvariablen des Schreibunterrichts auch keine Rückschlüsse auf unterrichtliche Einflussfaktoren der Schreibleistung treffen lassen. Obgleich individuelle Merkmale von Schülerinnen und Schülern (etwa Intelligenz, Geschlecht oder der sozioökonomische Hintergrund) allgemein für etwa die Hälfte der Leistungsunterschiede prädiktiv sind und damit den höchsten Anteil an der Aufklärung der Leistungsvarianz einnehmen (z. B. Campbell, Kyriakides, Muijs & Robinson, 2004; Hattie, 2009), sind auch Aspekte auf der Unterrichts- und Klassenebene nachweislich für erfolgreiche schulische Lernprozesse von Bedeutung (Sammons, Hillman & Mortimore, 1995; Scheerens & Bosker, 1997). So beträgt der durch die Unterrichtsbedingungen erklärte Anteil der Leistungsvarianz im Mittel zwischen 10 % und 30 % (Campbell et al., 2004; Hattie, 2009). Neuere Befunde sprechen darüber hinaus für stärkere Effekte von Schule und Unterricht auf die individuelle Entwicklung im Primarbereich (Zfs. Lipowsky, 2007; Seidel & Shavelson, 2007). Ebenso wird angenommen, dass von domänenspezifischen und curricularen Unterrichtsmerkmalen noch entscheidendere Einflüsse auf Leistungsvariablen ausgehen (Hattie, 2009; Seidel & Shavelson, 2007; vgl. auch Lipowsky, 2015).

Demzufolge sollte bei einer Analyse der Schreibfähigkeiten von Grundschülerinnen und Grundschüler der domänenspezifische Schreibunterricht als Determinante der Leistungsperformanz berücksichtigt und in seiner Gestaltung und Prozessqualität systematisch untersucht werden:

Die konzeptuelle Entwicklung und empirische Überprüfung von Qualität im Sprachlichen Anfangsunterricht ist dabei vordringlichste Aufgabe der Fachdidaktik in enger Kooperation mit den Bezugsdisziplinen (wie z. B. Psychologie, Soziologie, Pädagogik). Aus einer fachdidaktischen Empirie müssen Referenzmodelle etabliert werden, auf deren Basis schließlich in einem zweiten Schritt die Wirksamkeit kompetenzorientierten Unterrichts überprüft wird. (Wildemann, 2011, S. 95)

1.2. Zielsetzung und Verortung der Dissertation im Projekt NaSch1

Vor dem Hintergrund des skizzierten Forschungsdesiderats besteht das übergeordnete Ziel der Dissertation darin, Schreibfähigkeiten in der Schuleingangsstufe empirisch-quantitativ zu analysieren und dabei Prozessmerkmale der schriftsprachlichen Unterrichtsgestaltung und -qualität zur ihrer Erklärung heranzuziehen. Diese empirischen Wirkungsanalysen setzen an ausgewählten Gelenkstellen des Schreibunterrichts an, deren Sicht- und Tiefenstrukturen umfangreich dokumentiert werden.

Die Modellierung und Operationalisierung des latenten Konstrukts der Schreibkompetenz sowie der Prozessvariablen von Unterricht erfolgen auf Basis fachdidaktischer Grundlagen,

die mit psychometrischen Anforderungen an die Messung verbunden werden. Im Hinblick auf die Integration theoriebasierter, schreibdidaktischer Überlegungen mit der empirischen Unterrichtsforschung lässt sich das Forschungsvorhaben interdisziplinär verorten. Dabei liegt mit der Verwendung von Videoanalysen zur Beschreibung und Bewertung unterrichtlicher Prozesse die videobasierte Unterrichtsqualitätsforschung als spezieller Ansatz zugrunde. Den übergeordneten Kontext für das Dissertationsvorhaben bildet das DFG-Projekt NaSch1³ (Narrative Schreibkompetenz in Klasse 1), das die Schreibfähigkeiten von Erstklässlerinnen und Erstklässlern in Abhängigkeit von individuellen, familiären sowie unterrichtlichen Einflussfaktoren untersucht.

Im Einzelnen wird in NaSch1 folgenden übergreifenden Fragestellungen nachgegangen:

- 1) Welche Qualität weisen die Texte der Lernenden in Klasse 1 auf?
- 2) Welche individuellen Merkmale zu Schulbeginn beeinflussen die Textqualität im Anfangsunterricht?
- 3) Auf welche Weise beeinflusst die Unterrichtsqualität die Textqualität?
- 4) Wie lassen sich Best-Practice-Klassen im Hinblick auf Text- und Unterrichtsqualität charakterisieren?

In diesen Fragestellungen werden linguistische, deutschdidaktische und grundschulpädagogische Perspektiven im Rahmen der empirischen Bildungsforschung verknüpft, wobei ein vereinfachtes theoretisches Modell zu Grunde gelegt wird, das sich an konstruktivistischen Lerntheorien⁴ (z. B. Reusser, 2001, 2006) sowie am systemischen Angebots-Nutzungsmodell von Unterricht (Fend, 1981; Helmke, 2014) orientiert. Im Zuge der Untersuchung von Zusammenhangsmustern zwischen Unterrichtswirksamkeit und Textqualität berücksichtigt das Rahmenmodell individuelle Charakteristika der Schülerinnen und Schüler (z. B. schriftsprachbezogene Vorläuferfähigkeiten, Geschlecht), Kontextfaktoren wie die Klassenzusammensetzung, Merkmale der Lehrpersönlichkeit bzw. Einstellungen von Lehrpersonen sowie unterrichtliche Rahmenbedingungen und Voraussetzungen. Unterricht wird in diesem Zusammenhang als sozialer Prozess verstanden, in dem das Handeln der Lehrpersonen eine Lernumgebung kreiert, die im Sinne eines Angebots individuell von den Lernenden wahrgenommen werden kann (Klieme, 2006, S. 765). Mit dem Fokus auf Unterrichtsprozesse bildet das Angebots-Nutzungsmodell auch die theoretische Grundlage für die videogestützte Unterrichtsforschung (Pauli & Reusser, 2006), deren Ansatz für die

³ Leitung: Sanna Pohlmann-Rother & Gabriele Faust (†), Förderkennzeichen PO 1739/1-1.

⁴ Das sozial-konstruktivistische Lehr-Lernverständnis basiert auf Arbeiten von Piaget (z. B. 1935/1969, 1936/1969) und Wygotsky (1978, 1986).

Erfassung der unterrichtlichen Prozessvariablen im Rahmen des Dissertationsvorhabens konstitutiv ist.

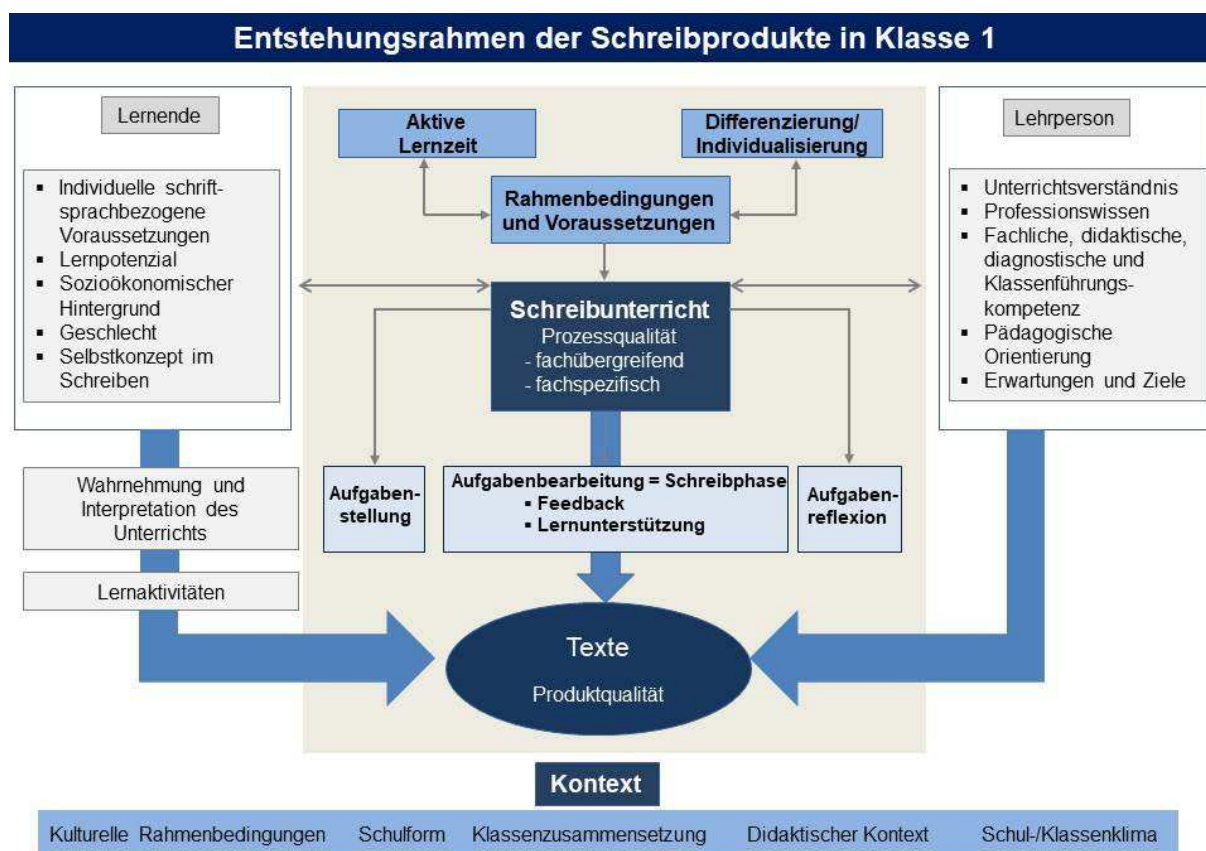


Abb. 1: Rahmenbedingungen und Einflussfaktoren auf die Textqualität (in Anlehnung an Lotz, 2016; S. 76 und Helmke, 2014, S. 71)⁵

Wie Abbildung 1 verdeutlicht, wird der Prozess des Schreibunterrichts im Projekt NaSch1 entlang der Gelenkstellen Aufgabenstellung, Aufgabenbearbeitung sowie Reflexion differenziert aufgeschlüsselt⁶. Es wird angenommen, dass sowohl situative fachdidaktische (z. B. Vermittlung von Schreibstrategien) als auch fachübergreifende (z. B. Klarheit) Wirkprinzipien von Unterricht im Kontext dieser drei Gelenkstellen auf die Schreibperformanz

⁵ Dieses Modell zeigt lediglich eine verkürzte Sicht auf die komplexen Einflussfaktoren auf Textqualität und die vielschichtigen Wirkungsmechanismen von Unterricht. Damit erhebt es keinen Anspruch auf Vollständigkeit, sondern dient ausschließlich der Veranschaulichung ausgewählter Prädiktoren von Schreibkompetenz. In Anlehnung an das Angebots-Nutzungsmodell wird der Unterricht als Angebot verstanden, wobei Mediationsprozesse (Wahrnehmung und Interpretation der Schülerinnen und Schüler) die Lernaktivitäten und die Schreibkompetenz beeinflussen (Helmke, 2014).

⁶ Stein, Grover und Henningsen (1996, S. 459) grenzen Aufgabenstellung und Aufgabenbearbeitung im Unterrichtsprozess voneinander ab. Blömeke, Risse, Müller, Eichler & Schulz (2006) erweitern diese Einteilung um die Auswertungs- und Reflexionsphase.

Einfluss nehmen⁷. Eine Beurteilung der Textqualität ohne Berücksichtigung des Schreibunterrichts greift auch aus didaktischer Perspektive zu kurz, da in deren Verständnis Texte der Lernenden als „Reaktion auf eine konkrete schreibdidaktische Situierung“ entstehen und nicht losgelöst vom Unterrichtsgeschehen bewertet werden können (Pohl & Steinhoff, 2010).

Basierend auf diesen Prämissen werden Interaktionsprozesse zwischen den Lehrpersonen und Lernenden innerhalb der Gelenkstellen systematisch erfasst sowie ihr Zusammenhang mit den Schreibfähigkeiten der Kinder analysiert.

Den Ausgangspunkt der unterrichtlichen Gelenkstellen bildet die Aufgabenstellung, in der die Schreibaufgabe zunächst von der Lehrkraft eingeführt und erläutert wird. Mit Aufgaben lässt sich das Unterrichtsgeschehen gestalten, planen und reflektieren (Bromme, 1981; Jordan et al., 2006), weswegen das Formulieren von Schreibaufgaben allgemein als „...das genuine Geschäft der Didaktik und das der alltäglichen Konzeption von Unterricht“ (Dehn, 2009, S. 154) begriffen wird. In der Phase der Aufgabenstellung wird in der vorliegenden Untersuchung jedoch nicht die Qualität der Schreibaufgabe⁸, sondern ihre konkrete Situierung durch die Lehrperson im Unterricht untersucht. Diese Unterscheidung ist essenziell, zumal die Auswahl einer kompetenzförderlichen Schreibaufgabe nicht gleichbedeutend mit einer effektiven Umsetzung der Aufgabenstellung ist. Vielmehr kann sich dieses Verhältnis als diametral erweisen, etwa, wenn Lehrkräfte durch ihr unterrichtliches Handeln Aufgaben modifizieren und beispielsweise ihren ursprünglichen Anspruch reduzieren (siehe auch S. 19). In der Debatte um lernförderliche Schreibaufgaben, die in den letzten Jahren zunehmend in den didaktischen Fokus gerückt sind (z. B. Bachmann & Becker-Mrotzek, 2010; Hoppe & Metz, 2013; Metz, Maier, Kleinknecht, Bohl & Hoppe, 2012), wird die Implementierung von Schreibaufgaben im Unterricht kaum explizit analysiert oder von der Aufgabenqualität separiert. Angesichts dieses Forschungsdesiderats erscheint es umso bedeutender, die Umsetzung von Aufgabenstellungen im Unterricht zu beobachten und ihren Zusammenhang mit der Textqualität zu untersuchen.

Im Anschluss an den Prozess der Aufgabenstellung folgt im Unterricht die Schreibphase, in der die Schülerinnen und Schüler ihren Text entsprechend des Schreibarrangements verfassen. Trotz der selbstständigen Schüleraktivitäten lassen sich in dieser Arbeitsphase verschiedene Möglichkeiten des kompetenzförderlichen, aktiven Einwirkens der Lehrkraft realisieren, etwa in Form von individueller Lernunterstützung und Rückmeldungen für die

⁷ Die Aufgabenstellung und Aufgabenbearbeitung können sich direkt auf einen entstehenden Text auswirken, während die Aufgabenreflexion nach dem Verfassen des Textes einsetzt und daher indirekte Effekte auf die Entwicklung der Schreibkompetenz zu besitzen vermag.

⁸ Durch die Vorgabe eines standardisierten Schreibansatzes für alle teilnehmenden Klassen im Projekt PERLE (siehe Kapitel 2.1), kann die Auswahl einer geeigneten Schreibaufgabe als Kernkompetenz von Lehrkräften nicht erfasst werden.

Lernenden. Dies gilt vor allem in Anbetracht der Tatsache, dass die Texte nicht im Rahmen einer klassischen Testsituation entstanden sind, sondern direkt in den Schreibunterricht eingebettet wurden. So stellen die Schülerarbeitsphasen Zeitfenster (Kobarg, 2004) dar, in denen Lehrpersonen den Lernstand und die Verstehenstiefe der Schreibanfängerinnen und Schreibanfänger diagnostizieren und individuell unterstützen können (Schrader, 1997; Reusser & Pauli, 2010). Aus dieser Perspektive trägt das Unterrichtsmodell auch dem konstruktivistischen Lernverständnis Rechnung: „Dieses Konzept geht verkürzt gesagt davon aus, dass Lernen ein zwar eigenständiger, konstruktiver Prozess ist, der aber unabdingbar auf die Anregung und Unterstützung von außen angewiesen ist“ (Böttcher & Becker-Mrotzek, 2009, S. 49).

Nach der Schreibphase können Reflexionsphasen, von denen beispielsweise Blömeke und Kolleginnen und Kollegen (2006) eine hohe Lernwirkung erwarten, den Abschluss der Lerneinheit bilden. Ihr potentieller Beitrag zur Kompetenzentwicklung der Kinder kann sich darauf gründen, dass sie dazu anregen, über eigene Lernwege vertiefend nachzudenken (Wyss, 2013; Lotz, 2016).

Innerhalb der Einbettung in das Projekt NaSch1 und unter Rückgriff auf das Wirkungsmodell von Schreibunterricht fokussiert die vorliegende Dissertation auf die Aufgabenstellung sowie die Lernunterstützung⁹ in der Schülerarbeitsphase als zentrale unterrichtliche Gelenkstellen, die als Prädiktoren für die Ausgestaltung individueller Schreibfertigkeiten herangezogen werden. Für beide Phasen werden Messinstrumente entwickelt, anhand derer die Lehr-Lernprozesse erfasst und mit der Qualität von Kindertexten in Verbindung gebracht werden.

⁹ Das Feedback-Verhalten von Lehrkräften wird im Projekt NaSch1 als ein Aspekt von Lernunterstützungen separat analysiert und ist kein Bestandteil der vorliegenden Arbeit.

1.3. Forschungsfragen der Dissertation

Ausgehend von dem unter 1.2 vorgestellten Rahmenmodell zu den unterrichtlichen Einflussfaktoren auf Textqualität und der beschriebenen Zielsetzung nimmt die Dissertation auf folgende Forschungsfragen Bezug, die in zwei umfassende Themenkomplexe unterteilt sind¹⁰:

Fragestellung 1): Textqualität

- a) Wie lässt sich das Konstrukt Schreibkompetenz im Anfangsunterricht operationalisieren und nach empirischen Gütekriterien erfassen?
- b) Welche Qualität weisen die untersuchten Texte der Lernenden im Anfangsunterricht auf?
- c) Lässt sich die Schreibkompetenz im ersten Schuljahr in Teilkompetenzen aufschlüsseln oder stellt sie sich als eine globale Dimension dar?
- d) Welche Zusammenhänge zeigen sich zwischen den beiden Textbewertungsinstrumenten (holistisches und kriteriales Messinstrument)?

Fragestellung 2): Schreibunterricht und seine Effekte

- a) Wie lassen sich Aufgabenstellung und individuelle Lernunterstützung als Gelenkstellen des Schreibunterrichts in Klasse 1 operationalisieren und nach empirischen Gütekriterien videogestützt erfassen?
- b) Welche Qualität weisen die Aufgabenstellung und individuelle Lernunterstützung im Schreibunterricht der ersten Klasse auf?
- c) Wie wirken sich Aufgabenstellung und individuelle Lernunterstützung unter Kontrolle individueller Prädiktoren der Textqualität (Geschlecht, schriftsprachbezogene Vorläuferfähigkeit) auf die Schreibleistungen der Schülerinnen und Schüler aus?

Der erste Themenkomplex wird innerhalb der Beiträge 1 und 2 behandelt, während die Fragestellungen zum Schreibunterricht und seinen Zusammenhängen mit der Textqualität in Beitrag 3 bis 5 aufgegriffen werden.

¹⁰ Detaillierte Forschungsfragen und entsprechende Hypothesen können den einzelnen Beiträgen entnommen werden.

2. Synopsis

2.1. Skizzierung der Erhebungssituation

Die Daten, die dem Forschungsvorhaben und dem Projekt NaSch1 zugrunde liegen, stammen aus der Längsschnittstudie PERLE¹¹ (Persönlichkeits- und Lernentwicklung von Grundschulkindern, vgl. Lipowsky, Faust, Kastens & Post, 2013) und wurden innerhalb eines Videomoduls im März der Klassenstufe 1 erhoben. Im BMBF-Projekt PERLE wurde jeweils eine ca. 90-minütige Unterrichtseinheit im Fach Deutsch auf Grundlage eines Kameraskripts mit zwei digitalen Kameras (Lehrerkamera, Klassenkamera) in 37 Klassen aufgezeichnet (Lotz & Corvacho del Toro, 2013). Derartige Videoaufnahmen stellen grundsätzlich ein sinnvolles Instrument für die Untersuchung von Mikroprozessen im Unterricht dar. Ihr Mehrwert gegenüber der Live-Beobachtung gründet sich auf der Authentizität und Ganzheitlichkeit des Datenmaterials als auch auf der Wiederholbarkeit und Re-Analysierbarkeit (Pauli & Reusser, 2006, S. 787).

In den untersuchten Videomodulen unterrichteten die Lehrkräfte auf Basis von Vorgaben, da diese inhaltliche Standardisierung einen Vergleich der Unterrichtsbeobachtungen zwischen den Klassen ermöglicht (z. B. Dalehefte, 2006; Krammer, 2009) und vor allem auch eine Voraussetzung für fachdidaktische Unterrichtsanalysen darstellt (Pauli, 2008).

Zwei Vorgaben, mit denen die Lehrkräfte in den 37 teilnehmenden Klassen beauftragt wurden, sind für die Analyse der Text- und Unterrichtsqualität von Bedeutung: So sollte zunächst der Anfangsteil des fantastischen Bilderbuchs „Lucy rettet Mama Krok“ (Doucet & Wilsdorf, 2005) vorgestellt werden. Im Anschluss an die rezipierte Geschichte sollte von den Lernenden ein Brief aus der Perspektive der literarischen Hauptfigur Lucy an eine andere literarische Figur (Krokodilmutter „Mama Krok“) verfasst werden. Die Schreibaufgabe knüpft unmittelbar an das Geschehen im Bilderbuch an, indem die Schülerinnen und Schüler angehalten werden, einen Abschiedsbrief zu formulieren, als das Mädchen Lucy seine Krokodilfamilie verlässt. Mit diesen Vorgaben wurde die Textproduktion in ein handlungs- und produktionsorientiertes Verfahren eingebettet.

Eine Besonderheit der Stichprobe liegt in ihrem Einbezug staatlicher und privater Grundschulen, sogenannter BIP-Kreativitätsgrundschulen (Begabung – Intelligenz – Persönlichkeit, Melhorn & Melhorn, 2003)¹². Aufgrund des in fast allen Klassen der BIP-Schulen praktizierten Team-Teachings, bei dem die Lehrkräfte die Klasse teilweise in zwei Teilgruppen in unterschiedlichen Räumen aufteilen, sind aus insgesamt 37 Klassen 49

¹¹ Leitung: Frank Lipowsky und Gabriele Faust (†).

¹² Über alle fünf Beiträge konnten für die erste Klasse keine signifikanten Unterschiede zwischen den staatlichen und den privaten Schulen festgestellt werden, sodass eine Gegenüberstellung der Schularten im Folgenden nicht weiter thematisiert wird.

Unterrichtsvideos (staatliche Klassen $N = 20$; BIP-Klassen $N = 29$) entstanden. Davon werden für die schreibdidaktische Unterrichtsbeschreibung während der Aufgabenstellung 43 Videos und für die Textbewertung 540 Schreibprodukte berücksichtigt¹³. Im Rahmen der Wirkungsanalyse des Unterrichts verringert sich diese Stichprobe auf 41 Videos und 524 Schülerinnen und Schüler¹⁴. Die deskriptiven Analysen der Lernunterstützung beziehen sich auf alle 49 Videos, während sich die Mehrebenenanalysen in dieser Phase auf ebenfalls 41 Videos stützen.

Zum Zeitpunkt der Einschulung beträgt das Alter der Kinder ($N = 524$ ¹⁵) im Mittel 6 Jahre und 9 Monate ($SD = 4.7$ Monate). 44.1 % der Kinder sind männlich, 55.9 % weiblich, wobei 58 % staatliche Schulen und 42 % eine private BIP-Kreativitätsgrundschule besuchen.

Zusätzlich zu dem videografierten Unterricht kann auf umfangreiche Individualdaten der Schülerinnen und Schüler zurückgegriffen werden (u.a. Vorläuferfähigkeiten Schriftsprache) sowie auf Daten aus einer in Klasse 1 durchgeführten Elternbefragung (u.a. sozialer Hintergrund) und einer Befragung der Lehrkräfte (u.a. „Constructivist“ vs. „Transmission View“).

¹³ Sechs Aufnahmen kommen nicht in Betracht, da die Aufgabenstellung in den privaten Schulen von beiden Lehrkräften zunächst gemeinsam erklärt und im Anschluss daran die Klassen geteilt wurden.

¹⁴ Die Stichprobe reduziert sich um zwei weitere Klassen in der mehrebenenanalytischen Untersuchung der Aufgabenstellung und der Lernunterstützung aufgrund geringer Fallzahlen in den Teilklassen bzw. aufgrund von Missings.

¹⁵ Altersangaben liegen nicht für alle Schülerinnen und Schüler vor.

2.2. Überblick über die Publikationen

Das Dissertationsvorhaben umfasst insgesamt fünf Beiträge, in denen die zugrundeliegende Zielsetzung (siehe 1.2) sowie die Forschungsfragen (siehe 1.3) untersucht werden.

Bezugnehmend auf das unter 1.2 vorgestellte theoretische Modell zu den Einflussfaktoren von Textqualität tragen die Publikationen zur Beantwortung der Frage bei, ob und inwieweit der Schreibunterricht in der Schuleingangsstufe einen Effekt auf die Schreibleistung von Grundschülerinnen und Grundschülern besitzt.

Während sich die ersten beiden Publikationen mit der Operationalisierung und Diagnose von Textqualität in der Primarstufe auseinandersetzen, beziehen sich die drei weiteren Manuskripte auf die Aufgabenstellung (Beitrag 3 und 4) sowie die individuelle Lernunterstützung (Beitrag 5) als unterrichtliche Bedingungsfaktoren der Schreibfähigkeit. Alle Beiträge stellen für sich genommen abgeschlossene Analysen dar, bauen aber im Rahmen der übergeordneten Fragestellung chronologisch wie inhaltlich stark aufeinander auf, was auch in der folgenden zusammenfassenden Darstellung deutlich wird.

In Tabelle 1 (S. 15) ist ein Überblick über die Beiträge und deren Veröffentlichungsstatus enthalten. Ebenfalls wird der eigene Anteil im Rahmen der Dissertation gegenüber den Co-Autorinnen und Co-Autoren der Publikationen ausgewiesen.

Tab 1: Publikationen im Überblick

Literaturangaben	Beschreibung des eigenen Anteils	Status
Kürzinger, A. & Pohlmann-Rother, S. (2015). Möglichkeiten einer objektiven und reliablen Bestimmung von Textqualität im Anfangsunterricht – Methodisches Vorgehen und deskriptive Befunde aus dem Projekt NaSch1. <i>Didaktik Deutsch</i> , 20 (38), 60–79.	Konzeption, Durchführung und statistische Auswertung der Teilstudie 1 im Rahmen der Dissertation sowie ihre federführende Darstellung in Beitrag 1; Beratung und Supervision bei der Konzeption und der Verfassung des Beitrages durch Frau Prof. Dr. Pohlmann-Rother	Veröffentlicht; im Anhang beigefügt
Pohlmann-Rother, S., Schoreit, E. & Kürzinger, A. (2016). Schreibkompetenzen von Erstklässlern quantitativ-empirisch erfassen – Herausforderungen und Zugewinn eines analytisch-kriterialen Vorgehens gegenüber einer holistischen Bewertung. <i>Journal for Educational Research Online</i> . 8 (2), 107–135.	Konzeption, Durchführung und statistische Auswertung der Texte (Teilstudie 1) im Rahmen der Dissertation als Vorarbeit für Beitrag 2; Berechnung von Rasch-Modellen als Vorarbeit für Beitrag 2; Unterstützung bei der Aufarbeitung des Theorie- und Forschungshintergrundes; Verfassung der Stichproben-, Erhebungs- und Instrumentenbeschreibung im Manuskript; Berechnung und Darstellung der statistischen Analysen im Manuskript durch Herrn Dr. Schoreit; Verfassung des Theorie- und Forschungshintergrundes durch Frau Prof. Dr. Pohlmann-Rother	Veröffentlicht; im Anhang beigefügt
Kürzinger, A., Pohlmann-Rother, S. & Hess, M. (2017). Aufgabenstellung als zentrale Gelenkstelle im Schreibunterricht – Eine videogestützte Analyse. <i>Zeitschrift für Pädagogik</i> , 63 (5), 636–656.	Konzeption, Durchführung und statistische Auswertung der Teilstudie 3 im Rahmen der Dissertation sowie ihre federführende Darstellung in Beitrag 3; Beratung und Supervision bei der Konzeption und der Verfassung des Beitrages durch Frau Prof. Dr. Pohlmann-Rother; Unterstützung bei der Aufbereitung des Theoriehintergrundes durch Frau Prof. Dr. Pohlmann-Rother; Beratung bei der Verfassung des Manuskripts durch Frau Dr. Hess	Veröffentlicht; im Anhang beigefügt
Kürzinger, A., Pohlmann-Rother, S., Hess, M. & Lipowsky, F. (2018). Aufgabenstellungen und ihre Bedeutung für die Textqualität von Schülerinnen und Schülern im Anfangsunterricht. <i>Unterrichtswissenschaft</i> , 46 (3), 327–343. doi: https://doi.org/10.1007/s42010-018-0016-x .	Konzeption, Durchführung und statistische Auswertung der Teilstudie 4 im Rahmen der Dissertation sowie ihre federführende Darstellung in Beitrag 4; Beratung und Supervision bei der Konzeption und der Verfassung des Beitrages durch Frau Prof. Dr. Pohlmann-Rother; Methodische Unterstützung bei der Berechnung der Mehrebenenanalysen durch Frau Dr. Hess; Beratung durch Herrn Prof. Dr. Lipowsky	Veröffentlicht; nicht im Anhang beigefügt
Pohlmann-Rother, S., Kürzinger, A. & Lipowsky, F. (2018). Individuelle Lernunterstützung im schriftsprachlichen Anfangsunterricht: Formen, Verteilungsmuster und Wirksamkeit. <i>Zeitschrift für Grundschulforschung</i> , 11 (2), 315–332. doi: https://doi.org/10.1007/s42278-018-0024-2 .	Konzeption, Durchführung und statistische Auswertung der Teilstudie 5 im Rahmen der Dissertation; Berechnung der Analysen und Darstellung des empirischen Teils (Methodik und Ergebnisse) sowie der Zusammenfassung und Diskussion in Beitrag 5; Supervision bei der Konzeption sowie Darstellung des Theorie- und Forschungsstandes im Manuskript durch Frau Prof. Dr. Pohlmann-Rother	Veröffentlicht; nicht im Anhang beigefügt

2.3. Zusammenfassende Darstellung der Publikationen innerhalb des Forschungsvorhabens

Beitrag 1: Möglichkeiten einer objektiven und reliablen Bestimmung von Textqualität im Anfangsunterricht – Methodisches Vorgehen und deskriptive Befunde aus dem Projekt NaSch1

Der erste Beitrag verfolgt das Ziel, Schreibfähigkeiten von 540 Erstklässlerinnen und Erstklässlern¹⁶ unter Berücksichtigung fachdidaktischer sowie empirischer Anforderungen zur Modellierung und Erfassung des Konstrukts Schreibkompetenz zu untersuchen. Im Fokus steht dabei die Frage, wie sich Schreibkompetenz im Anfangsunterricht operationalisieren, messen und ausweisen lässt. Vor dem Hintergrund testtheoretischer Annahmen und der Interpretation von Schreibaufgaben als Schulleistungstests können Schreibprodukten eine gewisse Qualität und den Schreibenden gewisse Fähigkeiten (Kompetenzen) zugeschrieben werden (Harsch et al., 2007; Neumann, 2007; Knopp et al., 2012). In Anlehnung an dieses Kompetenz-Performanz-Modell wird in der vorliegenden Arbeit eine gezeigte Leistung (Textqualität) als Indikator für eine entsprechende dahinterliegende Schreibkompetenz angenommen (Harsch et al., 2007). Mehrheitlich lässt sich Schreibkompetenz als Zusammenspiel verschiedener Teilkompetenzen, als Kombination von Wissen und Fähigkeiten und zugleich als Teil einer umfassenderen Sprachkompetenz verstehen (Becker-Mrotzek, 2014; Becker-Mrotzek & Schindler, 2007, S. 12-17; Knopp et al., 2012, S. 47-49; Fix, 2008, S. 24). Im Hinblick auf diese Komplexität erscheint es sinnvoll, das jeweilige Verständnis von Schreibkompetenz an dem Schreibanlass und der darauf bezogenen Textfunktion auszurichten (vgl. auch Harsch et al., 2007). Vor dem Hintergrund dieses Postulats wird Schreibkompetenz im Rahmen des Dissertationsprojekts als altersgemäße (auf das „Schreibalter“ bezogen, vgl. Feilke, 1996, S. 1181¹⁷) Fertigkeit begriffen, die mit einer Schreibaufgabe verbundenen Anforderungen zu bewältigen. Darauf aufbauend wurde ein vereinfachtes Modell von Textqualität entwickelt, das vor allem auf sprachsystematische und kohärenzstiftende Kriterien abzielt sowie die geringe Schreiberfahrung der Kinder und den marginalen Textumfang¹⁸ ($M = 17.5$ Wörter; $SD = 9.59$) berücksichtigt. Die Operationalisierung des Konstrukts erfolgte anhand eines Kriterienkatalogs und eines holistischen Ratings auf Basis zweier Testinstrumente, die aufgrund der möglichen Sicherung der Qualitätseinschätzungen über Beurteilerübereinstimmungen bzw. Raterreliabilitäten in der Literatur als empirisch fundiert

¹⁶ Aus der Stichprobe von 657 Schülerinnen und Schülern (siehe 2.1) lassen sich nur 540 Texte analysieren.

¹⁷ Mit diesem Terminus wird auf die Schreiberfahrung der Schülerinnen und Schüler Bezug genommen, nicht auf das biologische Alter.

¹⁸ Der geringe Umfang der Texte erwies sich als besondere Herausforderung für die Entwicklung geeigneter Kriterien, da die Ausprägung inhaltliche-sprachlicher Merkmale wie etwa Implizitheit von der Textlänge bedingt werden.

gelten (z. B. Baumann, 2006; Grzesik & Fischer, 1984). In Anbetracht des Umstandes, dass die beiden Instrumente bislang überwiegend für schreiberfahrenere Schülerinnen und Schüler konzipiert wurden, erfolgte eine deutliche Modifikation der Testinstrumente an die zugrundeliegenden Rahmenbedingungen des Forschungskontexts. So setzt sich das holistische Rating, das sich an den NAEP Holistic Scoring Guide (Persky, Daane & Jin, 2003) sowie an das holistische Analyseinstrument von Böhme und Kollegen (2009) orientiert, beispielsweise aus einer vier- statt einer sechsstufigen Skala zusammen. Auch die formalstrukturellen, inhaltlichen, sprachlichen und sprachsystematischen Merkmale des Kriterienkatalogs, der sich z. B. auf den Arbeiten von Kruse und Kollegen (2012), Nussbaumer und Sieber (1994, 1995) sowie Steinig, Betzel, Geider und Herbold (2009) stützt, wurden an die kaum ausgeprägte Schreiberfahrung der Schülerinnen und Schüler angepasst und dementsprechend in ihrer Komplexität reduziert.

Kennwerten der Beobachterübereinstimmung sowie Inter- bzw. Intraraterreliabilität in Beitrag 1 zufolge lässt sich mittels der beiden Messinstrumente eine reliable Einschätzung der Textqualität umsetzen. Gerade vor dem Hintergrund der zum Teil als problematisch dokumentierten Sicherung der Gütekriterien bei der Textbewertung (z. B. Blatt et al., 2009) kann dieses Ergebnis als überraschend betrachtet werden¹⁹. Die Befunde zu den Schreibfähigkeiten der Erstklässlerinnen und Erstklässler belegen in Konsens mit zwei qualitativ ausgerichteten Untersuchungen zur Schreibkompetenz im Anfangsunterricht (Schulte, 2000; Weinhold, 2000), dass bereits Schreibanfängerinnen und Schreibanfänger über funktional-pragmatische Schreibfähigkeiten verfügen und eine inhaltlich wie sprachlich kohärente Textwelt aufbauen können. So verfassen die Schülerinnen und Schüler in Klasse 1 – auf Basis der kriterialen Analyse der Textqualität – durchschnittlich gut verständliche ($M = 2.9$; $SD = 0.96$; Skala 1 - 4; $N = 540$), sprachlich angemessene ($M = 2.7$; $SD = 0.78$; Skala 1 - 4; $N = 540$) sowie differenzierte ($M = 2.7$; $SD = 0.80$; Skala 1 - 4; $N = 540$) Texte. Die holistische Erfassung der Textqualität kommt mit einem Mittelwert von $M = 2.4$ ($SD = 0.89$; Skala 1 - 4; $N = 540$) zu einem ähnlichen Ergebnis.

Beitrag 2: Schreibkompetenzen von Erstklässlern quantitativ-empirisch erfassen – Herausforderungen und Zugewinn eines analytisch-kriterialen Vorgehens gegenüber einer holistischen Bewertung

Aufbauend auf diesen Erkenntnissen zur Operationalisierung und Charakterisierung von Schreibfähigkeiten im Anfangsunterricht steht im zweiten Beitrag²⁰ – neben der

¹⁹ Die (marginale) Textlänge als potenzieller Einflussfaktor für die Übereinstimmung zwischen den Beobachtern wird im Beitrag 1 ausführlich thematisiert und Rahmen der methodischen Diskussion auf S. 38-39 knapp dargelegt.

²⁰ Alle folgenden Ausführungen zu Beitrag 2 beziehen sich auf Pohlmann-Rother, Schoreit & Kürzinger (2016).

Gegenüberstellung der beiden Textbewertungsinstrumente – die Frage nach der Dimensionalität der Schreibkompetenz im Vordergrund. Qua dem Verständnis von Schreibhandlung als Konglomerat verschiedener Teilkompetenzen bestehen für Schülerinnen und Schüler divergente Herausforderungen beim Verfassen von Texten, die sich etwa in motorischen und orthografischen Anforderungen einerseits und sprachlichen, konzeptionellen Ansprüchen andererseits manifestieren. Während die Ausprägung dieser unterschiedlichen Teilkompetenzen bei schreiberfahrenen Schülerinnen und Schülern gut dokumentiert ist (z. B. IEA-Aufsatzstudie, vgl. Lehmann, 1987,1990; DESI-Studie, vgl. Neumann, 2007), wurde die Struktur der Schreibkompetenz bei Schreibanfängerinnen und Schreibanfängern dagegen kaum erfasst.

Um diese Dimensionalität zu analysieren, wurden in der vorliegenden Stichprobe auf Basis des kriterialen Textbewertungsinstruments (15 Kriterien) verschiedene ein- und zweidimensionale Partial-Credit-Modelle in ConQuest 2 (Wu, Adams, Wilson & Haldane, 2007) gegenübergestellt. In Einklang mit der empirisch wiederholt aufgefundenen Unterteilung in zweidimensionale Schreibfähigkeiten bei älteren Schülerinnen und Schülern (Lehmann, 1987, 1990; Neumann, 2007; Böhme et al., 2009), legen auch die latenten Korrelationen und entsprechende Fit-Indizes der Modelle eine zweidimensionale Struktur in Klasse 1 nahe. Rein technische Aspekte des Schreiben-Könnens (sprachsystematischer Faktor, z. B. Grammatik, Orthografie) lassen sich demnach bereits bei Schreibanfängerinnen und Schreibanfängern von einer eher inhaltlich-stilistischen Sprachbeherrschung (semantisch-pragmatischer Faktor, z. B. Textaufbau, Wortwahl) differenzieren. Aus schreibdidaktischer Perspektive deutet dieser interessante Befund darauf hin, dass höherrangige Prozesse der Ideenentwicklung und -strukturierung sowie der sprachlichen Gestaltung bei Schreibanfängerinnen und Schreibanfängern demnach nicht automatisch von basalen technischen Ausführhandlungen des Textschreibens eingeschränkt werden, sodass die Lernenden durchaus über freie mentale Kapazitäten für die inhaltlich-stilistische Sprachbeherrschung verfügen.

In dem Potenzial der kriterialen Textbewertung, Teildimensionen von Schreibkompetenz identifizieren zu können, besteht ein Vorzug gegenüber dem holistischen Instrument, das Textqualität lediglich eindimensional abzubilden vermag. Die Gegenüberstellung dieser beiden Instrumente als weiterer zentraler Gegenstand in Beitrag 2 wurde über multiple Regressionen mit den beiden Skalen des Kriterienkatalogs als Prädiktoren und der holistischen Skala als Kriterium in Mplus (Muthèn & Muthèn, 1998-2010) untersucht. Dabei zeigt sich zunächst eine besonders hohe Übereinstimmung ($r = .78$) zwischen der semantisch-pragmatischen Dimension und der holistischen Auswertungsstrategie. Aber auch die sprachsystematische Dimension weist einen mittleren bis hohen Zusammenhang ($r = .47$) mit dem holistischen Instrument auf. Angesichts der knappen textuellen Entfaltung

der Schreibprodukte ($M = 17.5$ Wörter; $SD = 9.59$) und der Beziehung zwischen Textlänge und Benotung durch Lehrkräfte (Birkel & Birkel, 2002) wird in diesem Kontext auch ausführlich die Rolle der Textlänge als Einflussgröße beleuchtet, die bereits in Beitrag 1 im Zusammenhang mit den hohen Beobachterübereinstimmungen diskutiert wurde²¹. Während die Textlänge eine gleichermaßen hohe Übereinstimmung mit dem semantisch-pragmatischen Faktor ($r = .63$) und dem holistischen Rating ($r = .62$) aufweist, lässt sich jedoch kein bedeutsamer Zusammenhang mit der sprachsystematischen Schreibkompetenz ($r = .09$) feststellen. Dieses Ergebnis lässt sich auf Basis der vorhandenen Daten nicht weiter systematisieren. Denkbar wäre jedoch, dass etwa kurze, weitgehend „fehlerfreie“ Texte gerade aufgrund ihrer Kürze fehlerfrei blieben, womit die geringe Textlänge sozusagen das Resultat einer erfolgreichen Fehlervermeidung abbildet. In diesem Kontext könnte für weitere Untersuchungen auch die Frage, inwiefern die Textlänge das Resultat der Schreibkompetenz bedeutet oder als Limitation der Schreibkompetenz fungiert, einen interessanten Aspekt darstellen.

Insgesamt verweisen die Befunde darauf, dass beide Messinstrumente geeignet erscheinen, um Schreibfähigkeiten von Erstklässlerinnen und Erstklässlern reliabel zu erfassen. So deutet nichts darauf hin, dass die jeweilige Art der Bewertung (holistisch oder kriterial) eine unterschiedliche Anfälligkeit auf die Beeinflussung durch die Textlänge (als entscheidendes Moment bei Texten in Klasse 1) aufweist. Jedoch lässt sich mit dem Nachweis einer zweidimensionalen Struktur der Schreibkompetenz auf Basis der kriterialen Textbewertung ein klarer Zugewinn gegenüber dem holistischen Rating konstatieren.

Beitrag 3: Aufgabenstellung als zentrale Gelenkstelle im Schreibunterricht – Eine videogestützte Analyse

Auf welche Weise sich Aufgabenstellungen im schriftsprachbezogenen Anfangsunterricht videogestützt erfassen und charakterisieren lassen, wird in Beitrag 3 untersucht. Mit dieser Fragestellung wird Bezug auf die aktuelle schreibdidaktische Diskussion genommen, in der die Bedeutung von Aufgaben für die Entwicklung der Schreibkompetenz derzeit eine große Resonanz erfährt (z. B. Bachmann & Becker-Mrotzek, 2010; Hoppe & Metz, 2013; Metz, Maier, Kleinknecht, Bohl & Hoppe, 2012). Im Zentrum der Debatte stehen jedoch vor allem Fragen, wie Aufgaben generiert und kategorisiert werden können sowie, welche Merkmale lernförderliche Schreibaufgaben ausweisen. Damit wird hauptsächlich die Qualität von Schreibaufgaben in den Blick genommen, während ihre konkrete Umsetzung im Unterricht weniger erforscht ist. Doch die Qualität von Schreibaufgaben ist unmittelbar an die Implementierung im Unterricht gekoppelt. So können Schreibaufgaben durch instruktionales Lehrerhandeln verändert werden, etwa indem sie „klein gearbeitet“ und dadurch ihre

²¹ In der methodischen Diskussion wird auf S. 38-39 knapp auf diese Thematik eingegangen.

Komplexität und die damit verbundenen Anforderungen verringert werden (Neubrand, 2002; Stein et al., 1996; Blömeke et al, 2006; vgl. auch Bohl & Kleinknecht, 2009).

Beitrag 3 stellt in diesem Kontext die Analyse der Aufgabenstellung im Anfangsunterricht in den Mittelpunkt, worunter in einem weiten Begriffsverständnis die Implementierung und didaktisch-methodische Aufbereitung der Schreibaufgabe als auch die konkrete Formulierung des Schreibauftrags durch die Lehrperson verstanden wird. Die Entwicklung des videobasierten Beobachtungsinstruments, das einer Überprüfung der Reliabilität standhält, stützt sich vor allem auf die kompetenzorientierte Schreibdidaktik und auf „Aufgaben mit Profil“ (Bachmann & Becker-Mrotzek, 2010). Diese didaktischen Entwürfe betonen neben der Bedeutung des Lebensweltbezugs bzw. der Authentizität und der Bereitstellung von Gelegenheiten zur Wissensgenerierung über den Schreibinhalt vor allem die Integration der Schreibaufgabe in eine soziale Interaktion sowie eine Zielorientierung (z. B. Abraham & Kupfer-Schreiner, 2007; Bachmann & Becker-Mrotzek, 2010; Metz, Maier & Kleinknecht, 2009; Metz et al., 2012).

In dem vorliegenden Dissertationsvorhaben wird daher insbesondere auf die Zielorientierung und die Einbettung der Schreibaufgabe in eine soziale Interaktion fokussiert²². Die Zielorientierung wird an einer klar definierten Kommunikationssituation, in der der Zweck des Schreibens, der Autor, Rezipient und Textart vermittelt werden (Schneuwly, 1996), sowie an dem Präzisionsgrad und der Verständlichkeit des Schreibauftrags festgemacht und weist damit Berührungspunkte mit Klarheit und Strukturiertheit als fachübergreifende Merkmale der Unterrichtsqualität auf (z. B. Helmke, 2014; Helmke, Helmke & Schrader, 2007). Mit der Einbettung der Schreibaufgabe in eine soziale Interaktion wird auf kommunikativen Austausch im Unterricht verwiesen, der vor allem für ungeübte Schreibende essenziell ist, die die „Zerdehnung der Sprechsituation“ (Ehlich, 1983), also die Trennung von Empfänger und Sender beim Schreiben, überwinden müssen (vgl. auch Bachmann & Becker-Mrotzek, 2010).

Basierend auf den deskriptiven Befunden lässt sich festhalten, dass die Zielorientierung von Lehrpersonen in der Stichprobe vermehrt umgesetzt wird, während soziale Interaktionen im Rahmen der Aufgabenstellung kaum realisiert werden (Ausnahme: Sammeln von Schreibideen). Diese kontrastierenden Ergebnisse zeichnen sich auch bei weiteren Merkmalen ab: So weisen allgemeine Kriterien der Unterrichtsqualität wie Präzision und Verständlichkeit, Kohärenz und Fokussierung sowie die Organisation des Schreibarrangements eine durchschnittlich ausgeprägte Qualität auf, wohingegen kompetenz- und schreibdidaktisch orientierte Kriterien wie der Lebensweltbezug oder

²² Die Zielorientierung wird über vier einzelne Items (Schreibfunktion, Verständlichkeit, Präzisionsgrad sowie Ideensammlung zum Abschiedsthema) erfasst. Die Einbettung der Aufgabenstellung in eine soziale Interaktion setzt sich aus drei Einzelitems (Austausch über Schreibideen, soziale Interaktion im engeren Sinn, Wechsel der Sozialform) zusammen.

Differenzierungsmöglichkeiten teilweise vernachlässigt werden. Diese erwartungswidrigen Ergebnisse zur schreibdidaktischen Gestaltung der Aufgabenstellung lassen sich unter Umständen in Bezug auf das fachdidaktische Wissen von Lehrkräften interpretieren: Möglicherweise fehlt den Lehrpersonen das Wissen bzw. das Bewusstsein darüber, welche schreibdidaktischen Kriterien lernförderlich sind und auf welche Weise sie im Schreibsetting integriert werden können. Insgesamt gründet sich der Erkenntnisgewinn des dritten Beitrags darauf, dass für den Anfangsunterricht empirisch dokumentiert wird, wie Lehrkräfte Aufgabenstellungen in Klasse 1 gestalten, welche Schwerpunkte sie setzen und welchen Merkmalen sie weniger Gewicht verleihen. Zudem wurde eine Möglichkeit aufgezeigt, wie sich Aufgabenstellungen im Anfangsunterricht konzeptualisieren und operationalisieren lassen.

Beitrag 4: Aufgabenstellungen und ihre Bedeutung für die Textqualität von Schülerinnen und Schülern im Anfangsunterricht

Anknüpfend an theoretische Bezüge und an das Analyseinstrument der Aufgabenstellung in Beitrag 3 ergänzt der vierte Beitrag die deskriptiven Befunde um mehrebenenanalytische Untersuchungen zum Zusammenhang zwischen der Aufgabenstellung und der Textqualität von Erstklässlerinnen und Erstklässlern²³. Vor dem Hintergrund des Forschungsdesiderats und der Bedeutung klar definierter, strukturierter Aufgabenstellungen wird unter Kontrolle der schriftsprachbezogenen Vorläuferfähigkeiten (Buchstabenkenntnis²⁴), des Geschlechts sowie der mittleren Leistungsstärke der Klasse im Lesen untersucht, wie sich die Erläuterung der Schreibfunktion (Abschied von Lucy) sowie die Verständlichkeit und der Präzisionsgrad einer Aufgabenstellung auf die Schreibfertigkeiten der Schülerinnen und Schüler auswirken. Um diese Zusammenhänge zu erfassen, wurden Mehrebenenanalysen mit HLM 7.0 (Raudenbush, Bryk & Congdon, 2008) berechnet, die die hierarchische Struktur der Daten (Schüler- und Klassenebene) berücksichtigen. Als abhängige Variable auf Ebene 1 wurde die Textqualität (holistisches Rating) modelliert, während als unabhängige Variablen auf Ebene 2 der Präzisionsgrad, die Verständlichkeit sowie die beiden Kriterien zur Schreibfunktion (Thematisierung der Schreibfunktion durch die Lehrkraft, Thematisierung der

²³ Im Unterschied zu Beitrag 3 werden die vier Kriterien der „Zielorientierung“ (Schreibfunktion, Verständlichkeit, Präzisionsgrad sowie Ideensammlung zum Abschiedsthema) einzeln analysiert und ihre Bezeichnung teilweise prägnanter umformuliert. So wird die Schreibfunktion in Beitrag 4 als „Thematisierung der Schreibfunktion durch die Lehrkraft“ benannt und die „Ideensammlung zum Abschiedsthema“ als „Thematisierung der Schreibfunktion durch die Lernenden“ bezeichnet. Die Kriterien der sozialen Interaktion werden in Beitrag 4 nicht weiter untersucht, da die deskriptiven Analysen in Beitrag 3 zeigen, dass soziale Interaktionen während der Aufgabenstellung kaum realisiert wurden.

²⁴ Die Buchstabenkenntnis wurde im Projekt PERLE über den Untertest „Buchstaben lesen“ des LEst 4-7 (Moser, Berweger & Lüchinger-Hutter, 2004) erfasst (Gresser, Pohl, Corvacho del Toro, Greb & Faust, 2009), und kann als Merkmal einer allgemeinen Buchstabenkenntnis betrachtet werden, die auch für die Textproduktion erforderlich ist (Lotz, Schoreit & Kempter, 2013d).

Schreibfunktion durch die Lernenden) schrittweise eingeflossen sind. Auf Ebene 2 wurde der gemittelte und aggregierte Wert für die Buchstabenkenntnis als Indikator für die mittlere Leistungsstärke der Lerngruppe im Lesen berücksichtigt, während auf Ebene 1 das Geschlecht sowie die Buchstabenkenntnis kontrolliert wurden.

Die Befunde des Individualmodells deuten darauf hin, dass sich das Geschlecht der Kinder ($\beta = .18$; $p < .01$) zugunsten der Mädchen sowie die Buchstabenkenntnis ($\beta = .35$; $p < .01$) signifikant auf die Textqualität auswirken. In Einklang zu den Erwartungen kristallisiert sich heraus, dass eine präzise ($\beta = .14$; $p = .01$) und verständlich ($\beta = .20$; $p = .01$) formulierte Aufgabenstellung, bei der die Schreibfunktion klar kommuniziert wird, einen signifikant positiven Effekt auf die Schreibleistung besitzt. Dieses Ergebnis steht auch in Konsens zu Studien, die als Merkmale dieser Facetten Klarheit und Strukturiertheit als entscheidende Qualitätsmerkmale guten Unterrichts hervorheben (z. B. Helmke, 2014). Die Verständlichkeit der Aufgabenstellung manifestiert sich beispielsweise in der einfachen Formulierung eines Arbeitsauftrags und in der Anpassung an den sprachlogischen Wortschatz von Erstklässlerinnen und Erstklässlern. Ein präzise formulierter Schreibauftrag wird an einer exakten Erläuterung des Adressaten, des Absenders und der Textsorte deutlich, womit zum Verständnis von Arbeitsaufträgen und zur Klarheit der Unterrichtsgestaltung beigetragen wird.

Ein überraschendes Resultat tritt im Zusammenhang mit den Akteuren auf, welche die Schreibfunktion während der Aufgabenstellung thematisieren: Während die Explikation der Schreibfunktion durch die Lehrkraft keinen Effekt auf die Schreibleistung ($\beta = -.00$; $p = .97$) aufweist, zeigt sich bei der Besprechung der Schreibfunktion durch die Schülerinnen und Schüler ein signifikant positiver Zusammenhang ($\beta = .26$; $p < .05$). Dies könnte darauf hindeuten, dass die aktive, unter den Lernenden kommunizierte Auseinandersetzung mit dem Abschied (= Schreibfunktion) ein intensiveres Nachdenken bewirken und zu einer tiefergehenden Verarbeitung der Inhalte führen kann. Daraus zu schließen, dass die Rolle der Lehrperson weniger bedeutsam ist, wäre jedoch verkürzt. So könnte vor allem eine entsprechende Anregung durch die Lehrkräfte die Besprechung des Abschieds bei den Schülerinnen und Schülern auslösen.

Als Resümee lässt sich festhalten, dass sich aus den Befunden, die den domänenspezifischen Forschungsdesiderata im schriftsprachlichen Anfangsunterricht Rechnung tragen, Erkenntnisse über die Gestaltung kompetenzorientierter Aufgabenstellungen ableiten lassen.

Beitrag 5: Individuelle Lernunterstützung im schriftsprachlichen Anfangsunterricht: Formen, Verteilungsmuster und Wirksamkeit

Ausgehend von der Bedeutung unterrichtlicher Prozessvariablen auf die Schreibfähigkeiten wird in Beitrag 5 die Schülerarbeitsphase in den Blick genommen, indem die Formen und Verteilungsmuster individueller Lernunterstützungen sowie deren Einfluss auf die Textqualität analysiert werden. In der aktuellen Lehr-Lernforschung gilt die Unterstützung individueller Lernprozesse als eine Form der Differenzierung, nämlich als Mikroadaptation des Unterrichts, die den Fokus auf eine optimale Förderung der Kinder richtet (Krammer, 2009; Lotz, 2016; vgl. auch Pohlmann-Rother & Kürzinger, 2015). Im vorliegenden Forschungsvorhaben wird unter individueller Lernunterstützung in Anlehnung an Krammer (2009) die

... Unterstützung von einzelnen oder einer kleinen Gruppe von Lernenden durch eine Person mit mehr Expertise beim Aufbau von Wissen respektive beim Erwerb von Kompetenzen zum Lösen einer Aufgabe oder eines Problems im Hinblick auf das zukünftige selbständige Bewältigen analoger Aufgaben und Probleme durch die Lernenden ... (S. 89)

verstanden. In Einklang mit dem vorherrschenden Forschungsdefizit im schriftsprachlichen Anfangsunterricht der Grundschule lässt sich auch für individuelle Lernunterstützungen konstatieren, dass gegenwärtig kaum eine systematische, quantitative Untersuchung für den Schreibunterricht in der Schuleingangsstufe im deutschsprachigen Raum vorliegt²⁵. Existierende Studien beziehen sich überwiegend auf den Sekundarbereich und deuten darauf hin, dass die individualisierte Unterstützung von Lernenden positive Effekte auf ihre schulische Leistung sowie auf Einstellungen, Interesse und Motivation besitzt (Zsf. van de Pol et al., 2010; Kobarg & Seidel, 2007; Seidel & Shavelson, 2007). Kompetenzförderlich ist dabei vor allem die Anregung von Denkprozessen, etwa durch Hinweise (z. B. Chi et al., 2001; Krammer, 2009; Lepper, Drake & O'Donnell-Johnson, 1997; Pauli & Reusser, 2000). Zur Erfassung des Unterstützungsverhaltens in der vorliegenden Arbeit wurde ein Beobachtungssystem mit 15 Kriterien entwickelt, welches an Krammer (2009) sowie Lotz (2016) angelehnt ist. Die Analyse basiert auf einem detaillierten, zweistufigen Verfahren, das jede Lehrer-Schüler-Interaktionen dahingehend charakterisiert, ob bzw. welche Art der Lernunterstützung enthalten ist. Der Zusammenhang zwischen individueller Lernunterstützung und Schreibleistung ist – ähnlich wie in Beitrag 4 – auf Basis eines Mehrebenenmodells berechnet worden. Als abhängige Variable auf Ebene 1 wurde erneut das holistische Rating der Textqualität modelliert, während als unabhängige Variablen auf Ebene 2 verschiedene Items der individuellen Lernunterstützung sowie der gemittelte und

²⁵ Im Grundschulbereich wurden aktuell Studien zur individuellen Lernunterstützung im naturwissenschaftlichen Sachunterricht (Schnebel & Wagner, 2016) und zum Leseunterricht des ersten Schuljahres (Lotz, 2016) veröffentlicht.

aggregierte Wert der Buchstabenkenntnis als Indikator für die mittlere Leistungsstärke der Klasse im Lesen eingeflossen sind. Bei den Kriterien der Lernunterstützung wurden diejenigen Formen und inhaltlichen Schwerpunkte ausgewählt, die sich im Rahmen der deskriptiven Analyse als dominierend herausgestellt haben (Evaluationen, einfache Hinweise, Lösungen sowie die inhaltlichen Schwerpunkte Orthografie und Schreibinhalt). Da die Aufgabenstellung die Qualität der Texte beeinflusst (Beitrag 4), wurde zudem die Verständlichkeit der Aufgabenstellung als Einflussfaktor berücksichtigt. Als Kontrollvariablen auf Ebene 1 fungieren die Buchstabenkenntnis sowie das Geschlecht der Kinder.

Neben der reliablen Messqualität des Instruments zeigt sich aus deskriptiver Perspektive zunächst, dass die Lehrkräfte in der Stichprobe die Erstklässlerinnen und Erstklässler beim Schreiben eigener Texte relativ häufig unterstützen ($M = 1.64$ Hilfestellungen pro Minute; $SD = 0.71$) und überwiegend Lösungen (33.4 %), einfache Hinweise (29.3 %) sowie evaluative Unterstützungsformen (13.1 %) erteilen. Hinsichtlich des inhaltlichen Fokus werden Lernunterstützungen vor allem zu orthografischen Fragen (41.7 %) und zum Schreibinhalt (37.9 %) gegeben. Zudem lassen sich deutliche Unterschiede zwischen den Lehrpersonen bezüglich Häufigkeit, Form und Inhalt der Lernunterstützung feststellen.

Erwartungskonforme Befunde kristallisieren sich bei der Verteilung der Unterstützungsmaßnahmen auf einzelne Schülerinnen und Schüler heraus. So erhalten Jungen, die im Durchschnitt eine geringere Textqualität als Mädchen aufweisen²⁶, tendenziell mehr Hilfestellungen ($r_s = -.11^{27}$; $p = .01$; $N = 657$, Spearman-Rho-Korrelation) sowie eher einfache Hinweise ($M = 1.4$; $SD = 2.27$; t ($df. 427$) = 2.25; $p = .01$) als ihre Mitschülerinnen ($M = .93$; $SD = 2.81$). Ebenso beziehen sich die Hilfestellungen für Jungen ($M = 1.9$; $SD = 2.81$; t ($df. 427$) = 3.6; $p = .00$) signifikant häufiger auf Schreibinhalte als bei den Mädchen ($M = 1.1$; $SD = 1.81$).

Interessanterweise werden die Lernunterstützungen in der Stichprobe mehrheitlich von den Lernenden (60.16 %) selbst initiiert, sodass die Schülerinnen und Schüler vielmehr selbst die Verteilung von Lernunterstützungen und ihre Teilhabe daran steuern. Dabei lassen sich keine Zusammenhänge zwischen der Initiierung durch die Lernenden und der schriftsprachbezogenen Vorläuferfähigkeit (Buchstabenkenntnis) nachweisen. Im Gegensatz dazu orientieren sich die Lehrkräfte, sofern die Initiative für eine Lernunterstützung von ihnen ausgeht ($M = 38.58$ %), an den Lernausgangslagen der Schülerinnen und Schüler (Spearman-Rho-Korrelation mit Buchstabenkenntnis: $r_s = -.14$; $p = .01$; $N = 418$). Bei geringer Leistungsfähigkeit der Kinder ergreift eine Lehrkraft somit tendenziell eher die Initiative für eine Lernunterstützung.

²⁶ In der vorliegenden Stichprobe ist die Textqualität bei Mädchen ($M = 2.4$; $N = 293$; $SD = 0.88$) signifikant höher als bei Jungen ($M = 2.3$; $N = 231$; $SD = 0.89$). Die Effektstärke d_{cohen} beträgt .22 (Kürzinger et al., 2018).

²⁷ Die Korrelation bezieht sich auf den Zusammenhang zwischen der Anzahl der Hilfestellungen und dem Geschlecht (Kodierung Jungen = 0; Mädchen = 1).

Auf Basis der Mehrebenenanalysen zeichnet sich ab, dass das Benennen der korrekten Lösung²⁸ ($\beta = .16$; $p = .04$) sowie eine evaluative Lernunterstützung ($\beta = .14$; $p = .05$) die Schreibleistung positiv beeinflussen. Einflüsse der Aufgabenstellung (hier mit dem Prädiktor Verständlichkeit) bleiben trotz der Effekte der individuellen Lernunterstützung auf die Schreibleistung entweder auf dem 1 %- oder 5 %-Niveau signifikant bestehen ($\beta = .20$). Ähnlich wie in Beitrag 4 zeichnen sich signifikant positive Effekte der Kontrollvariablen Geschlecht (z. B. $\beta = .23$; $p = .01$) und Buchstabenkenntnis (z. B. $\beta = .39$; $p = .00$) ab. Einfache Hinweise, orthografische Unterstützungen und Hilfestellungen zum Schreibinhalt weisen dagegen keinen Zusammenhang mit der Schreibleistung auf.

Auf Basis der deskriptiven und mehrebenenanalytischen Befunde ermöglicht der Beitrag einen vertieften Einblick in das Unterstützungsverhalten von Lehrkräften im schriftsprachlichen Anfangsunterricht.

²⁸ Die Vergabe der korrekten Lösung tritt vor allem in Verbindung mit der Orthografie ($r_s = .28$; $p = .05$) und der schreibstrategischen Unterstützungen ($r_s = .35$; $p = .01$) auf.

3. Gesamtdiskussion

3.1. Diskussion der zentralen Befunde

Die Dissertation verfolgt das übergeordnete Ziel, unterrichtliche Prozessvariablen als Determinanten der individuellen Schreibleistungen in der Primarstufe zu untersuchen. Resümierend lassen sich die Befunde zur Unterrichtswirksamkeit auf die Schreibperformanz der Schülerinnen und Schüler als überwiegend erwartungskonform bezeichnen. Es erscheint weder überraschend, dass Schreibleistungen von präzisen und verständlichen Aufgabenstellungen positiv beeinflusst werden, noch, dass die Effekte lösungsorientierter Lernunterstützungen auf Textqualität empirisch evident sind. Aus dem Blickwinkel der Unterrichtsforschung werden damit zum Teil bereits bekannte, allgemeindidaktische Befunde zur Prozessqualität von Unterricht gestützt. Jedoch erweitert das Dissertationsprojekt bisherige Erkenntnisse, indem es sich interdisziplinär innerhalb der Schreibdidaktik sowie empirischen Bildungsforschung positioniert und durch den Fokus auf den schriftsprachlichen Anfangsunterricht dezidiert auf deren domänenspezifische Forschungsdesiderata reagiert. Das innovative Moment der Arbeit beschränkt sich damit nicht ausschließlich auf die empirische Wirkungsanalyse von Unterricht auf den bis dato kaum erforschten Kompetenzbereich Schreiben in der Primarstufe. Vielmehr besteht zudem – insbesondere aus didaktischer Sicht – ein konkreter Erkenntnisgewinn der Arbeit darin, sowohl die Ausprägung der Schreibfähigkeiten von Erstklässlerinnen und Erstklässler als auch die Alltagspraxis von Schreibunterricht der ersten Klasse auf breiter, quantitativer Basis zu beschreiben. Damit geht eine Modellierung und Operationalisierung des Konstrukts Schreibkompetenz im Anfangsunterricht einher, die auf theoretischen, schreibdidaktischen Überlegungen fußt und sich im Sinne der mit den Bildungsstandards geforderten Kompetenzorientierung (Klieme et al., 2007) mit einem psychometrischen Ansatz auseinandersetzt. Basierend auf psychologischen Kompetenz-Performanz-Modellen²⁹ (z. B. Erpenbeck & Rosenstiel, 2003; Weinert, 2001) wird dabei die Annahme vertreten, dass die gezeigte Schreibleistung in den Texten einen Indikator für die dahinter liegende Schreibkompetenz darstellt (Harsch et al., 2007; Knopp et al., 2012; Neumann, 2007). Dementsprechend positioniert sich das Verhältnis zwischen manifester Textqualität und latenter Schreibkompetenz in der vorliegenden Arbeit wie folgt³⁰:

²⁹ Die Unterscheidung zwischen Kompetenz und Performanz lässt sich auf die sprachwissenschaftliche Forschungsrichtung unter Chomsky (1968) zurückführen.

³⁰ Im Dissertationsvorhaben wurde die Schreibkompetenz über einen Text erfasst, den die Schülerinnen und Schüler im Rahmen der PERLE-Videostudie geschrieben haben. Die Schreibfähigkeiten der Kinder wurden dabei durch die Bewertung ihres Textes im Rahmen zweier Messinstrumente (holistische und kriteriale Auswertungsstrategie) erfasst, die in Relation zu der Entwicklungsspanne bzw. der in ihr zu erwartenden Schreibleistungen (Pohl & Steinhoff, 2010) entwickelt wurden.

Wird Schreibkompetenz im schulischen Kontext untersucht, zeigt sich die Leistung üblicherweise im Bearbeiten von Schreibaufgaben. Die dabei produzierten Texte werden als Ausweis der erwarteten Fähigkeiten verstanden, einen angemessenen Text zu schreiben. Abhängig davon, ob und wie von Schreibern bestimmte Merkmale im Text realisiert werden, sprechen wir dem Text eine gewisse *Qualität*, dem Schreiber gewisse *Fähigkeiten* (Kompetenzen) zu. (Knopp et al., 2012, S. 51)

In Einklang mit der DESI-Studie liegt dem Forschungsvorhaben ein Verständnis von Schreibkompetenz zugrunde, das sich an dem jeweiligen Schreibanlass, der Textfunktion sowie dem Adressaten ausrichtet: „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. 45). Diese Interpretation wird innerhalb des Dissertationsprojekts vor dem Hintergrund der geringen Schreiberfahrung der Schülerinnen und Schüler um die Komponente des „Schreibalters“ (Feilke, 1996, S. 1181) ergänzt und in ein Modell der Textqualität überführt, dessen Definition im Wesentlichen durch das Untersuchungsinteresse, die Schreiberfahrung der Lernenden und das spezifische schreibdidaktische Setting bedingt werden. Unter Berücksichtigung der Komplexität von Schreibfähigkeiten basieren die 15 Kriterien, aus denen sich die Textqualität zusammensetzt, auf formal-strukturellen, inhaltlichen, sprachlichen sowie sprachsystematischen Teilbereichen. Damit wird auch dem Verständnis Rechnung getragen, das Schreibkompetenz mehrheitlich als Konglomerat verschiedener Teilkompetenzen auffasst (z. B. Becker-Mrotzek, 2014; Knopp et al., 2012), worauf sich auch das kognitionspsychologisch fundierte Schreibentwicklungsmodell Bereiters (1980) gründet. Verschiedenen Modellen der Schreibentwicklungsforschung zufolge, die einen theoretischen Zugang zu den erwartbaren Schreibleistungen ermöglichen, manifestiert sich die Ontogenese der Schreibkompetenz in zwei gegensätzlichen Ansätzen. Während Stufenmodelle (Augst, Disselhoff, Henrich, Pohl & Völzing, 2007; Feilke, 1995; Feilke & Augst, 1989) eine lineare, sukzessive Entfaltung der Schreibfähigkeiten proklamieren, gehen andere Modelle von einer integrativen Entwicklung der Teilkompetenzen (z. B. Bereiter, 1980) aus, nach der sich „... ein einfaches mentales Kompetenzmodell immer weiter ausdifferenziert“ (Blatt, 2011, S. 91).

Wie die Ergebnisse in Beitrag 1 belegen, können bereits Schreibanfängerinnen und Schreibanfänger inhaltlich wie sprachlich kohärente Texte verfassen und damit funktional-pragmatische Schreibfähigkeiten ausbilden, was tendenziell die Vorstellung einer parallelen, mehrdimensionalen Entwicklung der Schreibkompetenz (Becker-Mrotzek & Böttcher, 2011) untermauert: „Die Entwicklung beginnt nicht mit einzelnen Teilfertigkeiten, die dann zusammengefügt werden, sondern parallel in allen Dimensionen. Schreibaufgaben werden

mit den je vorhandenen Kompetenzen bearbeitet, die dann vom Kern her ausgeweitet und ausdifferenziert werden“ (S. 76).

So erweisen sich die Texte der Erstklässlerinnen und Erstklässler auf Basis des kriterialen Messinstruments als durchschnittlich verständlich, sprachlich angemessen sowie sprachlich differenziert. Es zeigt sich im Hinblick auf die Itemschwierigkeiten, dass es den Lernenden vergleichsweise schwerfällt, syntaktisch komplexere Sätze zu verfassen ($\sigma_i = 0.76$; $N = 540$) und Kohäsionsmittel einzusetzen ($\sigma_i = 0.48$; $N = 540$). Dagegen gelingt ihnen beispielsweise die Übernahme einer fremden Perspektive ($\sigma_i = -3.39$; $N = 540$) sowie die Verwendung von Anredeformeln ($\sigma_i = -3.13$; $N = 540$) sehr leicht.

Der in Beitrag 1 dokumentierte Nachweis einfacher wie komplexerer Schreibfähigkeiten im Anfangsunterricht legt den Schluss nahe, Schreibkompetenz bereits von Beginn an zu fördern und den Fokus im Anfangsunterricht nicht nur auf motorische und orthografische Aspekte des Schreibenlernens zu setzen. Diese didaktische Implikation wird auch durch den innovativen Befund in Beitrag 2 gestützt, der auf die Ausprägung der Schreibfähigkeiten bei Erstklässlerinnen und Erstklässlern in zwei Teilkompetenzen (semantisch-pragmatischer Faktor und sprachsystematischer Faktor) verweist. So deutet sich an, dass selbst bei Schreibanfängerinnen und Schreibanfängern die wiederholt vorgefundene Unterteilung der Schreibkompetenz in einen technischen Aspekt (Rechtschreibung) und eine inhaltlich-stilistische Sprachbeherrschung (Inhalt, Aufbau, Sprache) identifiziert werden kann. Die kognitive Aufmerksamkeit und mentale Ressourcen der Kinder sind demzufolge nicht nur auf technisch-handwerkliche Schreibfertigkeiten reduziert (vgl. auch Ellis, 1982), sondern lassen tendenziell auch Kapazitäten für inhaltlich-stilistische Textqualitätsmerkmale frei.

Grundsätzlich stellt das Schreiben von Texten aus kognitionspsychologischer Perspektive eine komplexe Leistung dar, die sowohl das Arbeitsgedächtnis, als auch die metakognitiven Fähigkeiten von Kindern beansprucht. Um sich ausreichend auf die konzeptionelle Verschriftlichung konzentrieren zu können, müssen Schreibende auf graphematische und motorische Schreibroutinen zurückgreifen, über die Erstklässlerinnen und Erstklässler nur bedingt verfügen. Vor dem Hintergrund dieser begrenzten Kapazitäten können Schreibanfängerinnen und Schreibanfänger beim Verfassen eigener Texte entlastet werden, indem beispielsweise die korrekte Orthografie eine untergeordnete Rolle spielt, und eher die Konzentration auf die höherrangigen Prozesse realisiert wird (vgl. Blatt et al., 2005; Kruse et al., 2012). Für die curriculare Praxis des Schreibunterrichts bedeutet dieser Erkenntnisgewinn jedoch nicht, orthografischen und motorischen Lehr-Lernprozessen weniger Gewicht zu verleihen. Diese Interpretation wäre stark verkürzt und in Bezug auf die Ontogenese der Schreibkompetenz fatal, zumal orthografische und motorische Fähigkeiten Effekte auf die Textproduktion besitzen (z. B. Alamargot & Fayol, 2009; Christensen, 2009; McCutchen, 2006). Vielmehr lässt sich aus dem Befund der zweidimensionalen Struktur der

Schreibkompetenz in Klasse 1 die Schlussfolgerung ziehen, dass bereits bei Schreibanfängerinnen und Schreibanfängern Schreibaufgaben eingesetzt werden sollten, die neben dem Erwerb orthografischer Kompetenzen auch auf die textuelle Entfaltung inhaltlich-sprachlicher Fähigkeiten abzielen.

Dies kann beispielsweise dadurch realisiert werden, dass Lehrkräfte Unterstützungsmaßnahmen bei orthografischen und grafomotorischen Schwierigkeiten anbieten. Aus diesem Blickwinkel hat die dokumentierte Fokussierung der Lehrkräfte auf orthografischen Hilfestellungen während des Schreibprozesses (Beitrag 5) vermutlich auch dazu beigetragen, dass sich die Erstklässlerinnen und Erstklässler auch auf die konzeptionelle Verschriftlichung ihrer Ideen konzentrieren konnten, anstatt ihre mentalen Kapazitäten ausschließlich für Motorik und Rechtschreibung einzusetzen.

Dieses komplexe Beziehungsgeflecht demonstriert einmal mehr, dass eine Analyse der Schreibfähigkeiten, die unterrichtliche Prozessvariablen nicht berücksichtigt, zu kurz greift. So kristallisiert sich auch in Bezug auf die dimensionale Struktur der Schreibkompetenz ein Verweis auf die Bedeutung der Aufgabenstellung für Textqualität heraus: Die Differenzierung von sprachsystematischen und semantisch-pragmatischen Fähigkeitsausprägungen deckt sich auch mit der von Quellmalz, Capell und Chou (1982) getroffenen, lediglich anders formulierten Unterscheidung zwischen einem aufgabenspezifischen und aufgabenübergreifenden Faktor: „These data cast doubt on the assumption that ‚a good writer is a good writer‘ regardless of the assignment“ (S. 255).

Im Rahmen der Aufgabenstellung (Beitrag 3 und 4³¹) wirkt sich insbesondere ein präziser und verständlich kommunizierter Schreibauftrag, bei der die Schreibfunktion besprochen wird, signifikant positiv auf die Schreibleistung der Kinder aus. Eine höhere Qualität der Texte kann demnach durch eine klare Definition der Kommunikationssituation, in der das Schreibziel, der Autor, der Rezipient und die Textart (Schneuwly, 1996, S. 31) in einer für Grundschülerinnen und Grundschüler angemessenen sprachlogischen Komplexität formuliert wird, bewerkstelligt werden. Bei den zugrundeliegenden Kriterien Verständlichkeit und Präzisionsgrad eines Schreibauftrags handelt es sich um allgemeinere bzw. fachübergreifende Aspekte der Unterrichtsqualität, die oftmals mit dem Merkmal Klarheit und Strukturiertheit in Relation gesetzt werden (z. B. Helmke, 2014; Helmke, Helmke & Schrader, 2007). Während die Verständlichkeit eher auf die akustische und sprachliche Komponente von Klarheit³² abzielt, wird anhand des Präzisionsgrades die inhaltliche Klarheit erfasst. Schreibdidaktische Kriterien wurden in der vorliegenden Stichprobe von den Lehrkräften

³¹ In Beitrag 3 wurde die Aufgabenstellung deskriptiv untersucht, in Beitrag 4 wurden ihre Effekte auf die Textqualität fokussiert. Aufgrund des gemeinsamen Theoriedesigns und der Konzentration auf spezielle Kriterien der Aufgabenstellung in beiden Beiträgen wird auf eine separate Darstellung in der Diskussion verzichtet.

³² Helmke (2014) unterscheidet vier Komponenten von Klarheit: akustisch, sprachlich, inhaltlich und fachlich.

tendenziell weniger realisiert. So wurden beispielsweise kaum Differenzierungsmaßnahmen aufgegriffen oder Lebensweltbezüge hergestellt. Auch die Einbettung in eine soziale Interaktion, die in vielen Konzeptionen effektiver Schreibsettings bzw. zur Aufgabenqualität (z. B. Bachmann & Becker-Mrotzek, 2010; Blömeke et al., 2006) betont wird, wurde kaum umgesetzt (mit Ausnahme der Besprechung von Schreibideen). Daraus lässt sich jedoch nicht ableiten, dass die Lehrkräfte bewusst auf die Umsetzung dieser Merkmale verzichtet haben. Unter Umständen sind sie im Zusammenhang mit dem speziellen Setting in der vorliegenden Stichprobe bereits indirekt angeklungen. So ist vorstellbar, dass sich aufgrund der Integration der Schreibaufgabe in einen handlungs- und produktionsorientierten Schreibunterricht in anderen Unterrichtsphasen (z. B. bei der Besprechung des Bilderbuchs) Gelegenheiten zur sozialen Interaktion und Lebensweltbezüge ergeben haben, die während der Aufgabenstellung nicht erneut aufgegriffen wurden.

In Bezug auf die individuellen Lernunterstützungen als entscheidende unterrichtliche Einflussfaktoren auf Textqualität zeichnet sich ab, dass das Benennen der korrekten Lösung sowie evaluative Lernunterstützungen die Schreibleistung signifikant positiv beeinflussen (Beitrag 5). Hinweise, die durch die Anregung von Denkprozessen als dezidiert kompetenzförderliche Form von Lernunterstützungen gelten, zeigen keinen Zusammenhang mit der Textqualität.

Diese Befunde widersprechen zunächst Forschungsergebnissen, wonach das Anregen von Denk- und Verstehensprozessen von Lernenden, wie sie durch Hinweise initiiert werden können, allgemein als zielführender als das Erteilen von Instruktionen oder das bloße Benennen von Lösungen gilt: „Zentral ist, dass die Lehrperson die Überlegungen der Lernenden anstößt und auf deren Denkwege eingeht und ihnen nicht nur die Information zur richtigen Lösung des Problems gibt“ (Krammer, 2009, S. 129). Allerdings sollte die jeweilige Situationsangemessenheit der Lernunterstützung berücksichtigt werden, die entscheidend für die Wirksamkeit der Lernunterstützung ist (Krammer, Reusser & Pauli, 2010). So lässt sich vermuten, dass sich die Anregung von Denkprozessen gerade bei Schreibanfängerinnen und Schreibanfängern eher langfristig und weniger unmittelbar in den Schreibprodukten niederschlägt, während Lösungen kurzfristig die Qualität von Texten erhöhen. Dieses Ergebnis ist weniger als Plädoyer für die Erteilung von Lösungen zu verstehen. Vielmehr kann es dahingehend interpretiert werden, dass die Benennung einer korrekten Lösung gerade Schreibanfängerinnen und Schreibanfängern, die mit basalen Anforderungen beschäftigt sind (Grafomotorik, Phonem-Graphem-Korrespondenz), eine Hilfestellung bieten kann, zumal sich in der Stichprobe nachweisen lässt, dass die Lösungsvorgabe vor allem bei orthografischen Schwierigkeiten der Schülerinnen und Schüler sowie bei schreibstrategischen Unterstützungen erfolgt. So vermag die Lösungsvorgabe hinsichtlich der Rechtschreibung und Schreibstrategien den technischen

Schreibprozess entlasten und die Kapazitäten für Ideenentwicklung und die sprachliche Gestaltung erhöhen. Lösungen können folglich unter bestimmten Voraussetzungen und Zielsetzungen durchaus sinnvoll sein, vor allem, wenn sie kalkuliert eingesetzt werden und der Schwerpunkt auf kognitiv anregenden Unterstützungsformen liegt.

Dass sich in der Stichprobe die evaluativen Formen der Lernunterstützung als wirksam erweisen, deckt sich mit Studien, die zeigen, dass Impulse, die auf die spezifische Aufgabe, den aktuellen Arbeitsstand, die nächsten Schritte und die zu erreichenden Ziele fokussieren, effektiver sind als unspezifische personenbezogene Rückmeldungen (Hattie & Timperley, 2007).

Limitationen erfahren die unterrichtlichen Wirkungsanalysen auf Textqualität vor allem dadurch, dass im Kontext des Dissertationsprojekts in den Mehrebenenanalysen bislang lediglich die holistische und damit global erfasste eindimensionale Textqualität als abhängige Variable modelliert wurde. Dieses Vorgehen ist aufgrund der hohen Korrelation zwischen den beiden Messinstrumenten durchaus haltbar. Jedoch würden Analysen mit der sprachsystematischen und semantisch-pragmatischen Dimension von Schreibkompetenz differenziertere Einblicke erlauben, indem Aussagen über Einzelfacetten und deren Zusammenhänge getroffen werden können. Denkbar wäre beispielsweise, die Einflüsse von Aufgabenstellung und Lernunterstützungen auf die beiden Teilkompetenzen zu untersuchen. Möglicherweise lassen sich – je nach Art und Inhalt der Lernunterstützung bzw. Aufgabenstellung – stärkere Zusammenhänge mit einer der Dimensionen nachweisen, was präzisere Implikationen für die Unterrichtspraxis zur Folge haben könnte. So wirkt sich eine schreibfunktionsorientierte, präzise Aufgabenstellung vermutlich vor allem auf die Entfaltung der semantisch-pragmatischen Teilkompetenz aus, während lösungsorientierte Lernunterstützungen vor allem Effekte auf den sprachsystematischen Faktor besitzen sollten.

Eine Analyse der unterrichtlichen Determinanten auf individuelle Schreibleistungen, die auf der Datengrundlage der zweidimensionalen Schreibfähigkeiten erfolgt, könnte zudem sensiblere Rückschlüsse ermöglichen. So lassen sich unter Umständen Effekte bestimmter fachdidaktischer Kriterien (z. B. Schreibstrategien) auf Teildimensionen der Schreibkompetenz dokumentieren, die im Zusammenhang mit der globalen, eindimensionalen Textqualität nicht nachgewiesen werden konnten. Auch aus der schreibdidaktischen Perspektive ist ein stärkerer Fokus auf die zweidimensionale Struktur der Schreibkompetenz zu befürworten. Im Bereich des Texte-Verfassens müssen Schülerinnen und Schüler verschiedene Teilkompetenzen erwerben, die vor allem dann realisiert werden, wenn sie sich mit den Anforderungen auf semantisch-pragmatischer und sprachsystematischer Ebene bewusst auseinandersetzen (vgl. auch Neumann, 2007).

Dementsprechend werden im Projekt NaSch1 weitere Untersuchungen angestrebt, die sich auf die Teilkompetenzen der Schreibfähigkeiten beziehen.

Aus methodischer Perspektive ist zu berücksichtigen, dass die diskutierten Befunde vor dem Hintergrund der vorliegenden Stichprobe und des Forschungsdesigns zu interpretieren sind. So handelt es sich bei der aus PERLE herangezogenen Stichprobe um eine selektive Datenbasis (siehe 1.2). Während für Deutschland der mittlere HISEI in der PISA-Studie 2003 $M = 49.2$ ($SD = 15.9$) beträgt (Prenzel et al., 2005), übersteigt der mittlere HISEI in den Familien der PERLE-Studie diesen Wert mit $M = 62.54$ ($SD = 15.34$) um ein Vielfaches (Lipowsky et al., 2013). Die Ergebnisse zur Textqualität sind daher auf Grundschulen mit einer sozioökonomisch heterogeneren Zusammensetzung der Schülerschaft wahrscheinlich eher eingeschränkt übertragbar. Mit Blick auf gut strukturierte Aufgabenstellungen, qualitative Differenzierungsangebote und individuelle Unterstützungsmaßnahmen empfiehlt es sich jedoch auch bei heterogeneren Klassenzusammensetzungen, anspruchsvolle Schreibaufgaben einzusetzen und dadurch hierarchiehöhere Schreibkompetenzen zu lancieren.

Weitere Einschränkungen resultieren auch aus der relativ kleinen Stichprobengröße und den speziellen Herausforderungen videobasierter Datenquellen, etwa hinsichtlich der Selektivität von Videoaufnahmen (z. B. Brophy, 2009; Clausen, 2002; Praetorius, 2014). So sollen auf Basis einer Doppelstunde im Fach Deutsch möglichst Rückschlüsse über die Alltagspraxis im Schreibunterricht getroffen werden (Lotz, 2016). Auch die Repräsentativität von Videoanalysen wird kritisch betrachtet, da Videoaufnahmen des Unterrichts möglicherweise zu Reaktivitätseffekten bei Lehrkräften und Schülerinnen und Schülern führt und die Unterrichtssituationen somit entsprechend verfälscht werden (z. B. Clausen, 2002; Krammer, 2009; Kunter, 2005). Nachträgliche Befragungen von Lehrkräften belegen jedoch oftmals, dass die Alltagspraxis und die videografierten Stunden nur gering voneinander abweichen (z. B. Helmke, 2014; vgl. auch Lotz, 2016; Reusser & Pauli, 2003;). Auch im Projekt PERLE lässt sich feststellen, dass die teilnehmenden Lehrpersonen ihren Unterricht im Rahmen des Lehrerinterviews zu 84.8 % als alltägliche Unterrichtspraxis bezeichnen³³ (Gabriel, 2014, S. 152).

Berücksichtigt werden sollte zudem, dass auch die Prämisse des Kompetenz-Performanz-Modells (vgl. S. 26f) als kritisch betrachtet werden kann. Die Befunde im Rahmen des Beitrags 4 können nicht zwangsläufig dahingehend gedeutet werden, dass eine präzise, verständliche und schreibfunktionsorientierte Aufgabenstellung die Kompetenz der Kinder beeinflusst. Möglicherweise gelingt es den Schülerinnen und Schülern erst durch eine klare,

³³ Fraglich ist vor dem Hintergrund der sozialen Erwünschtheit und weiteren potenziellen Verzerrungen bei Selbstauskünften von Lehrkräften, ob sie ihren eigenen Unterricht objektiv und angemessen beurteilen können (Gabriel, 2014, S. 152).

strukturierte Aufgabenstellung, ihre Kompetenz (Performanz) auch tatsächlich zu demonstrieren.

Als limitierend für die Ergebnisse in Beitrag 5 gilt die Tatsache, dass die fachliche und fachdidaktische Qualität der Lernunterstützungen nicht berücksichtigt wurde.

In Bezug auf die Wirkungsanalysen zwischen Textqualität und den Prozessvariablen des Schreibunterrichts ist zudem als problematisch zu betrachten, dass die Bilderbuchrezeption in die Untersuchung bislang nicht integriert wurde. Durch die Einbettung der Schreibaufgabe in einen handlungs- und produktionsorientierten Schreibunterricht ist diese Phase jedoch auch mit der Textproduktion verbunden und sollte als potentieller Einflussfaktor auf die Textqualität untersucht werden. So können während der Vorstellung und Besprechung des Bilderbuchs zentrale Aspekte aufgegriffen worden sein, die auch die Aufgabenstellung betreffen. Denkbar ist in diesem Zusammenhang, dass während der Bilderbuchrezeption bereits eine Interaktion zwischen den Kindern stattgefunden hat, etwa in Form von Rollenspielen und Diskussionen, die den Kindern die Perspektivenübernahme beim Schreiben erleichtern kann.

Trotz dieser methodischen und inhaltlichen Einschränkungen lässt sich resümierend festhalten, dass auf Grundlage des Forschungsvorhabens erstmals im deutschsprachigen Raum Schreibfähigkeiten in der Primarstufe auf breiter Basis quantitativ-empirisch analysiert wurden (Beitrag 1) und Aussagen über die Dimensionalität der Schreibkompetenz von Erstklässlerinnen und Erstklässlern getroffen werden können (Beitrag 2). Weiterhin liegen empirische Ergebnisse darüber vor, welche fachdidaktischen und fachübergreifenden Aspekte Lehrkräfte in ihrer Aufgabenstellung heranziehen (Beitrag 3), wie sich ihr Unterstützungsverhalten charakterisieren lässt (Beitrag 5) und welche Konsequenzen aus der Aufgabenstellung und Lernunterstützung für die Qualität von Schreibprodukten resultieren (Beitrag 4 und 5).

3.2. Diskussion der methodischen Vorgehensweise

Ergänzend zu der inhaltlichen Erörterung der zentralen Befunde im vorherigen Kapitel wird nachfolgend das methodologische Vorgehen im Rahmen des Dissertationsprojekts kritisch reflektiert. Neben der Skalierung des kriterialen Textbewertungsinstruments und den Mehrebenenanalysen zu den Effekten des Schreibunterrichts auf Textqualität steht vor allem die Entwicklung und Güte der vier verwendeten Messinstrumente im Zentrum der Diskussion. Auf diese Weise wird auf ein Forschungsdesiderat der Unterrichtsqualitätsforschung reagiert, die nach Ansicht Praetorius (2014) „[...] eher am Anfang steht, was die Konzeption und Entwicklung von Messinstrumenten angeht“ (S. 31). Dies gilt umso mehr für die domänenspezifische Forschung im Kompetenzbereich Schreiben der Primarstufe. So mangelt es im deutschsprachigen Raum selbst an empirischen Studien, die die Prozessqualität des Schreibunterrichts lediglich deskriptiv untersuchen. Dementsprechend liegen bislang auch eher vage Beschreibungen vor, wie sich fachdidaktische Konstrukte operationalisieren und Kriterien wie soziale Interaktionen vor dem Hintergrund des Anfangsunterrichts entwickeln lassen.

In der vorliegenden Arbeit wurden sowohl die Qualität der Unterrichtsprozesse (Aufgabenstellung und Lernunterstützungen im Schreibunterricht) als auch die Qualität der Unterrichtsprodukte (Texte der Schülerinnen und Schüler) anhand von Einschätzungen geschulter Beobachterinnen und Beobachter erfasst. Diese Erhebungsmethode gilt als „Königsweg“ (Helmke, 2014, S. 292; vgl. auch Petko, Waldis, Pauli & Reusser, 2003) und beruht auf der Annahme, dass externe Beobachterinnen und Beobachter aufgrund eines speziellen Trainings eine angemessene Beobachtungs- und Urteilskompetenz entfalten (Praetorius, 2014). Gerade im Vergleich zu Schüler- und Lehrerbefragungen sollen Beobachtungen und Videoanalysen im Hinblick auf die Untersuchung der Prozessqualität höhere Effekte zur Folge haben (Seidel & Shavelson, 2007).

Helmke (2014) argumentiert in diesem Kontext, dass sich die hohe Dynamik und Komplexität des Unterrichtsgeschehens über Beobachtungen am adäquatesten erfassen lässt. So werden auch deren Defizite, etwa die hohen Kosten und Selektivität (z. B. Clausen, 2002), im Vergleich zu den Nachteilen der Lehrer- und Schülerbeobachtungen als weniger gravierend eingestuft (Helmke, 2014; vgl. auch Praetorius, 2014).

Entwicklung und Erprobung der Messinstrumente

Die Entwicklung der Messinstrumente zur Erfassung der Schreibfähigkeiten und der Prozessqualität des Schreibunterrichts beruht hauptsächlich auf methodischen Ansätzen der videobasierten Unterrichtsforschung, wie sie beispielsweise auch in der IPN Videostudie „Lehr-Lernprozesse im Physikunterricht“ (Seidel, Prenzel, Duit & Lehrke, 2003), der

schweizerisch-deutschen „Pythagoras“-Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis" (Klieme, Pauli & Reusser, 2006) sowie der PERLE-Studie (Lotz, Lipowsky & Faust, 2013c) zum Tragen gekommen sind³⁴.

So wurde sowohl für die beiden Messinstrumente der Textqualität (kriteriales und holistisches Rating) als auch für die Instrumente der Aufgabenstellung und individuellen Lernunterstützung ein deduktiv-induktives Verfahren in Anlehnung an Hugener (2006), Jacobs und Kollegen (2003), Lotz, Berner und Gabriel (2013a) sowie Seidel (2003) kombiniert. Dementsprechend fußt die Entwicklung der jeweiligen Items einerseits auf theoriebasierten fachdidaktischen Grundlagen und bestehenden Beobachtungssystemen, andererseits auch auf zufällig ausgewählten Teilstichproben der zu analysierenden Texte bzw. Unterrichtsvideos. Letzteres ist besonders im Kontext der zum Teil fehlenden, adaptierbaren Instrumente essenziell (vgl. auch Lotz et al., 2013a): Da dezidiert für das Fach Deutsch in der Primarstufe keine angemessenen Beobachtungssysteme vorliegen, wurden teilweise Instrumente (bzw. vielmehr einzelne Items) für die Entwicklung der eigenen Beobachtungssysteme herangezogen, die für höhere Jahrgangsstufen oder andere Fächer konzipiert wurden. Durch den Einbezug der Teilstichproben des eigenen Materials lassen sich zentrale grundschul- und deutschdidaktische Facetten in die Entwicklung integrieren, wodurch eine differenzierte, angemessene Erfassung der Konstrukte gewährleistet wird. Die präzise, vielschichtige Analyse der Texte und des Unterrichts kann auch dadurch gesichert werden, indem in den meisten der vorliegenden Messinstrumente eine Kombination aus niedrig, mittel und hoch inferenten Items³⁵ realisiert wurde.

Nach Abschluss dieses Entwicklungsschrittes, der von zahlreichen Klassifikations-, Selektions- und Abstraktionsprozessen (z. B. Bortz & Döring, 2006; Hugener, 2006; Seidel, 2003; vgl. auch Lotz et al., 2013a) gekennzeichnet ist, liegen die Messinstrumente in Form sogenannter Beobachtungsmanuale vor, die neben den einzelnen Kriterien auch Rating- und Kodierregeln sowie umfangreiche Ankerbeispiele für die Beobachterinnen und Beobachter enthalten. Dementsprechend bilden die Manuale die Grundlage für die Einschätzung der Texte und des Unterrichts und wurden in der Schulung der Beobachterinnen und Beobachter eingesetzt. Die mehrtägigen Trainings zielten unter Einsatz einer hohen Anzahl von Übungstexten bzw. Videos³⁶ (z. B. 80 Texte, 5 Videos) darauf ab, die Beobachterinnen und

³⁴ Bei der Bewertung der Texte durch externe Beobachterinnen und Beobachter wurde das methodische Vorgehen ebenfalls an die videogestützte Unterrichtsforschung und deren Begrifflichkeiten angepasst.

³⁵ Das Messinstrument zur Erfassung der individuellen Lernunterstützung enthält keine hoch inferenten Items. Zudem besteht das holistische Rating aus einer einzelnen intervallskalierten Dimension, da das Ziel darin besteht, Schreibfähigkeiten als globale Kompetenz zu erfassen.

³⁶ Die Übungstexte bzw. Videos, die für die Schulungen verwendet und daher von der Übereinstimmungsberechnung ausgeschlossen wurden, sind nach der Auswertung von den kodierenden/ratenden Personen noch einmal unter Kontrolle der Übereinstimmung untereinander bzw. mit der Manualentwicklerin eingeschätzt worden. Dadurch können sie der Stichprobe zugerechnet werden und gehen in weitere Analysen ein.

Beobachter mit den Instrumenten vertraut zu machen und eine Beurteilungsexpertise zu entwickeln.

Im Anschluss an die Schulung erfolgte die jeweilige Einschätzung der Textqualität bzw. des Schreibunterrichts auf Basis der entwickelten Instrumente. Zur Vermeidung von Konsistenztendenzen wurde jedes Beobachtungssystem getrennt voneinander und anhand unterschiedlicher Beobachterinnen und Beobachter erfasst. Zudem wurden im Sinne einer effektiven Kontrolle des Auswertungsprozesses die Einschätzungen der Beobachterinnen und Beobachter fortlaufend durch die Berechnung von Beobachterübereinstimmungen bzw. Reliabilitäten zu mehreren, vorab festgelegten Zeitpunkten kontrolliert. Dieses Prozedere schafft die Voraussetzung, um Abweichungen und Schwierigkeiten bei der Text- bzw. Unterrichtsbeurteilung direkt zu identifizieren und Defizite gegebenenfalls in einer Nachschulung zu beheben.

Dokumentation der Güte der Messinstrumente

Im Sinne der Gütekriterien besteht das vordergründige Ziel der empirischen Analyse darin, die individuellen Schreibleistungen der Schülerinnen und Schüler und die Prozessvariablen des Schreibunterrichts auf Grundlage der Messinstrumente objektiv und reliabel zu erfassen. Allerdings birgt diese Art der Datenerhebung, die auf Einschätzungen externer Beobachterinnen und Beobachter basiert, grundsätzlich die Gefahr von Verzerrungen. So lassen sich verschiedene Arten von Rater-Bias, also der Anteil an den Einschätzungen, der durch die Raterinnen und Rater verursacht wird und nicht von dem zu beurteilenden Kriterium ausgeht (z. B. Schaefer, 2008), unterscheiden, etwa die Tendenz zur Strenge bzw. Milde, die Tendenz zur Mitte oder die Tendenz zur konsistenten Bewertung (vgl. auch Böhme et al., 2009; Praetorius, 2014).

Um möglichst hohe Übereinstimmungen in der Einschätzung zwischen den Beurteilerinnen und Beurteilern zu erzielen, wurden daher zum einen Instrumente entwickelt, deren präzise und eindeutige Kodier- und Ratingregeln individuelle Verzerrungen aufgrund subjektiver Eindrücke minimieren können. Zum anderen wurde sichergestellt, dass die Beurteilerinnen und Beurteiler während der intensiven Schulungen mit dem Konstrukt und der Anwendung des Beobachtungsinstruments vertraut gemacht werden (vgl. auch Böhme et al., 2009).

Für alle im Rahmen des Dissertationsprojekts eingesetzten Messinstrumente liegen letztendlich Kennwerte vor, die die Güte der Einschätzungen dokumentieren³⁷. Dieses Prozedere orientiert sich an videobasierten Untersuchungen wie die bereits erwähnten IPN-

³⁷ Das Dissertationsprojekt adaptiert die in der PERLE-Studie festgelegten, relativ strengen Mindestkennwerte, die bei der prozentualen Übereinstimmung ($P\hat{U}$) $\geq 85.00\%$, bei Cohens Kappa (κ) $\geq .70$ und beim relativen Generalisierbarkeitskoeffizienten (g) $\geq .70$ betragen (Lotz et al., 2013a).

Videostudie (Seidel, 2003), die „Pythagoras“-Videostudie (Hugener, 2006) und die PERLE-Studie (Lotz et al., 2013c).

Wie sich den einzelnen Publikationen der Forschungsarbeit entnehmen lässt, wurde das zentrale Ziel, die Schreibleistungen der Schülerinnen und Schüler und die unterrichtlichen Prozessvariablen reliabel und objektiv zu erfassen, für alle Messinstrumente erreicht³⁸.

Während die Objektivität dadurch gesichert wurde, dass das Vorgehen unter kontrollierten Bedingungen und einer transparenten Darstellung stattgefunden hat (vgl. Hugener, 2006), wurde die Reliabilität über statistische Maßzahlen überprüft. Für jedes Item wurde in Abhängigkeit seiner Inferenz bzw. seines Skalenniveaus kontrolliert, ob die Einschätzungen der Rater die erforderlichen, vorab festgelegten Mindestkennwerte hinsichtlich der Übereinstimmung erfüllen. Bei nominalskalierten Kategoriensystemen (meist niedrig bis mittel inferente Items) wurden dementsprechend die prozentuale Übereinstimmung sowie Cohens Kappa ermittelt, bei Ratingskalen (hoch inferente Items) liegen Kennwerte der Interraterreliabilitäten auf Basis des absoluten und relativen Generalisierbarkeitskoeffizienten vor. Die Entscheidung über die Stichprobengröße für die Überprüfung der Übereinstimmungen wurde ebenfalls auf Grundlage der Inferenz bzw. des Skalenniveaus des jeweiligen Items getroffen. So wurden im Rahmen nominalskalierter Kategoriensystemen lediglich für einen Teil der Stichprobe Kennwerte berechnet (jeweils 1/3 der Texte und der Videos), während bei den intervallskalierten Ratingsystemen alle Einschätzungen der Rater über die gesamte Stichprobe bestimmt wurden.

Neben der Beobachterübereinstimmung bzw. Interraterreliabilität wurde auch die Intraraterreliabilität als Maß herangezogen, um die zeitliche Stabilität der Beurteilungen zu erfassen. Die Befunde deuten an, dass die Einschätzungen der Kodierer/Rater auch über verschiedene Messzeitpunkte hinweg als hinreichend konsistent zu bewerten sind.

Vor dem Hintergrund der Einwände, Interraterreliabilitäten als alleinigen Indikator für die Güte der Einschätzungen zu verwenden (Shavelson & Dempsey-Atwood, 1976; vgl. auch Praetorius, 2014), wurden in der Arbeit auch Varianzanteile im Rahmen der Generalisierbarkeitstheorie (Cronbach, Gleser, Nanda & Rajaratnam, 1972) angegeben.

Mittels dieser Theorie ist es möglich, die Varianz in Ratings in unterschiedliche Anteile zu zerlegen und so zu bestimmen, welcher Varianzanteil in Ratings auf den zu messenden Gegenstand, auf die Rater, auf die Items, sowie ggf. weitere Merkmale der Messung zurückzuführen ist (...). (Praetorius, 2014., S. 41)

³⁸ Bei einzelnen niedrig und mittel inferenten Kriterien resultieren bei Cohens Kappa Ausreißer, die den Mindestkennwert unterschritten haben. Aufgrund der jeweils hohen prozentualen Übereinstimmung und der geringen Varianz in der Stichprobe als Ursache für den niedrigen Kappa-Wert wird schlussendlich von einer zufriedenstellenden Übereinstimmung ausgegangen (präzise Beschreibung z. B. bei Pohlmann-Rother & Kürzinger, i.V.).

Die Generalisierbarkeitstheorie nimmt unter anderem an, dass Messwerte durch mehrere Fehlerquellen (simultan sowie interagierend) verzerrt werden können (Lotz et al., 2013a). In Anlehnung an Lotz und Kolleginnen (2013a, S. 97) bilden in der vorliegenden Forschungsarbeit (1) die Unterrichtsvideos bzw. Kindertexte, (2) die Rater und (3) die Interaktion zwischen Video (Text) und Rater und die unsystematische Fehlervarianz die Varianzquellen einer Messung. Ein großer Anteil der Varianz sollte dabei auf reale Unterschiede zwischen den Videos bzw. Texten zurückgehen, weniger bzw. so gering wie möglich auf die systemische Fehlervarianz, die auf Unterschiede in den Urteilen basiert (Clausen et al., 2003; vgl. auch Lotz et al., 2013a), was sich für die intervallskalierten Ratingdimensionen der Messinstrumente belegen lässt. In der nachfolgenden Tabelle (Tabelle 2) sind exemplarisch die Varianzkomponenten für die Bewertung der Texte durch das holistische Rating sowie die intervallskalierten Ratingdimensionen des kriterialen Instruments angegeben³⁹

Tab 2: Varianzkomponenten und relativer Generalisierbarkeitskoeffizient für das holistische Rating und die hoch inferenten kriterialen Items ($N = 500$)

Ratingdimension	g_{relativ}	Varianzkomponenten (VK)		
		Brief	Rater	BxR+e
Holistisches Rating	.91	84 %	0 %	16 %
Textidee und -entfaltung	.99	98 %	0 %	2 %
Implizitheit	.99	99 %	0 %	1 %
Kohäsionsmittel	.99	99 %	0 %	1 %
Angemessenheit der Sprachmittel ⁴⁰	.98	95 %	0 %	4 %
Wortschatzverwendung	.97	96 %	0 %	4 %

Anmerkungen: g_{relativ} = relativer Generalisierbarkeitskoeffizient; BxR+e = Varianzkomponente Interaktion Brief und Rater sowie unsystematische Variation.

Resümierend kann der Nachweis der Reliabilität und Objektivität vor allem hinsichtlich der beiden Messinstrumente der Textqualität als zufriedenstellend betrachtet werden, insofern Studien zur Erfassung von Schreibkompetenz teilweise unzureichende Übereinstimmungen zwischen den Beurteilerinnen und Beurteilern berichten (z. B. Blatt et al., 2009; Bremerich-Voss & Possmayer, 2011). Allerdings ist im Rahmen der vorliegenden Stichprobe kritisch einzuräumen, dass die vergleichsweise sehr hohen Beurteilerübereinstimmungen auch der geringen Textlänge geschuldet sein können. Schreibanfängerinnen und Schreibanfänger

³⁹ Die Berechnungen wurden mit dem GT-Programm für Generalisierbarkeitsstudien (Ysewijn, 1996) durchgeführt.

⁴⁰ Bei der Berechnung der Varianzkomponenten mit dem GT-Programm für Generalisierbarkeitsstudien wird automatisch gerundet, weswegen Rundungsdifferenzen von bis zu 1 % auftreten können (vgl. auch Clausen et al., 2003).

verfassen kürzere, sprachlich und inhaltlich geringer entfaltete Texte, was vermutlich zu einer einheitlicheren Einschätzung führt. So weisen Befunde darauf hin, dass längere Texte weniger konsistent bewertet werden (vgl. auch Baurmann, 2006; Kötter & Grau, 1965). Analysen zu ausgewählten Teilstichproben (Texte mit $N > 17$ Wörter vs. Texte mit $N < 17$ Wörter) innerhalb des Forschungsvorhabens belegen zumindest im Ansatz, dass die Textlänge keinen messbaren Einfluss auf die Beurteilerübereinstimmung besitzt⁴¹.

Indessen gestaltet sich ein in statistische Kennzahlen überführter Nachweis der Validität für alle Instrumente des Dissertationsprojekts als problematisch. So gilt etwa die Ähnlichkeit der Urteile lediglich als Minimalvoraussetzung für Validität (Wirtz & Caspar, 2002). Im Projekt PERLE erfolgte die Validierung der eingesetzten Instrumente beispielsweise im Sinne einer interpersonellen Konsensbildung (Bortz & Döring, 2006) vorwiegend im Entwicklungsprozess der Kriterien durch das Bemühen um solide theoretische Grundlagen und wechselseitigen Austausch (Lotz et al., 2013a). Die Sicherung der Validität kann sich dabei durch ein gemeinsames theoretisches Verständnis über die einzelnen Kriterien fortsetzen, das bei der Schulung der beobachtenden Personen entwickelt werden sollte und sich anhand von Übereinstimmungen der Beurteilerinnen und Beurteiler mit der Entwicklerin bzw. dem Entwickler der Beobachtungsverfahren überprüfen lässt (vgl. auch Früh, 2007). In der vorliegenden Arbeit wurde eine Validierung der Instrumente angestrebt, indem das Prozedere des PERLE-Projekts adaptiert wurde. Die interpersonelle Konsensbildung wurde dadurch kontrolliert, dass nicht nur Übereinstimmungen zwischen den Beurteilerinnen und Beurteilern berechnet wurden, sondern für einen Teil der Stichprobe auch zwischen den einzelnen Beurteilerinnen und Beurteilern und der Manualentwicklerin. Zusätzlich wurden – sofern möglich – in einigen Instrumenten dieselben Aspekte sowohl niedrig inferent als auch hoch inferent erfasst, so dass sich die Kriterien gegenseitig validieren können.

Lotz, Gabriel und Lipowsky (2013b) führen in diesem Kontext folgendes an:

Die Kombination beider Verfahren erlaubt zum einen eine gegenseitige Validierung der Ergebnisse und kann zum anderen dazu beitragen, ein umfassendes Bild des Unterrichts zu erhalten, indem sowohl übergeordnete Qualitätsmerkmale des Unterrichts bewertet als auch genaue Analysen einzelner Interaktionen zwischen den verschiedenen Akteuren im unterrichtlichen Geschehen vorgenommen werden. (S. 377)

Aus dieser Perspektive lässt sich die hohe Übereinstimmung zwischen dem holistischen Rating und dem kriterialen Textbewertungsinstrument (vgl. Beitrag 2) vermutlich als Indikator einer gelungenen Validierung der Instrumente betrachten.

⁴¹ Da es sich jedoch auch bei der Gegenüberstellung der Teilstichproben nach wie vor um Schreibprodukte mit einer geringen textuellen Entfaltung handelt, sind weitere Untersuchungen mit unterschiedlich langen Textkorpora nötig, um den Einfluss der Textlänge differenzierter erfassen zu können.

Dennoch bleibt zu betonen, dass im Gegensatz zur Reliabilität kein statistischer Nachweis in Form von Kennzahlen für die Validität ermittelt wurde.

Skalierung der kriterial ermittelten Schreibleistungen

Als Ausgangspunkt für weitere Analysen wurden die Schreibleistungs-Daten der kriterialen Einschätzung einer Skalierung auf Basis der Item-Response-Theorie (IRT) (z. B. Rost, 2004) unterzogen. Diese Skalierungsmethode zeichnet sich im Gegensatz zu Skalierungen auf Grundlage der klassischen Testtheorie (KTT) dadurch aus, dass sie eine Kontrolle der Konstruktvalidität mehrerer Items, die als Indikatoren des Konstrukts betrachtet werden, zulässt (Karst, Kastens, Mösko, Lipowsky, 2013).

Grundsätzlich nehmen IRT-Verfahren an, dass ein Zusammenhang zwischen der latenten Fähigkeit einer Person (Personenparameter) und der Wahrscheinlichkeit besteht, eine Aufgabe zu lösen, wobei die Schwierigkeit der jeweiligen Aufgabe berücksichtigt wird. So erlaubt die IRT-Modellierung gleichermaßen Rückschlüsse über die Itemschwierigkeit und die Personenfähigkeiten, indem zum einen für jedes Item festgestellt wird, welche Anzahl an Kindern das Kriterium richtig bzw. falsch gelöst hat. Zum anderen wird für jeden Lernenden bestimmt, welche Anzahl an Items von ihm korrekt gelöst wurden (vgl. auch Neumann, 2007).

Aufgrund von dichotomen⁴² und mehrstufigen Merkmalsausprägungen innerhalb des kriterialen Instruments erfolgte die Skalierung in Beitrag 2 auf Basis von Partial-Credit-Modellen (Einparameter-Logistisches-Modell), die die Lösungswahrscheinlichkeit jedes Items in Abhängigkeit der latenten Personenfähigkeit modellieren⁴³.

Um die Dimensionalität der Schreibkompetenz zu überprüfen, wurden ein- und zweidimensionale Partial-Credit-Modelle im Rahmen der Statistiksoftware ConQuest2 (Wu et al., 2007) gegenübergestellt, wobei der Nullpunkt der Skala als mittlere Personenfähigkeit festgelegt wurde⁴⁴.

Die Entscheidung für das zweidimensionale Modell von Schreibkompetenz (semantisch-pragmatischer und sprachsystematischer Faktor) beruht dabei auf verschiedenen Fit-Indizes (WMNSQ⁴⁵ sowie BIC-, AIC- und CAIC-Informationskriterien) und der vergleichsweise geringen latenten Korrelation ($r = .49$) zwischen den beiden Subdimensionen des zweidimensionalen Modells. So sprechen etwa die niedrigeren Informationskriterien des BIC,

⁴² Ausschließlich dichotom erfasste Merkmalsausprägungen lassen sich anhand von Rasch-Modellen (Rasch, 1960) skalieren.

⁴³ Die Beschreibung der Skalierung entstammt aus Beitrag 2 (Pohlmann-Rother et al., 2016). Die Skalierung auf Basis von Partial-Credit-Modellen wurde dabei von dem Co-Autor Schoreit berechnet (siehe auch Tabelle 1, S. 14). In Vorarbeiten zu Beitrag 2 wurde zunächst eine Gegenüberstellung der beiden Textbewertungsinstrumente anhand von Rasch-Modellen von Kürzinger berechnet.

⁴⁴ Die Skalenbildung erfolgte anhand einer Schätzung der WLE-Personenparameter (Weighted Likelihood Estimator).

⁴⁵ Der angestrebte Bereich des WMNSQ-Index umfasst Werte von 0.75- 1.33 (Adams & Khoo, 1996).

AIC und CAIC gegen das eindimensionale Modell, bei dem auch die Infit-Werte bei den Items „Leserlichkeit“ und „Orthografie“ mit WMNSQ-Werten von 1.43 und 1.34 oberhalb des angestrebten Bereichs liegen.

Auf Basis der Skalierung wurde zudem ein Vergleich zwischen den beiden Textbewertungsinstrumenten ermöglicht, der darüber Aufschluss geben soll, welches Instrument adäquater für eine reliable Erfassung der Schreibfähigkeiten von Schülerinnen und Schülern geeignet ist. In diesem Kontext wurde auch dem Einfluss der Textlänge Rechnung getragen. Den Befunden zufolge kristallisiert sich keine deutliche Überlegenheit eines der Instrumente heraus. So korreliert das holistische Rating sowohl mit der sprachsystematischen ($r = .47$) Subdimension als auch mit der semantisch-pragmatischen Schreibkompetenz ($r = .78$) hinreichend. Die Textlänge erweist sich nicht als spezifische Fehlerquelle des holistischen Ratings.

Aus messmethodischer Perspektive gilt die Bewertung des Gesamteindrucks im Vergleich zur kriterialen Einschätzung allgemein aufgrund der gemittelten Urteile über beide Raterinnen bzw. Rater⁴⁶ als reliabler und valider (Bortz & Döring, 2006). Allerdings lassen sich mit ihr inhaltsbezogene Dimensionen weniger adäquat erfassen (Beck, 1979; Hofen, 1980, vgl. auch Neumann, 2007). Aufgrund der Tatsache, dass bislang keine Überlegenheit eines der beiden Verfahren festgestellt werden konnte (Neumann, 2007), ergänzen sich die beiden Instrumente und werden in vielen Studien kombiniert (z. B. Böhme et al., 2009; Kruse et al., 2012).

In der vorliegenden Forschungsarbeit ließ sich die Güte für beide Instrumente gleichermaßen nachweisen. Allerdings ist vor dem Hintergrund der theoretisch modellierten wie empirisch belegten Mehrdimensionalität von Schreibkompetenz zu berücksichtigen, dass das holistische Rating lediglich eine Globaldimension abzubilden vermag. Diese Charakteristik stellt jedoch sowohl aus forschungsbezogener Sicht als auch aus dem pädagogischen Blickwinkel eine beträchtliche Einschränkung dar (vgl. auch S. 31). So lassen sich mit Teildimensionen der Schreibkompetenz differenziertere Analysen realisieren, etwa, wenn Effekte bestimmter unterrichtlicher Prozessvariablen auf einzelne Fähigkeitsbereiche nachgewiesen werden sollen. Damit geht auch eine Limitation der vorliegenden Forschungsarbeit einher, bei der lediglich das holistische Rating als abhängige Variable modelliert wurde.

Mehrebenenanalytische Untersuchung der Effekte des Schreibunterrichts auf die Textqualität

Das zentrale Ziel des vorliegenden Dissertationsprojekts, unterrichtliche Prädiktorvariablen zur Erklärung individueller Schreibleistungen heranzuziehen, wurde im Rahmen eines mehrebenenanalytischen Designs mit der Software HLM 7.0 (Raudenbush et al., 2008)

⁴⁶ In der vorliegenden Arbeit liegen für intervallskalierte und qualitätsbezogene Items ebenfalls für die komplette Stichprobe gemittelte Einschätzungen von beiden Ratern vor.

untersucht. Mehrebenenanalysen zeichnen sich durch die Berücksichtigung der hierarchischen Struktur von Daten aus, die dadurch vorliegt, dass die Schülerinnen und Schüler in Klassen geschachtelt sind. Individuelle Daten (Schülerinnen und Schüler) werden demnach auf der unteren Mikroebene 1 modelliert, Gruppen (Klassen) auf der übergeordneten Ebene 2 dargestellt. Konventionelle Regressionsanalysen können dagegen weder die hierarchische Ordnung noch die Interaktionseffekte zwischen Individual- und Klassenmerkmalen abbilden.

In der Arbeit wurden jeweils separat für die Untersuchung der Effekte der Aufgabenstellung und der Lernunterstützung Modelle berechnet, in der die verschiedenen Prädiktoren schrittweise hinzugeführt wurden⁴⁷. In allen Analysen wurde als abhängige Variable auf Ebene 1 das holistische Rating modelliert, während als unabhängige Variablen auf Ebene 2 die verschiedenen unterrichtlichen Prozessvariablen eingeflossen sind. Zusätzlich wurde auf Ebene 2 als Indikator für die mittlere Leistungsstärke der Klasse im Lesen der gemittelte und aggregierte Wert der Buchstabenkenntnis einbezogen, da vorstellbar ist, dass auch die Leistungsstärke der Klasse im Lesen die Qualität der individuellen Texte beeinflusst. Als Kontrollvariablen auf Ebene 1 wurden die relevanten individuellen Einflussfaktoren Buchstabenkenntnis sowie das Geschlecht der Schülerinnen und Schüler herangezogen. Bis auf die dichotomen Merkmale wurden alle Items vor der Mehrebenenanalyse z-transformiert, um die Regressionskoeffizienten miteinander vergleichen zu können (vgl. Hartig & Rakoczy, 2010). Zentriert wurde dabei um den Gesamtmittelwert.

Die Entscheidung für einen mehrebenenanalytischen Ansatz wird auch durch die anhand des Nullmodells (One-Way-ANOVA) in HLM geschätzte Intraklassenkorrelation *ICC* gestützt, die den Anteil der Klassenzugehörigkeit an der Gesamtvarianz angibt. Diese verweist mit einem Wert von 22 %⁴⁸ der Schreibleistung auf markante Unterschiede zwischen den Klassen, was bedeutet, dass die Klassenzugehörigkeit einen Effekt auf die individuelle Schreibleistung der Schülerinnen und Schüler in der Primarstufe besitzt.

Nach der Berechnung des Nullmodells wurde in einem zweiten Schritt das Individualebenenmodell (Random-Coefficient-Model) in HLM spezifiziert, das die Varianz auf der Individualebene und damit die Effekte der schriftsprachlichen Vorläuferfähigkeiten (Buchstaben lesen) und des Geschlechts auf die Textqualität erklärt. In einem letzten Schritt wurden im Rahmen des Intercepts-and-Slopes-as-Outcomes-Modell die verschiedenen unterrichtlichen Prädiktoren einzeln auf Ebene 2 hinzugefügt.

Wie die Befunde aus den Beiträgen 4 und 5 zeigen, lassen sich individuelle Schreibleistungen in Klasse 1 durch bestimmte Prozessvariablen des Schreibunterrichts (Aufgabenstellung, Lernunterstützung) erklären.

⁴⁷ Das Modell zur Analyse der Lernunterstützung enthält auch einen Prädiktor der Aufgabenstellung.

⁴⁸ Im Nullmodell zur Analyse der Lernunterstützung wurde eine *ICC* von 22.8 %, im Nullmodell zur Analyse der Aufgabenstellung ein Wert von 22.2 % ermittelt.

Kritisch ist in diesem Zusammenhang anzufügen, dass die Erklärungskraft von Werten im Bereich $\beta = .14$ $-.28$ als eher gering bis mittel ausgeprägt zu bewerten ist. Dieses Ergebnis steht jedoch in Einklang zu wiederholt dokumentierten geringeren Effekten des Unterrichts auf Leistungsvariablen von Schülerinnen und Schülern (z. B. Klieme, Pauli & Reusser, 2009; Lipowsky et al., 2009; Seidel & Shavelson, 2007), was beispielsweise durch die Bedeutung individueller Charakteristika wie die jeweiligen Lernvoraussetzungen für den Schulerfolg (Seidel & Shavelson, 2007) verursacht wird.

Bedenklich erscheint auch im Hinblick auf eine solide Parameter- und Standardfehler-schätzung die geringe Stichproben- bzw. Klassengröße in der vorliegenden Forschungsarbeit. So sollte die erforderliche Anzahl der Untersuchungseinheiten auf Ebene 2 $N = 50$ (Hox, 1998; Maas & Hox, 2004) betragen, was in der Untersuchung mit $N = 42$ unterschritten wird. Daher besteht die Gefahr, dass Zufallseffekte auf Ebene 2 unterschätzt wurden.

Insgesamt lässt sich festhalten, dass aufgrund der hierarchischen Datenstruktur und der hohen ICC ein mehrebenenanalytisches Design trotz der geringeren Stichprobengröße gerechtfertigt ist. Dennoch sind weitere Analysen mit größeren Untersuchungseinheiten auf Ebene 2 wünschenswert. Zudem sollten zukünftige Analysen auch die kriterial erfasste Schreibkompetenz als abhängige Variable modellieren, um differenziertere Ergebnisse zu erhalten.

3.3. Implikationen für die Schreibdidaktik und Schreib- bzw. Unterrichtsforschung

„Schreibenkönnen ist eine zentrale Schlüsselkompetenz in Schule, Beruf und Freizeit – Schreibenlernen bleibt ein Praxisproblem, das noch viele Fragen aufwirft“ (Fix, 2000, S. 52).

Ausgehend von den zentralen Befunden des Dissertationsprojekts, die in den vorherigen Kapiteln vorgestellt und diskutiert wurden, lassen sich sowohl auf inhaltlicher wie auch auf methodologischer Ebene Implikationen für die Schreibdidaktik und Schreibforschung ableiten, die in der Diskussion der zentralen Befunde zum Teil bereits angeklungen sind. Aus inhaltlicher, didaktischer Perspektive resultiert vor allem das Plädoyer, anspruchsvolle und herausfordernde Schreibsettings bereits im Anfangsunterricht einzusetzen. Merz-Grötsch (2005) kritisiert beispielsweise die zum Teil vorherrschende Praxis im schriftsprachlichen Anfangsunterricht, die lediglich auf den Erwerb der Orthografie und Grafomotorik setzt:

Es stellt sich die Frage, ob und wie sich Schreibkompetenzen überhaupt ausbilden und weiterentwickeln können in einem Schreibunterricht, der über den Erwerb grafomotorischer und normorientierter Teilkompetenzen kaum hinausgeht und davon ausgeht, dass weiterführende kommunikative Funktionen den älteren

Schreibern vorbehalten bleiben müssen, in der Annahme kommunikatives Schreiben überfordere die jungen Schreibanfänger. (S. 120)

Die Ergebnisse des Promotionsprojekts zur Textqualität bringen zum Ausdruck, dass bereits in Klasse 1 anspruchsvolle Schreibaufgaben bewältigt und komplexere Schreibfähigkeiten realisiert werden können. Die konzeptionellen, sprachlichen Anforderungen an die Textproduktion werden nicht automatisch von motorischen und orthografiebezogenen Beanspruchungen überlagert, sondern lassen sich als Teilkompetenzen von Schreibkompetenz konstruieren. Dieses Potenzial sollte von Lehrkräften im Anfangsunterricht stärker wahrgenommen werden und durch zielführende, klare Aufgabenstellungen sowie durch individuelle Unterstützung der Kinder während der Schreibhandlung gefördert werden. In den Schülerarbeitsphasen nehmen vor allem evaluative Unterstützungsformen eine zentrale Rolle ein. Auf Grundlage dieser Hilfestellungen erhalten die Lernenden Rückmeldungen zu ihrem Arbeitsstand und zum weiteren Vorgehen, indem ihnen die Ziele ihres Arbeitsprozesses vergegenwärtigt werden, was auf Seiten der Lehrpersonen eine hohe diagnostische Kompetenz voraussetzt. Künftige Untersuchungen sollten im Hinblick auf individuelle Lernunterstützungen auch die Passung zwischen dem kognitiven Anspruchsniveau der Hilfestellung und dem Vorwissen der Schülerinnen und Schüler fokussieren.

Befunde der Dissertation liefern auch Ansatzpunkte dafür, dass sich weitere Studien mit dem fachdidaktischen Wissen von Deutschlehrern und ihrem Einfluss auf die Gestaltung des Schreibunterrichts auseinandersetzen sollten. Bei der Erfassung der Aufgabenstellung hat sich gezeigt, dass fachdidaktische Kriterien tendenziell weniger berücksichtigt werden. Ein Grund für die Vernachlässigung dieser Kriterien könnte ein Defizit an fachdidaktischem Wissen bzw. dem fehlenden Bewusstsein über seine Relevanz sein. Ergebnissen der COACTIV-Studie zufolge variiert das fachdidaktische Wissen unter Lehrkräften, was sich auch im Praxishandeln niederschlägt (Kunter et al., 2011). Mit Blick auf die Lehrerbildung und -fortbildung wird deutlich, dass Lehrkräfte in höherem Maße für eine fachdidaktische Gestaltung des Unterrichts sensibilisiert und ihre professionelle Kompetenz dahingehend ausgebildet werden sollte. Ein höheres fachdidaktisches Wissen könnte ein ausgewogenes Zusammenspiel zwischen Schreibdidaktik und fachübergreifenden Merkmalen der Unterrichtsqualität im Schreibunterricht fördern. Ebenso lässt sich annehmen, dass eine fundierte fachdidaktische Expertise zu einer Kompetenzsteigerung der Schülerinnen und Schüler führen kann. Winkler (2005) konnte etwa nachweisen, dass das fachdidaktische Wissen Effekte auf den Lernerfolg von Schülerinnen und Schülern im Deutschunterricht besitzt.

Stärker methodisch verankerte Implikationen lassen sich aus der Entwicklung der Messinstrumente für die Text- und Unterrichtsqualität in der Primarstufe gewinnen. Die

Konzeption fußt jeweils auf einer theoriebezogenen, fachdidaktischen Herleitung der Konstrukte und Kriterien, die entsprechend empirisch operationalisiert wurden.

Jedes der Instrumente integriert spezifische, aufgabenbezogene Kriterien, die vor dem Hintergrund des didaktischen Settings modelliert wurden, und aufgabenübergreifende Aspekte, die sich auf verschiedene Schreibaufgaben bzw. schriftsprachbezogene Unterrichtsentwürfe übertragen lassen. Damit liefert die Arbeit entscheidende Impulse für künftige Untersuchungen, die sich mit Textqualität oder der Praxis im Deutschunterricht auseinandersetzen und einen ebenfalls quantitativ forschungsmethodischen Zugang wählen.

Zusammenfassend lässt sich festhalten, dass das interdisziplinäre Forschungsvorhaben mit seiner Zielsetzung und seinen Befunden einen Beitrag dazu leisten kann, die kontroversen Perspektiven von Schreibdidaktik und empirischer Bildungsforschung zu verbinden, indem es Forschungsdesiderate der Didaktik adressiert und bei der empirischen Operationalisierung der Konstrukte fachdidaktische Expertisen berücksichtigt (vgl. auch Klieme & Rakoczy, 2008). Eine Annäherung der beiden Disziplinen, deren Diskrepanz vor allem auf dem unterschiedlichen wissenschaftlichen Selbstverständnis sowie dem unterschiedlichen Zugang zu Forschungsmethoden beruht, hat bisher zu wenig bzw. zu einseitig stattgefunden. In diesem Kontext könnten Impulse für weitere Forschungsvorhaben ausgehen, die insbesondere deutschdidaktische Fragestellungen mit quantitativ-empirischen Zugängen verbinden.

3.4. Ausblick auf zukünftige Forschung zu individuellen Schreibfähigkeiten im Grundschulalter

„Die empirische deutschdidaktische Unterrichtsforschung steht noch am Anfang, blickt jedoch auf ein weites und viel versprechendes Arbeitsfeld“ (Kleinbub, 2016, S. 34).

Im Rahmen des Dissertationsprojekts wurden ausgewählte Gelenkstellen des Schreibunterrichts in den Blick genommen, die in künftigen Analysen innerhalb des Projekts NaSch1 vertieft werden sollen, um die gewonnenen Befunde unter Rückgriff auf weitere individuelle, familiäre und unterrichtliche Kontextdaten von Schreibkompetenz zu verdichten. So wurde beispielsweise Feedback bereits als zweiter zentraler Einflussfaktor innerhalb der Gelenkstelle der Schreibphase operationalisiert und kann als weitere unterrichtliche Determinante auf die Schreibleistung der Schülerinnen und Schüler untersucht werden.

Analysen, die die Bilderbuchrezeption vor der Aufgabenstellung sowie die Reflexionsphasen im Anschluss an die Textproduktion einbeziehen, sowie Untersuchungen auf Basis der

kriterial erfassten Schreibfähigkeiten vervollständigen die Beschreibung der domänenspezifischen Unterrichtspraxis in der ersten Klasse.

Um die Schreibkompetenzentwicklung als essenzielles Forschungsdesiderat im Kontext von empirischer Bildungsforschung und Schreibdidaktik in den Blick zu nehmen, wird in einem Folgeprojekt von NaSch1 eine längsschnittliche Untersuchung der Textqualität im zweiten und vierten Schuljahr angestrebt. Erneut soll unter fachwissenschaftlicher Verankerung eine Definition für das Konstrukt der Schreibkompetenz in Klasse 2 und 4 hergeleitet werden, auf deren Grundlage Messinstrumente konzipiert und latente Wachstumskurvenmodelle zur Erfassung der Kompetenzentwicklung berechnet werden. Von dieser Studie, die den Verlauf der Schreibkompetenz in der Grundschulzeit abbilden soll, werden zentrale Impulse für die Deutschdidaktik und Schreibforschung erwartet.

4. Literaturverzeichnis

- Abraham, U. & Kupfer-Schreiner, C. (Hrsg.). (2007). *Schreibaufgaben. Für die Klassen 1 bis 4*. Berlin: Cornelsen.
- Adams, R. J. & Khoo, S. T. (1996). *Quest*. Melbourne: Australian Council for Educational Research.
- Alamargot, D. & Fayol, M. (2009). Modelling the development of written composition. In R. Beard, D. Myhill, J. Riley & M. Nystrand (Eds.), *The SAGE Handbook of Writing Development* (pp. 23–47). London: SAGE.
- 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.
- Bachmann, T. & Becker-Mrotzek, M. (2010). Schreibaufgaben situieren und profilieren. In T. Pohl & T. Steinhoff (Hrsg.), *Textformen als Lernformen* (S. 191–210). Duisburg: Gilles & Francke.
- Bachmann, T. Ospelt-Geiger, B. Ospelt, K. & Vital, N. (2007). *Aufgaben mit Profil. Frühe Förderung funktional-pragmatischer Schreibfähigkeiten*. Unveröffentlichter Schlussbericht zum internen Projekt der Pädagogischen Hochschule Zürich.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (Hrsg.). (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske & Budrich.
- Baurmann, J. (2006). *Schreiben - Überarbeiten - Beurteilen. Ein Arbeitsbuch zur Schreibdidaktik*. Seelze (Velber): Kallmeyer.
- Beck, O. (1979). *Theorie und Praxis der Aufsatzbeurteilung*. Bochum: Ferdinand Kamp.
- 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, Berlin & Toronto: Budrich.
- Becker-Mrotzek, M. & Böttcher, I. (2011). *Schreibkompetenz entwickeln und beurteilen. Praxishandbuch für die Sekundarstufe I und II*. Berlin: Cornelsen-Scriptor.
- Becker-Mrotzek, M. & Schindler, K. (2007). Schreibkompetenz modellieren. In M. Becker-Mrotzek & K. Schindler (Hrsg.), *Texte schreiben* (S. 7–26). Duisburg: Gilles & Francke Verlag.
- Becker-Mrotzek, M. & Schindler, K. (2008). Schreibkompetenz modellieren, entwickeln und testen. *Didaktik Deutsch, Sonderheft 2*, 94–106.
- Bereiter, C. (1980). Development in Writing. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 73–93). Hillsdale, N.J.: Lawrence Erlbaum Associates Publishers.

- 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, 219–224.
- Blatt, I. (2011). Wie lässt sich Textqualität ermitteln und lernförderlich rückmelden? In J. Berning (Hrsg.), *Textwissen und Schreibbewusstsein. Beiträge aus Forschung und Praxis* (S. 89–114). Berlin: LIT Verlag.
- Blatt, I., Ramm, G. & Voss, A. (2009). Modellierung und Messung der Textkompetenz im Rahmen einer Lernstandserhebung in Klasse 6. *Didaktik Deutsch* (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.
- Blömeke, S., Risse, J., Müller, C., Eichler, D. & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht. Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft*, 34 (4), 330–357.
- Bohl, T. & Kleinknecht, M. (2009). Aufgabenkultur. In S. Blömeke (Hrsg.), *Handbuch Schule. Theorie - Organisation - Entwicklung* (UTB, Bd. 8392, S. 331–334). Bad Heilbrunn: Klinkhardt.
- Böhme, K., Bremerich-Vos, A. & Robitzsch, A. (2009). Aspekte der Kodierung von Schreibaufgaben. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 290–329). Weinheim: Beltz.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Böttcher, I. & Becker-Mrotzek, M. (2009). *Texte bearbeiten, bewerten und benoten*. Berlin: Cornelsen Verlag Scriptor.
- Bremerich-Vos, A. & Possmayer, M. (2011). Zur Reliabilität eines Modells der Entwicklung von Textkompetenz im Grundschulalter. *Didaktik Deutsch*, 31, 30–49.
- Brophy, J. E. (2009). Observational research on generic aspects of classroom teaching. In P.A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (pp. 755–780). New York: Routledge.
- Campbell, J., Kyriakides, L., Muijs, D. & Robinson, W. (2004). *Assessing teacher effectiveness. Developing a differentiated model*. London: Routledge.
- Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T. & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25, 471–533.
- Chomsky, N. (1968). *Language and mind*. Cambridge: Cambridge University Press.

- Christensen, C. A. (2009). The Critical Role Handwriting Plays in the Ability to Produce High-Quality Written Text. In R. Beard, D. Myhill, J. Riley & M. Nystrand (Eds.), *The SAGE Handbook of Writing Development* (pp. 284–299). London: SAGE.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen. Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31, 122–141.
- Cronbach, L., Gleser, G., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Dalehfeite, I. M. (2006). *Unterrichtsskripts – ein multikriterialer Ansatz. Eine Videostudie zum Zusammenspiel von Mustern unterrichtlicher Aktivitäten, Zielorientierung und prozessorientierter Lernbegleitung*. Dissertation, Christian-Albrechts-Universität Kiel. Online verfügbar: <http://d-nb.info/1019669373/34> (Zugriff: 04.10.2016).
- Dehn, M. (2009). Zur Funktion der Aufgabe für den Schülertext. In B. Hoffmann & R. Valtin (Hrsg.), *Projekte, Positionen, Perspektiven. 40 Jahre DGLS* (S. 145–175). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- Doucet, S. A. & Wilsdorf, A. (2005). *Lucy rettet Mama Kroko*. Hamburg: Oetinger.
- 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.
- Ellis, A. W. (1982). Spelling and writing (and reading and speaking). In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 113–146). New York: Academic Press.
- Erpenbeck, J. & Rosenstiel, L. v. (2003). *Handbuch Kompetenzmessung: Erkennen, Verstehen und Bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. Stuttgart: Schäffer-Poeschel.
- Feilke, H. (1995). Auf dem Weg zum Text. Die Entwicklung der Textkompetenz im Grundschulalter. In G. Augst (Hrsg.), *Frühes Schreiben. Studien zur Ontogenese der Literalität* (S. 69–88). Essen: Die Blaue Eule.
- 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: Max Niemeyer Verlag.

- Fend, H. (1981). *Theorie der Schule*. München: Urban und Schwarzenberg.
- Fix, M. (2000). *Textrevisionen in der Schule*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Fix, M. (2008). *Texte schreiben. Schreibprozesse im Deutschunterricht*. Paderborn: Schöningh.
- Früh, W. (2007). *Inhaltsanalyse. Theorie und Praxis* (6. überarbeitete und ergänzte Auflage). Konstanz: UVK-Verlag.
- Gabriel, K. (2014). *Videobasierte Erfassung von Unterrichtsqualität im Anfangsunterricht der Grundschule – Klassenführung und Unterrichtsklima in Deutsch und Mathematik*. Kassel: University Press.
- Gräsel, C. & Göbel, K. (2015). Unterrichtsqualität. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (2. Aufl., S. 107–119). Wiesbaden: VS Verlag.
- Gresser, A., Pohl, K., Corvacho del Toro, I., Greb, K. & Faust, G. (2011). Modul 2: Deutsch. In K. Greb, S. Poloczek, F. Lipowsky, G. Faust (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 1*. (Materialien zur Bildungsforschung Bd.23/1., S. 37–43). Frankfurt am Main: GFPP.
- 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* (Forschungsberichte des Landes Nordrhein-Westfalen, Nr. 3192, Fachgruppe Geisteswissenschaften). 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 (Deutsch Englisch Schülerleistungen International)* (S. 42–62). Weinheim: Beltz.
- Hartig, J. & Rakoczy, K. (2010). Mehrebenenanalyse. In H. Holling & B. Schmitz (Hrsg.), *Handbuch Statistik, Methoden und Evaluation* (S. 538–547). Göttingen: Hogrefe.
- Hattie, J. A. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J. & Timperley, H. (2007). The power of Feedback. *Review of Educational Research*, 77 (1), 81–112.
- Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Kallmeyer/Klett.
- Helmke, A., Helmke, T. & Schrader, F.-W. (2007). Qualität von Unterricht: Aktuelle Tendenzen und Herausforderungen im Hinblick auf die Evaluation und Entwicklung von Schule und Unterricht. *Pädagogische Rundschau*, 61 (5), 527–543.

- Hofen, N. (1980). *Messen und Beurteilen sprachlicher-produktiver Leistungen im Deutschaufsatz*. Dissertation. Mannheim.
- Hoppe, H. & Metz, K. (2013). Fachdidaktische Analyse von Aufgaben in Deutsch. In M. Kleinknecht, T. Bohl, U. Maier & K. Metz (Hrsg.), *Lern- und Leistungsaufgaben im Unterricht. Fächerübergreifende Kriterien zur Auswahl und Analyse* (S. 47–73). Bad Heilbrunn: Julius Klinkhardt.
- Hox, J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar & M. Schader (Eds.), *Classification, data analysis and data highways* (pp. 147–154). New York: Springer.
- Hugener, I. (2006). Überblick über die Beobachtungsinstrumente. In I. Hugener, C. Pauli & K. Reusser (Hrsg.), *Videoanalysen*. In E. Klieme, C. Pauli. & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Teil 3. Materialien zur Bildungsforschung. Bd. 15 (S. 45–54). Frankfurt am Main: Gesellschaft zur Förderung pädagogischer Forschung (GFPF).
- Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Givvin, K. B., Rust, K. et al. (2003). *Third international mathematics and science study 1999 video study technical report: Volume 1: Mathematics*. Washington: National Center for Education Statistics, Institute of Education Statistics, U. S. Department of Education.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M. et al. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Kammler, C. & Knapp, W. (2002). Empirische Unterrichtsforschung als Aufgabe der Deutschdidaktik. In C. Kammler & W. Knapp (Hrsg.), *Empirische Unterrichtsforschung und Deutschdidaktik* (S. 2–14). Baltmannsweiler: Schneider.
- Karst, K., Kastens, C., Mösko, E. & Lipowsky, F. (2013). Skalierung der Vor- und Nachtests der Videostudie im Fach Mathematik. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. In F. Lipowsky & G. Faust (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschulkindern" (PERLE)* (Materialien zur Bildungsforschung, 23/3, S. 423–481). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Kleinbub, I. (2016). Ein-Blick ins Klassenzimmer: deutschdidaktische Unterrichtsforschung am Beispiel der Videostudie VERA – Gute Unterrichtspraxis. In M. Krelle & W. Senn (Hrsg.), *Qualitäten von Deutschunterricht* (S. 15–36). Stuttgart: Fillibach bei Klett.

- Kleinknecht, M. (2010). *Aufgabenkultur im Unterricht. Eine empirisch-didaktische Video- und Interviewstudie an Hauptschulen*. Baltmannsweiler: Schneider.
- Klieme, E. (2005). Bildungsqualität und Standards. In G. Becker, A. Bremerich-Vos, M. Demmer, K. Maag Merki, B. Priebe, K. Schwippert et al. (Hrsg.), *Standards – Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten*. Friedrich Jahresheft 2005 (S. 5–6). Seelze: Friedrich.
- Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 52 (6), 765–773.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2007). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF), Referat Öffentlichkeitsarbeit.
- Klieme, E., Jude, N., Rauch, D., Ehlers, H., Helmke, A., Eichler, W. et al. (2008). Alltagspraxis, Qualität und Wirksamkeit des Deutschunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 319–344). Weinheim: Beltz.
- Klieme, E., Pauli, C. & Reusser, K. (Hrsg.) (2006). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Teil 3: Videoanalysen. Materialien zur Bildungsforschung. Bd. 15. Frankfurt am Main: Gesellschaft zur Förderung pädagogischer Forschung (GFPPF).
- Klieme, E., Pauli, C. & Reusser, K. (2009). Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik, T. Seidel & P. Najvar (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54 (2), 222–237.
- 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 et al. (Hrsg.), *Formate fachdidaktischer Forschung. Empirische Projekte – historische Analysen – theoretische Grundlegungen* (S. 49–65). Münster: Waxmann.
- Kobarg, M. (2004). *Die Bedeutung prozessorientierter Lernbegleitung für kognitive und motivationale Prozesse im Physikunterricht – eine Videostudie*. Institut für Psychologie der Christian-Albrechts-Universität Kiel: Unveröffentlichte Diplomarbeit.

- Kobarg, M. & Seidel, T. (2007). Prozessorientierte Lernbegleitung - Videoanalysen im Physikunterricht der Sekundarstufe I. *Unterrichtswissenschaft*, 35 (2), 148–168.
- Kötter, L. & Grau, U. (1965). Zur Bedingtheit der uneinheitlichen Benotung von Schüleraufsätzen (Nacherzählung). *Zeitschrift für experimentelle und angewandte Psychologie*, 12, 278–301.
- Krammer, K. (2009). *Individuelle Lernunterstützung in Schülerarbeitsphasen. Eine videobasierte Analyse des Unterstützungsverhaltens von Lehrpersonen im Mathematikunterricht*. Münster: Waxmann.
- Krammer, K. (2010). Individuelle Unterstützung im Unterricht mit 4- bis 8-jährigen Kindern. In M. Leuchter (Hrsg.), *Didaktik der ersten Bildungsjahre* (S.112–127). Seelze: Klett/Kallmeyer.
- Krammer, K., Reusser, K. & Pauli, C. (2010). Individuelle Unterstützung der Schülerinnen und Schüler durch die Lehrperson während der Schülerarbeitsphasen. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 107–122). Münster: Waxmann.
- Kruse, N. Reichardt, A. Herrmann, M. Heinzl, F. & Lipowsky, F. (2012). Zur Qualität von Kindertexten. Entwicklung eines Bewertungsinstrumentes in der Grundschule. *Didaktik Deutsch*, 32, 87–110.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kürzinger, A. & Pohlmann-Rother, S. (2015). Möglichkeiten einer objektiven und reliablen Bestimmung von Textqualität im Anfangsunterricht – Methodisches Vorgehen und deskriptive Befunde aus dem Projekt NaSch1. *Didaktik Deutsch*, 20 (38), 60–79.
- Kürzinger, A., Pohlmann-Rother, S. & Hess, M. (2017). Aufgabenstellung als zentrale Gelenkstelle im Schreibunterricht – Eine videogestützte Analyse. *Zeitschrift für Pädagogik*, 63 (5), 636–656.
- Kürzinger, A., Pohlmann-Rother, S., Hess, M. & Lipowsky, F. (2018). Aufgabenstellungen und ihre Bedeutung für die Textqualität von Schülerinnen und Schülern im Anfangsunterricht. *Unterrichtswissenschaft*, 46 (3), 327–343.
- Lehmann, R. H. (1987). The Scoring of Student Compositions. In W. Hartmann und 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.
- Lepper, M. R., Drake M. F. & O'Donnell-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning* (pp. 108–144). Cambridge: Brookline Books.
- Lipowsky, F. (2007). Unterrichtsqualität in der Grundschule – Ansätze und Befunde der nationalen und internationalen Forschung. In K. Möller, P. Hanke, Ch. Beinbrech, Th. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht entwickeln, erfassen und bewerten* (S. 35–49). Wiesbaden: Verlag für Sozialwissenschaften.
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 69-105; 2. überarbeitete Auflage). Heidelberg: Springer
- Lipowsky, F., Faust, G. Kastens, C. & Post, S. (2013). Die PERLE-Studie: Überblick und Hintergründe. In F. Lipowsky, G. Faust & C. Kastens (Hrsg.), *Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren* (S. 9–28). Münster: Waxmann.
- Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule. Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Wiesbaden: Springer VS Verlag für Sozialwissenschaften.
- Lotz, M., Berner, N. E. & Gabriel, K. (2013a). Auswertung der PERLE-Videostudien und Überblick über die Beobachtungsinstrumente. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), Technischer Bericht zu den PERLE-Videostudien. In F. Lipowsky & G. Faust (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschulkindern" (PERLE)* (Materialien zur Bildungsforschung, 23/3, S. 83–103). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lotz, M. & Corvacho del Toro, I. (2013). Die Videostudie im Fach Deutsch: „Lucy rettet Mama Kroko“. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), Technischer Bericht zu den PERLE-Videostudien. In F. Lipowsky & G. Faust (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschulkindern" (PERLE)* (Materialien zur Bildungsforschung, 23/3, S. 29–36). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Lotz, M., Gabriel, K. & Lipowsky, F. (2013b). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitigen Validierung. *Zeitschrift für Pädagogik*, 59 (3), 357–380.
- Lotz, M., Lipowsky, F. & Faust, G. (Hrsg.). (2013c). *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von*

- Grundschulkindern“ (PERLE) - Teil 3. Technischer Bericht zu den PERLE-Videoanalysen. Materialien zur Bildungsforschung, Band 23/3, Frankfurt am Main: GPF.*
- Lotz, M., Schoreit, E. & Kempfer, I. (2013d). Die Entwicklung des Leseverständnisses und der Rechtschreibfähigkeit. In F. Lipowsky, G. Faust & C. Kastens (Hrsg.), *Persönlichkeits- und Lernentwicklungen an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren* (S. 77–97). Münster: Waxmann.
- Maas, C. J. M. & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, 58, 127–137.
- McCutchen, D. (2006). Cognitive Factors in the Development of Children's Writing. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *The handbook of writing research* (pp. 115–130). New York: Guilford Press.
- Melhorn, G. & Melhorn, H.-G. (2003). Kreativitätspädagogik – Entwicklung eines Konzepts in Theorie und Praxis. *Bildung und Erziehung*, 56 (1), 23–45.
- Merz-Grötsch, J. (2005). *Schreiben als System. Band 1, Schreibforschung und Schreibdidaktik. Ein Überblick* (2. Aufl.). Freiburg i. Br.: Fillibach Verlag.
- Metz, K., Maier, U. & Kleinknecht, M. (2009). Kognitiver Anspruch von Aufgaben im Deutschunterricht. *ide*, 33 (3), 74–87.
- Metz, K., Meier, U., Kleinknecht, M., Bohl, T. & Hoppe, H. (2012). Einsatz eines fächerübergreifenden Kategoriensystems zur Analyse von Aufgaben im Fach Deutsch. In A. Ballis (Hrsg.), *Lernmedien und Lernaufgaben im Deutschunterricht. Konzeptionen und Analysen* (S. 25–47). Bad Heilbrunn: Klinkhardt.
- Moser, U., Berweger, S. & Lüchinger-Hutter, L. (2004). LEst 4-7. *Lern- und Entwicklungsstand bei 4- bis 7-Jährigen*. Unveröffentlichter Test.
- Muthén, L. & Muthén, B. (1998–2010). *Mplus user's guide*. (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen. Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie*. Hildesheim: Franzbecker.
- Neumann, A. (2007). *Briefe schreiben in Klasse 9 und 11. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen*. Münster: Waxmann.
- Neumann, A. (2014). Großuntersuchungen zur Schreibleistungsmessung. In H. Feilke & T. Pohl (Hrsg.), *Schriftlicher Sprachgebrauch - Texte verfassen* (S. 514–531). Baltmannsweiler: Schneider Verlag Hohengehren.
- Neumann, A. & Lehmann, R. H. (2008). Schreiben Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 89–103). Weinheim: Beltz.

- 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*, 141, 36–52.
- Pauli, C. (2008). Unterrichtsbeobachtung. In F. Hellmich (Hrsg.), *Lehr-Lernforschung und Grundschulpädagogik* (S. 143–155). Bad Heilbrunn: Klinkhardt.
- Pauli C. & Reusser, K. (2000). Zur Rolle der Lehrperson beim kooperativen Lernen. *Schweizerische Zeitschrift für Bildungswissenschaften*, 22 (3), 421–442.
- Pauli C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52 (6), 774–798.
- Persky, H. R., Daane, M. C. & Jin, Y. (U.S. Department of Education, Eds.). (2003). The Nations' Report Card - NAEP. Writing 2002, Institute of Education Sciences. Online verfügbar: <http://nces.ed.gov/nationsreportcard/pdf/main2002/2003529.pdf> (Zugriff: 04.10.2016).
- Petko, D., Waldis, M. Pauli, C. & Reusser, K. (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. *Zentralblatt für Didaktik der Mathematik*, 35 (6), 265–280.
- Philipp, M. (2012). Wirksame Schreibförderung, Metaanalytische Befunde im Überblick. *Didaktik Deutsch* (33), 59–73.
- Piaget, J. (1935/1998). Psychologische Anmerkungen zur Gruppenarbeit. In J. Piaget (Hrsg.), *Über Pädagogik* (S. 179-198). Weinheim: Beltz.
- Piaget, J. (1936/1969). *Das Erwachen der Intelligenz beim Kinde*. Stuttgart: Klett.
- Pohl, T. (2014). Entwicklung der Schreibkompetenz. In H. Feilke & T. Pohl (Hrsg.), *Schriftlicher Sprachgebrauch. Texte verfassen* (S. 101–140). Baltmannsweiler: Schneider.
- Pohl, T. & Steinhoff, T. (2010). Textformen als Lernformen. In T. Pohl, & T. Steinhoff (Hrsg.), *Textformen als Lernformen*. (S. 5–26). Duisburg: Gilles & Francke.
- Pohlmann-Rother, S. & Kürzinger, A. (2015). Individuelle Lernunterstützung in Schülerarbeitsphasen im Schreibunterricht der Grundschule. In K. Liebers, B. Landwehr, A. Marquardt & K. Schlotter (Hrsg.), *Lernprozessbegleitung und adaptive Lerngelegenheiten im Unterricht der Grundschule. Ergebnisse der 23. Jahrestagung der Kommission Grundschulforschung und Pädagogik der Primarstufe* (Bd. 1, S. 181–186). Wiesbaden: VS Verlag.
- Pohlmann-Rother, S., Kürzinger, A. & Lipowsky, F. (2018). Individuelle Lernunterstützung im schriftsprachlichen Anfangsunterricht: Formen, Verteilungsmuster und Wirksamkeit. *Zeitschrift für Grundschulforschung*, 11 (2), 315–332.

- Pohlmann-Rother, S. & Kürzinger, A. (i.V.). Die videobasierte Analyse individueller Lernunterstützungen im Schreibunterricht. In M. Lotz & F. Lipowsky (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien*. Band 2. In F. Lipowsky & M. Lotz (Hrsg.), Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschulkindern" (PERLE) (Materialien zur Bildungsforschung, 23/4). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPF).
- Pohlmann-Rother, S., Schoreit, E. & Kürzinger, A. (2016). Schreibkompetenzen von Erstklässlern quantitativ-empirisch erfassen – Herausforderungen und Zugewinn eines analytisch-kriterialen Vorgehens gegenüber einer holistischen Bewertung. *Journal for Educational Research Online*. 8 (2), 107–135.
- Pol, J. van de, Volman, M. & Beishuizen, J. (2010). Scaffolding in Teacher-Student Interaction: A Decade of Research. *Educational Psychology Review*, 22 (3), 271–296.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (Hrsg.). (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland - Was wissen und können Jugendliche?* Münster: Waxmann.
- 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.
- Rasch, G. (1960 [1980]). *Probabilistic models for some intelligence and attainment tests*. [Reprint, with a Foreword and Afterword by Benjamin D. Wright. Chicago: University of Chicago Press, 1980.]
- Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2008). *HLM 6.07 for Windows – student*. Lincolnwood, IL: Scientific Software International, Inc.
- Reusser, K. (2001) Co-constructivism in educational theory and practice. In N.J. Smelser, P.B. Baltes & F.E. Weinert (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 2058–2062). Oxford: Elsevier Science.
- Reusser, K. (2006). Konstruktivismus – Vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur. In M. Baer, M. Fuchs, P. Füglistner, K. Reusser & H. Wyss (Hrsg.), *Didaktik auf psychologischer Grundlage. Von Hans Aebli's kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung*. Bern: Hep.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität – Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S.9–32). Münster: Waxmann.

- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Sammons, P., Hillman, J. & Mortimore, P. (1995). *Key characteristics of effective schools. A review of school effectiveness research*. University of London, Institute of Education.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25, 465–493.
- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Schnebel, S. & Wagner, S. (2016). Kognitiv anregende Lernunterstützung in einem naturwissenschaftlichen Lernsetting. *Zeitschrift für Grundschulforschung* (9) 1, 107–120.
- Schneuwly, B. (1996). Der Nutzen psychologischer Schreibforschung für die Didaktik des Schreibens. In H. Feilke (Hrsg.), *Schreiben im Umbruch. Schreibforschung und schulisches Schreiben* (S. 29–39). Stuttgart: Klett.
- Schrader, F.-W. (1997). Lern- und Leistungsdiagnostik im Unterricht. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 559–661). Göttingen: Hofgrefe.
- Schulte, K. (2000). *Kohärenz. Eine Untersuchung an Texten aus Klasse 1*. Frankfurt a. M.: P. Lang.
- Seidel, T. (2003). Videobasierte Kodierverfahren in der IPN Videostudie Physik – Ein methodischer Überblick. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lernprozesse im Physikunterricht“* (S. 99–111). Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN).
- 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).
- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the last decade: role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77 (4), 454–499.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.) (2005). *Bildungsstandards im Fach Deutsch für den Primarbereich (4. Jahrgangsstufe)* – Beschluss vom 15.10.2004. München: Wolters Kluwer.
- Shavelson, R. & Dampsey-Atwood, N. (1976). Generalizability of measures of teaching behavior. *Review of Educational Research*, 46, 553–611.
- Stein, M. K., Grover, B. W. & Henningsen, M. (1996). Building Student Capacity for Mathematical Thinking and Reasoning: An Analysis of Mathematical Tasks Used in Reform Classrooms. *American Educational Research Journal*, 33 (2), 455–488.
- 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.

- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim und Basel: Beltz.
- Weinhold, S. (2000). *Text als Herausforderung. Zur Textkompetenz am Schulanfang. Mit 296 Schülertexten aus Klasse 1*. Freiburg i.B.: Fillibach.
- Weinhold, S. (2014). Schreiben in der Grundschule. In H. Feilke & T. Pohl (Hrsg.), *Schriftlicher Sprachgebrauch. Texte verfassen* (S. 143–158). Baltmannsweiler: Schneider.
- Wildemann, A. (2011). Qualität im Sprachlichen Anfangsunterricht – Fachdidaktische Perspektiven und empirische Notwendigkeiten. In K.-O. Bauer & N. Logemann (Hrsg.), *Unterrichtsqualität und fachdidaktische Forschung* (S. 75–100). Münster: Waxmann.
- Winkler, I. (2005). Zur Beziehung von Unterrichtsmaterial, -gestaltung und -erfolg: Drei Aufgaben zu Georg Brittings „Brudermord im Altwasser“ im Praxistest. In J. Stückrath & R. Strobel (Hrsg.), *Deutschunterricht empirisch*. (S. 177–196). Baltmannsweiler: Schneider Verlag Hohengehren.
- Wirtz, M. A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystem und Ratingskalen*. Göttingen: Hogrefe.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest Version 2.0. Generalized Item Response Modelling Software*. Melbourne: ACER.
- Wygotsky, L. S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Wygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: Harvard University Press.
- Wyss, C. (2013). *Unterricht und Reflexion. Eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften*. Münster: Waxmann.
- Ysewijn, P. (1996). *GT Programm für Generalisierbarkeitsstudien*. GT Handbuch.

5. Tabellenverzeichnis

Tabelle 1	Publikationen im Überblick.....	15
Tabelle 2	Varianzkomponenten und relativer Generalisierbarkeitskoeffizient für das holistische Rating und die hoch inferenten kriterialen Items (<i>N</i> = 500).....	38

6. Abbildungsverzeichnis

Abbildung 1	Rahmenbedingungen und Einflussfaktoren auf die Textqualität (in Anlehnung an Lotz, 2016; S. 76 und Helmke, 2014, S. 71).....	8
-------------	--	---

7. Anlagen

Anlage A

Kürzinger, A. & Pohlmann-Rother, S. (2015). Möglichkeiten einer objektiven und reliablen Bestimmung von Textqualität im Anfangsunterricht – Methodisches Vorgehen und deskriptive Befunde aus dem Projekt NaSch1. *Didaktik Deutsch*, 20 (38), 60-79.

Anlage B

Pohlmann-Rother, S., Schoreit, E. & Kürzinger, A. (2016). Schreibkompetenzen von Erstklässlern quantitativ-empirisch erfassen – Herausforderungen und Zugewinn eines analytisch-kriterialen Vorgehens gegenüber einer holistischen Bewertung. *Journal for Educational Research Online*, 8 (2), 107-135.

Anlage C

Kürzinger, A., Pohlmann-Rother, S. & Hess, M. (2017). Aufgabenstellung als zentrale Gelenkstelle im Schreibunterricht – Eine videogestützte Analyse. *Zeitschrift für Pädagogik*, 63 (5), 636-656.

1. Einleitung

Die Bedeutung von Sprachhandeln für den schulischen und beruflichen Erfolg ist unumstritten. Bereits in der Primarstufe werden die Weichen für das Schreibenlernen und die Entwicklung von Schreibfähigkeiten gestellt. In Klasse 1 werden die Schüler erstmals in einem institutionellen Rahmen mit dem Verfassen von Texten konfrontiert. Während die Schreibfähigkeiten im Grundschulalter vor allem von Klasse 2 bis 4 gut dokumentiert sind (z. B. Augst et al. 2007, Thonke et al. 2008, Böhme et al. 2009, Steinig et al. 2009, Bachmann 2012, Kruse et al. 2012), liegen insbesondere für den Anfangsunterricht kaum empirische Untersuchungen vor, die sich mit der Textqualität sowie ihrer objektiven, reliablen und validen Erhebung beschäftigen. Ein Forschungsdesiderat wird insbesondere für Befunde zum frühen Erwerb anspruchsvoller Schreibfähigkeiten konstatiert, obwohl die Notwendigkeit der frühen Förderung als unumstritten gilt (Bachmann et al. 2007).

Im DFG-Projekt NaSch1² (Narrative Schreibkompetenz in Klasse 1) werden Schreibfähigkeiten von Erstklässlern auf breiter Basis analysiert, indem die Textqualität von 540 Briefen anhand zweier fundierter Bewertungsinstrumente (analytisch und holistisch) eingeschätzt und in Abhängigkeit von individuellen, familiären und unterrichtlichen Bedingungen untersucht wird. Die Daten stammen aus dem BMBF-Projekt PERLE, das die Persönlichkeits- und Lernentwicklung von Grundschulkindern analysiert (Lipowsky et al. 2013)³.

Im vorliegenden Beitrag wird dargestellt, wie sich Textqualität im Anfangsunterricht nach empirischen Gütekriterien erheben lässt. Zudem werden erste deskriptive Befunde zur Qualität von Texten aus der ersten Jahrgangsstufe berichtet, die allerdings ausschließlich vor dem Hintergrund der selektiven PERLE-Stichprobe interpretiert werden können und sich nur bedingt auf konventionelle Grundschulen mit einer stärker heterogenen Zusammensetzung übertragen lassen.⁴

¹ Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift publizierten Originalbeitrag und kann daher nicht zitiert werden. Der Originalbeitrag ist erschienen als: Kürzinger, A. & Pohlmann-Rother, S. (2015). Möglichkeiten einer objektiven und reliablen Bestimmung von Textqualität im Anfangsunterricht – Methodisches Vorgehen und deskriptive Befunde aus dem Projekt NaSch1. *Didaktik Deutsch*, 20 (38), 60–79.

² Leitung: Sanna Pohlmann-Rother & Gabriele Faust (†), Förderkennzeichen PO 1739/1-1.

³ Leitung: Frank Lipowsky & Gabriele Faust.

⁴ Ein Ziel der PERLE-Studie bestand in dem Vergleich zwischen staatlichen und privaten Grundschulen. Für diese Gegenüberstellung wurden gezielt staatliche Schulen mit einem ähnlich

2. Zur Erfassung von Textqualität im Anfangsunterricht

Das Erfassen von Textqualität nach empirischen Gütekriterien weist in der Fachdidaktik keine lange Tradition auf (Becker-Mrotzek/Schindler 2008). Die Vielfältigkeit und Komplexität von Schülertexten, die auch bei stark strukturierten Aufgabenformaten vorliegen kann, erschwert dabei die Erfassung und Entwicklung theoriebasierter Kriterien, die sowohl Textqualität erfassen als auch den Kriterien der Objektivität, Reliabilität und Validität entsprechen (Blatt et al. 2009: 54f.). Sicher ist, dass die holistische sowie analytische Beurteilung empirisch fundierte Instrumente zur Bestimmung von Textqualität darstellen, da sie die Qualitätseinschätzungen über Beurteilerübereinstimmungen bzw. Raterreliabilitäten (Baurmann 2006, Grzesik/Fischer 1984) sichern können.

Die Ergebnisse einiger Untersuchungen dokumentieren jedoch, dass die Gütekriterien bei der Einschätzung von Schreibfähigkeiten zum Teil schwierig zu erfüllen sind. So werden einzelne Kriterien oftmals nicht ausreichend objektiv eingeschätzt (z. B. bei Blatt et al. 2009, hinreichende, aber eher niedrigere Kennwerte nach einer Modifikation bei Bremerich-Vos/Possmayer 2011). Neben den empirischen Gütekriterien ergeben sich auch aus didaktischer Perspektive Konsequenzen für die Entwicklung geeigneter Kriterien. Steinhoff (2010) plädiert im Kontext einer differenzierten Schülertextbeurteilung für die Berücksichtigung der von Pohl und Steinhoff (2010) konzipierten Entwicklungs-, Prozess- und Situierungsperspektive. Demnach wirken sich die Erwerbsphase eines Textes (z. B. Alter, Schulform), seine Produktion (z. B. erste vs. überarbeitete Version) sowie seine schreibdidaktisch-unterrichtliche Einbettung (z. B. Art der Schreibaufgabe, Lehr- und Lernziele) nachhaltig auf seine Qualität aus, so dass ein Schülertext nicht losgelöst vom Unterrichtsgeschehen betrachtet werden kann (Böttcher/Becker-Mrotzek 2003, Bachmann/Becker-Mrotzek 2010). Im Projekt NaSch1 wird diesen empirischen und didaktischen Forderungen Rechnung getragen, indem die Schreiberfahrungen der Schüler sowie das schreibdidaktische Setting bei der deduktiv-induktiven Entwicklung der Auswertungskriterien berücksichtigt werden.

3. Begriffsbestimmung und Entwicklung der Schreibkompetenz im Anfangsunterricht

Vor dem Hintergrund testtheoretischer Annahmen und der Interpretation von Schreibaufgaben als Schulleistungstests werden die produzierten Texte häufig als Ausweis der erwarteten Fähigkeit verstanden, einen angemessenen Text zu verfassen (Neumann 2007: 74, Böhme et al. 2009).

hohen sozioökonomischen Hintergrund der Schüler ausgewählt (Lipowsky et al. 2013). Für die Schüler der privaten Schulen zeigt sich ein höherer sozioökonomischer Status, der jedoch entschieden geringer ausfällt als es im Kontext der Privatschulforschung üblich ist (ebd.).

Dementsprechend wird dem Text eine gewisse Qualität, dem Schreiber werden gewisse Fähigkeiten (Kompetenzen) zugeschrieben (Knopp et al. 2012: 51). In NaSch1 wird im Sinne des Kompetenz-Performanz-Modells davon ausgegangen, dass eine gezeigte Leistung ein Indikator für eine dahinter liegende Schreibkompetenz ist (Harsch et al. 2007). 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: 12f., Knopp et al. 2012: 47-50, Fix 2008: 24). Aufgrund der Vielfalt und Komplexität von Texten und der verschiedenen möglichen Beurteilungskriterien erscheint es sinnvoll, das jeweilige Verständnis von Schreibkompetenz an dem Schreibanlass und der darauf bezogenen Textfunktion zu orientieren (Harsch et al. 2007: 54).

Einen Zugang zu den erwartbaren Schreibleistungen ermöglichen verschiedene Modelle der Schreibentwicklungsforschung. Dabei besteht vor allem hinsichtlich der Annahme eines linearen oder eines parallelen Verlaufs der Entwicklung der Schreibfähigkeiten Uneinigkeit. Stufenmodelle gehen davon aus, dass sich Schreibfähigkeiten sukzessive entfalten. Diese Annahmen werden von empirischen Befunden untermauert (Feilke/Augst 1989, Feilke 1995, Augst et al. 2007). Nach Augst et al. (2007: 351) entwickelt sich die Textsortenkompetenz in verschiedenen Stadien, wobei weg „die nächst höheren Stufen entwicklungslogisch aus der vorhergehenden ergeben.“ Das Überspringen von Entwicklungsstufen ist dabei nicht vorgesehen. Ungeübte Schreiber greifen vor allem auf die Schreibstrategie des „knowledge-tellings“ (Bereiter/Scardamalia 1987) zurück, bei der Ideen assoziativ generiert werden und das „Prinzip szenischer Kontiguität“ (Feilke 1996: 1186) dominiert. Eine Planung oder Leserorientierung findet nicht statt, vielmehr wird der Text aus dem Gedächtnis produziert, was inhaltliche Brüche zur Folge hat.

In anderen Studien zeichnet sich dagegen ab, dass auch Schreibanfänger – insbesondere unter bestimmten unterrichtlichen Voraussetzungen – Texte produzieren, die sprachlichen, inhaltlichen, strukturellen und formalen Gesichtspunkten genügen können (Schulte 2000, Weinhold 2000). Das bedeutet, dass sich sowohl einfache als auch komplexe Schreibfähigkeiten bereits von Anfang an in den Texten zeigen, sofern die Schüler mit entsprechenden Aufgaben betraut werden (Bachmann 2012). Die Ergebnisse von Schulte (2000) und Weinhold (2000) stützen die Modellierung der Schreibentwicklung von Becker-Mrotzek (1997), nach der sich die verschiedenen Dimensionen der Schreibfähigkeit nicht linear ausdifferenzieren, sondern zeitlich parallel entwickeln.

Entscheidend ist, dass die Entwicklung der Schreibkompetenz von der Schreiberfahrung (nicht vom biologischen Alter), von der Schreibmotivation und damit auch vom Unterricht in einem erheblichen Maß beeinflusst wird (Becker-Mrotzek/Böttcher 2011) und sich in Abhängigkeit von den an die Schüler gestellten Aufgaben weiter entfaltet (Feilke 1996).

4. Modellierung der Textqualität im Projekt NaSch1

Ausgehend von diesen Annahmen wird im Projekt NaSch1 ein vereinfachtes Modell der Textqualität zu Grunde gelegt, dessen Definition im Wesentlichen durch das Untersuchungsinteresse, die Schreiberfahrung der Schüler und das spezifische schreibdidaktische Setting bedingt werden. „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: 22). Schreibkompetenz wird vor diesem Hintergrund als altersgemäße (auf das „Schreibalter“ bezogen, vgl. Feilke 1996: 1181) Fertigkeit verstanden, die mit einer Schreibaufgabe verbundenen Anforderungen zu bewältigen. Die in NaSch1 entwickelten 12 Kriterien, die die Textqualität abbilden, lassen sich in formalstrukturelle (z. B. Anrede- und Schlussformel), sprachliche (z. B. Wortschatz) sowie inhaltliche Kriterien (z. B. Perspektivenübernahme) einteilen.

Briefmerkmale wie Anrede- und Schlussformeln sowie die von der Schreibaufgabe vorgegebene Perspektivenübernahme charakterisieren die schreibaufgaben- bzw. textsortenspezifischen Kriterien. Das Konzept der Perspektivenübernahme orientiert sich am Verfahren des perspektivischen Schreibens (Fritzsche 1989, Böttcher 2010: 169), indem die Schreiber in der Ich-Form einen Text aus der Perspektive einer literarischen Figur verfassen sollen. Im Mittelpunkt der Textanalyse steht die sprachliche Dimension, die sich in sprachsystematische (z. B. Wortschatz, Angemessenheit der Sprachmittel) und kohärenzstiftende Kriterien (Textidee und -entfaltung, Kohäsionsmittel, Implizitheit) einteilen lässt. Als diffizil bei der Analyse erweisen sich vor allem die wenig ausgeprägten Schreiberfahrungen der Schüler sowie der damit zusammenhängende geringe Textumfang. Aus diesem Grund ist etwa das Kriterium „Textidee und -entfaltung“ als eine Vorstufe zu den Kategorien „Gesamtidee, Thema, Absicht des Textes; Aufbau, Gliederung; thematische Entfaltung“ aus dem Zürcher Textanalyseraster (Nussbaumer/Sieber 1995: 45f.) zu 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. Nach Steinig et al. entwickelt sich Kohärenz nur dann,

wenn der Textproduzent einzuschätzen vermag, was er an Vorwissen bei Rezipienten voraussetzen kann und was nicht. Ist seine Einschätzung nicht adäquat, besteht die Gefahr, zu viele unnötige Informationen zu liefern. Wesentlich häufiger sind jedoch Informationslücken – Bruchstellen im Text, die der Leser nur mühsam oder gar nicht zu stopfen vermag. (2009: 160)

Das vorliegende Aufgabensetting sah vor, dass die Schüler einen Brief aus der Sicht einer literarischen Figur (Lucy) an eine andere literarische Figur und somit an einen fiktiven Adressaten bzw. Leser (Krokodilmutter) schreiben sollten. Insofern muss zwischen einem fiktiven und einem

realen Leser unterschieden werden. Was für den realen Leser eine Informationslücke sein kann, weil er kein Wissen über die gemeinsame Kommunikationsgeschichte zwischen beiden Figuren hat, muss für den fiktiven Leser keine Bruchstelle sein, sofern das von den literarischen Figuren geteilte Wissen berücksichtigt wird. Inkohärenzen in den Texten werden somit entsprechend der Aufgabenstellung ausschließlich im Kontext des gemeinsamen Wissens von fiktivem Absender und fiktivem Leser beurteilt, was mit einem aufgeweichten Kohärenzkriterium mit Blick auf den realen Leser einhergeht. Dieser der Bewertung des Kohärenzaspekts zugrunde liegende, entscheidende Gedanke wird von Steinig et al. (2009: 160) aufgegriffen: „Ein vertraulicher Brief an einen guten Freund, in dem sehr viel präsupponiert wird, kann sehr wohl kohärent sein, weil Schreiber und Leser über das zum Verständnis notwendige gemeinsame Wissen verfügen“.

Anhand des Items „Implizitheit“ wurde erhoben, inwieweit eine semantische Relation expliziert wird oder erst erschlossen werden muss. Mit der Konzipierung ergibt sich eine Abweichung vom ursprünglichen Kriterium im Zürcher Textanalyseraster. Der Grad an Implizitheit/Explizitheit wird dort im Sinne eines Kontinuums aufgefasst, in dem die goldene Mitte zwischen zu großer Implizitheit und zu großer Explizitheit (also zwischen zu wenig und zu viel Information) neutrale, unauffällige Texte umfasst. Dieses Konzept lässt sich auf das vorliegende Textkorpus insbesondere aufgrund des geringen Textumfangs der Schreibprodukte kaum übertragen. In der empirischen Analyse bestätigt sich die geringe Varianz des Items Explizitheit: Bis auf wenige Ausnahmen lassen sich die Texte nicht als zu explizit charakterisieren, weshalb es bei weiteren Analysen unberücksichtigt bleibt.

Darüber hinaus wurden auch Oberflächenmerkmale wie die Text- und Satzlänge sowie die Themen der Briefe erfasst. Sie vervollständigen die Darstellung der Schreibprodukte in Klasse 1, werden jedoch für die Ermittlung von Textqualität nicht herangezogen.

5. Schreibdidaktisches Setting *Lucy rettet Mama Kroko*

Charakteristisch für das Schreibarrangement ist ein handlungs- und produktionsorientiertes Verfahren, in dessen Kontext in 37 Klassen Briefe als Anschlussaufgabe an das rezipierte Bilderbuch *Lucy rettet Mama Kroko* (Doucet/Wilsdorf 2005) verfasst wurden. Der literarische Schreibanlass beinhaltet den Auftrag, einen Brief aus der Perspektive der Hauptfigur Lucy an die Krokodilmutter Mama Kroko zu dem Zeitpunkt zu schreiben, als das Mädchen ihre Krokodilfamilie verlässt. Die Orientierung an einem Adressaten erfordert bei den Schülern u. a. Empathie und das Vermögen, sich sowohl in die literarische Figur (Lucy) als auch in den fiktiven Adressaten (Mama Kroko) hineinzusetzen. Jüngeren und ungeübten Schreibern bereitet jedoch gerade die Trennung von Empfänger und Sender beim Verfassen von Texten, also die „Zerdehnung der Sprechsituation“

(Ehlich 1983), erhebliche Schwierigkeiten. Statt mit einem sichtbaren Gesprächspartner, mit dem man in direkten Dialog tritt, hat man beim Schreiben mit einem Leser zu tun, den man sich vorstellen muss. Möglicherweise lassen sich diese Anforderungen an die Schüler durch die situative Einbettung der Aufgabe in einen fiktiven Handlungsrahmen, den handlungs- und produktionsorientierten Literaturunterricht und durch die Verwendung der Textsorte Brief, die mit ihrem formalisierten Aufbau eine Orientierungshilfe für Schreibanfänger darstellen kann, relativieren. Die starke Strukturierung der Schreibaufgabe durch die Vorgaben erfüllt nicht nur empirische Anforderungen an die Leistungsmessung und an die Vergleichbarkeit der Auswertung, sondern kann insbesondere für leistungsschwächere Schüler unterstützend und lernförderlich wirken (May 2001: 183).

6. Modellierung der Textbewertungsinstrumente im Projekt NaSch1

In einem ersten Schritt wurden die Texte einem holistischen Rating unterzogen, das unter Rückgriff auf den NAEP Holistic Scoring Guide (Persky et al. 2003) sowie das holistische Analyseinstrument von Böhme et al. (2009) entwickelt und entsprechend modifiziert wurde.⁵ Als Bewertungsgrundlage für die Textqualität wurden folgende Teilaspekte bzw. Indikatoren herangezogen (vgl. auch Böhme et al. 2009):

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

Der folgenden Tabelle 1 kann eine verkürzte Übersicht über die Ratingstufen des Bewertungsinstruments entnommen werden.

⁵ Die holistische Auswertung der Texte wurde bereits im Projekt PERLE realisiert (vgl. Kürzinger et al. 2013).

Tab. 1: Die Ratingskala der holistischen Textbeurteilung⁶

Ausprägung	Gesamteindruck
Wert 4	<i>Hohe Textqualität:</i> Erfüllung der Aufgabenstellung durch Perspektivenübernahme von Lucy, Ansprache Mama Krokos und die Thematisierung des Abschieds (explizit oder implizit) sowie dessen Gründe. Text vergleichsweise detailreich, logisch nachvollziehbar und kohärent, gut leserliches Schriftbild, angemessene lautgetreue Verschriftlichung und adäquate sprachliche Ausdrucksweise; geringfügige Fehler möglich
Wert 3	<i>Durchschnittliche Textqualität:</i> Klarer Aufbau, Perspektivenübernahme und Thematisierung des Abschieds; Text insgesamt weniger detailliert, leserliches Schriftbild, teilweise nicht lautgetreue Schreibungen und wenige Fehler im Satzbau
Wert 2	<i>Eingeschränkte Textqualität:</i> Grundsätzlich Erfüllung der Aufgabenstellung; keine bzw. nur ansatzweise Thematisierung des Abschieds, kaum bzw. keine Details, sprachliche Defizite, schlechte Leserlichkeit sowie Mängel in der lautgetreuen Verschriftung; nur teilweise kohärent
Wert 1	<i>Schwache Textqualität:</i> Durch fundamentale Mängel in der Leserlichkeit des Schriftbilds, im Sprachgebrauch, Satzbau oder im kohärenten Aufbau kaum mehr verständlich; keine bzw. keine nachvollziehbare Erfüllung der Aufgabenstellung

Nach der holistischen Auswertung wurden die Texte anhand eines Kriterienkatalogs bewertet, der inhaltsbezogene, formal-strukturelle und sprachliche Facetten beinhaltet.⁷ In der nachfolgenden Tabelle 2 ist die stark verkürzte Ratinganleitung beispielhaft für zwei der insgesamt fünf vierstufigen Items angegeben.

⁶ Benchmarktexte sind in Tabelle 7 angegeben.

⁷ Im Kriterienkatalog (Kürzinger/Pohlmann-Rother 2014) sind alle Auswertungskriterien mit präziser Anleitung sowie Benchmarktexten aufgelistet.

Tab. 2: Vierstufige Items der analytischen Textbeurteilung

Wortschatz	
Grundidee	Das Item erfasst, wie differenziert die Wortschatzverwendung ist. Ein einfacher Wortschatz ist an umgangssprachlichen und/oder dialektalen Ausdrücken zu erkennen. Elaboriert ist der Wortschatz dann, wenn er abwechslungsreich und treffend ist.
Ratinganleitung mit Benchmarktexten	<p>Wert 1: sehr einfacher Wortschatz; Umgangssprache, mündlich geprägte bzw. dialekthafte Ausdrücke sowie eine stark vereinfachte Ausdrucksweise dominieren („Liebe Mama Kroko, ich hau ab“)</p> <p>Wert 2: Wortschatz ist einfach und zweckmäßig; Umgangssprache wird verwendet, dominiert jedoch nicht („Liebe Mama Kroko, ich bin abgehauen, weil deine Kinder mich ausgelacht haben“)</p> <p>Wert 3: Wortschatz ist weitgehend elaboriert, nahezu oder vollständig schriftsprachlich und abwechslungsreich („Liebe Mama Kroko, ich habe bei dir sehr viel gelernt. Ich möchte bitte eine Entschuldigung“)</p> <p>Wert 4: Wortschatz ist vollständig elaboriert; Norm der Schriftsprache ist erfüllt; es lassen sich elaborierte Verben bzw. Ausdrücke vorfinden („Liebe Mama Kroko, du musst dir keine Sorgen um mich machen, denn ich bin nur fort geschwommen, weil mich die anderen geärgert haben. Deine Lucy“)</p>
Textidee und -entfaltung	
Grundidee	Dieses Item bewertet die Idee eines Textes sowie seine Gliederung und thematische Entfaltung. Aufgrund der begrenzten Schreibentwicklung sind jedoch Einschränkungen zu treffen. Die thematische Entfaltung lässt sich aufgrund der kurzen Texte nach üblichen Maßstäben nur sehr schwer beurteilen. Als Textidee bzw. als Textthema wird daher die Proposition eines Satzes, also die Satzaussage, herangezogen. Diese Proposition, mit der ein Sachverhalt ausgedrückt wird, kann als Vorstufe zur Textidee betrachtet werden und ersetzt im Kontext der Texte aus Klasse 1 die Idee bzw. das Thema eines Briefs.
Ratinganleitung mit	Wert 1: Eine einfache, lediglich im Ansatz entfaltete Textidee oder keine Textidee und -entfaltung („[,] ich bin []Das ist Lucy kann schwimmen.“)

Benchmarktexten

Wert 2: Ansatzweise erkennbare Textidee und -entfaltung. Eine einzelne Proposition ist zu einer Gesamtidee entfaltet oder mehrere Propositionen bauen zum Teil aufeinander auf („Liebe Krokomama, ich wünsche dir viel Freude und viel Glück und viel Spaß und liebe Kinder. Tschüß deine Lucy.“)

Wert 3: Weitgehend erkennbare Textidee und -entfaltung. Gesamtidee aus mindestens zwei Propositionen, weitgehend aufeinander aufbauende Textteile („Liebe Mama Kroko, alle lachen mich aus. Kannst du mich nach Hause bringen?“)

Wert 4: Vollständig erkennbare Textidee und -entfaltung. Deutlich thematisch entfaltete Gesamtidee (mindestens drei Propositionen) von Anfang bis Ende; durchgängig aufeinander aufbauende Textteile („Liebe Mama Kroko, ich werde dich sehr vermissen, Mama Kroko. Bleib gut und gesund. Ich bedanke mich noch mal.“)

In der Tabelle 3 werden die niedrig und mittel inferenten Items der analytischen Textbeurteilung vorgestellt.

Tab. 3: Überblick über die niedrig und mittel inferenten Items der Textqualitätsanalyse

Item	Grundidee
Anrede	Das Item erfasst, ob der Brief eine korrekte Anrede enthält.
Schlussformel	Das Item erfasst, ob der Brief eine korrekte Schlussformel enthält.
Textende	Das Item erfasst, ob ein Brief vollständig zu Ende geführt wurde.
Aufgabenstellung	Das Item erfasst, ob der Schreibanlass berücksichtigt wird und sich die Briefe inhaltlich am Bilderbuch <i>Lucy rettet Mama Kroko</i> orientieren.
Perspektivenübernahme	Das Item erfasst, ob die Perspektive der literarischen Figur (Lucy) eingenommen wird.

7. Methodische Umsetzung der Textbewertung durch die beiden Instrumente

Die Auswertung der Texte anhand der beiden Bewertungsinstrumente wurde unabhängig voneinander und mit unterschiedlichen Ratern bzw. Kodierern umgesetzt. Auf diese Weise lassen sich Konsistenztendenzen vermeiden (vgl. Böhme et al. 2009).

Sowohl bei der analytischen als auch bei der holistischen Einschätzung wurden studentische Hilfskräfte geschult, die die Auswertung der Texte zusammen mit dem Entwickler der Instrumente vorgenommen haben. Bei den mehrtägigen Trainings wurde unter Einsatz einer hohen Anzahl von Übungstexten (ca. 100 Texte) auf ein gemeinsames theoretisches Verständnis der Kriterien fokussiert, was zur Validierung der Instrumente beitragen kann und die Basis für eine objektive Auswertung darstellt.

8. Ergebnisdarstellung

8.1 Beurteilerübereinstimmung und Reliabilitäten

Beide Textbewertungsinstrumente wurden auf die Übereinstimmung der Rater- bzw. Kodiererurteile überprüft. Die Güte der Einschätzung wird dabei anhand von Kennwerten abgebildet, die sich an denen des PERLE Projekts orientieren (Lotz et al. 2013): Sie betragen bei der prozentualen Übereinstimmung ($P\bar{U}$) $\geq 85.00\%$, bei Cohens Kappa (k) $\geq .70$ und beim relativen Generalisierbarkeitskoeffizienten (g) $\geq .70$. Bei hoch inferenten Kriterien liegen Kennwerte zu den Reliabilitäten über die gesamte Stichprobe vor, bei niedrig inferenten Beurteilerübereinstimmungen für jeden dritten Text. Darüber hinaus wurden auch für jeden dritten Brief der Stichprobe Übereinstimmungen der Rater/Kodierer mit dem Manualentwickler berechnet, um zu gewährleisten, dass die Beurteiler die Texte tatsächlich auf Basis des Manuals und des gemeinsam entwickelten Verständnisses eingeschätzt haben. In Tabelle 4 sind die Kennwerte zur Dokumentation der Interraterreliabilität des holistischen sowie analytischen Ratings dargestellt.

Tab 4: Varianzkomponenten und relativer Generalisierbarkeitskoeffizient für das holistische Rating und die hoch inferenten analytischen Items ($N = 500$)

Ratingdimension	g_{relativ}	Varianzkomponenten (VK)		
		Brief	Rater	BxR+e
Holistisches Rating	.91	84 %	0 %	16 %
Textidee und -entfaltung	.99	98 %	0 %	2 %
Implizitheit	.99	99 %	0 %	1 %
Kohäsionsmittel	.99	99 %	0 %	1 %
Angemessenheit der Sprachmittel ⁸	.98	95 %	0 %	4 %
Wortschatzverwendung	.97	96 %	0 %	4 %

Anmerkungen: g_{relativ} = relativer Generalisierbarkeitskoeffizient; BxR+e = Varianzkomponente Interaktion Brief und Rater sowie unsystematische Variation.

Der relative Generalisierbarkeitskoeffizient liegt für beide Instrumente deutlich über dem Richtwert und verweist auf eine hohe Übereinstimmung zwischen den Ratern. Dieser Befund wird durch die Angabe von Varianzkomponenten ergänzt, die bestätigen, dass die Unterschiede in den Einschätzungen weitgehend unabhängig von den Ratern zustande gekommen sind und überwiegend durch tatsächliche Unterschiede in den Briefen bedingt werden. Neben der Interraterreliabilität wurde zusätzlich die Intraraterreliabilität für beide Ratings bestimmt. Diese lässt Rückschlüsse auf die Reproduzierbarkeit der Urteile bzw. ihre zeitliche Stabilität zu, indem eine zufällige Auswahl an Briefen zu einem zweiten Messzeitpunkt eingeschätzt und die Übereinstimmung mit dem ersten Messzeitpunkt ermittelt wird. Zwischen beiden Daten lag eine Dauer von mindestens zwei Monaten. Die mit $g_{\text{relativ}} = .91-1.0$ belegen zumindest für einen Teil der Stichprobe ($N = 100$), dass die Urteile der Rater auch im Zeitverlauf stabil bleiben.

Bei den niedrig und mittel inferenten Items der analytischen Auswertung werden ebenfalls Übereinstimmungen erreicht, die über den Mindestkennwerten liegen (vgl. Tab. 5).

⁸ Bei der Berechnung der Varianzkomponenten mit dem GT-Programm für Generalisierbarkeitsstudien wird automatisch gerundet, weswegen Rundungsdifferenzen von bis zu 1 % auftreten können (vgl. auch Clausen et al. 2003).

Tab. 5: Beurteilerübereinstimmung der niedrig und mittel inferenten Items der Textqualitätsanalyse ($N = 179$)

Kriterium	Prozentuale Übereinstimmung	Cohens Kappa
Anrede	97,5 %	.79
Schlussformel	98,7 %	.97
Textende	94,9 %	.75
Aufgabenstellung	99,4 %	.88
Perspektivenübernahme	98 %	.79

Die Ergebnisse dieser Kodierung zur Intraraterreliabilität belegen für einen Teil der Stichprobe ($N = 67$) ebenso hohe Übereinstimmungen und damit zeitstabile Urteile. Bei den niedrig inferenten Items liegen die Werte bei Cohens Kappa zwischen $k = .83$ und 1.0, bei der prozentualen Übereinstimmung zwischen 90,2 %-100 %.

8.2 Deskriptive Ergebnisse zur Textqualität im Anfangsunterricht

Die Ergebnisdarstellung konzentriert sich auf ausgewählte deskriptive Analysen, anhand derer erste Aussagen über die Schreibfähigkeiten von Erstklässlern getroffen werden können. Allgemein lässt sich auf Basis der in Tabelle 6 dargestellten Mittelwerte zur Textqualität festhalten, dass es den Schülern durchschnittlich gelingt, verständliche (Textidee und -entfaltung, Impliztheit, Kohäsionsmittel) und sprachlich angemessene (Angemessenheit der Sprachmittel) sowie differenzierte (Wortschatzverwendung) Texte zu verfassen. In Bezug auf die Textverständlichkeit als Qualitätsaspekt scheint es den meisten Kindern nicht schwer zu fallen, eine Schreibidee zu entwickeln und die Beziehung zwischen Mama Krokodil und Lucy in ihren Briefen zu berücksichtigen. Allein die Verwendung von Kohäsionsmitteln ist für die Schreibanfänger anscheinend etwas schwieriger, aber trotzdem zu bewältigen.

Tab. 6: Textqualität auf Basis der hoch inferenten analytischen Items sowie des holistischen Ratings (vierstufig 1 bis 4; $N = 540$)

Ratingdimension	<i>M</i>	<i>SD</i>	Modus
Holistisches Rating	2.4	.89	2
Textidee und -entfaltung	2.6	.97	3

Implizitheit	2.9	.96	3
Kohäsionsmittel	2.3	1.0	2
Angemessenheit der Sprachmittel	2.7	.78	3
Wortschatzverwendung	2.7	.80	3

Auch die Umsetzung der schreibaufgaben- und textsortenspezifischen Merkmale in den Schülertexten ist als hoch zu bewerten. So wird beispielsweise das perspektivische Schreiben (Perspektivenübernahme) zu 92,2 % realisiert. Einen ähnlich hohen Wert weisen die Orientierung an der Aufgabenstellung (96,5 %) sowie die Textsortenmerkmale (66,1 % bis 90,6 %) auf. Ergänzende deskriptive Informationen liefert der „Null-Teil“ des Kriterienkatalogs, der nicht in die Qualitätsbewertung eingeht. Demnach schreiben die Erstklässler im Mittel 2.1 Sätze ($SD = 1.38$; Min-Max = 1-10) bei durchschnittlich 17,4 Wörter ($SD = 9.64$).

Einen eher didaktischen Zugang ermöglichen vier Briefbeispiele, die unterschiedliche Textqualitäten anhand der Einschätzung durch beide Instrumente dokumentieren (vgl. Tab. 7). Die Qualitätsabstufungen manifestieren sich dabei vor allem in den sprachlichen, kohärenzbezogenen Kriterien.

Tab. 7: Vergleich der Textqualität anhand ausgewählter Beispiele aus dem Textkorpus⁹

Nr.	Text	Wert Gesamt- eindruck (Wert 1- 4)	Werte der hoch inferenten Items der analytischen Beurteilung (Werte 1-4)				
			Wort- schatz	Textidee und -ent- faltung	Sprach- mittel	Implizit- heit	Kohäsions- mittel
1	„Liebe Mama, Beschwerde. Es ärgert. Entschuldigung. Viel gelernt: Futter, Fangen, Schwimmen. Schwimmt fort.“	1	1	1	1	1	1

⁹ Alle in der Tabelle 7 angegebenen Werte ergeben sich aus den gemittelten Raterurteilen. Der Wert 3,5 für den Wortschatz des Briefs Nr. 2 stellt eine Abweichung zwischen den Ratern dar (Rater 1 vergibt Wert 4, Rater 2 den Wert 3).

	Mama.“							
2	„Liebe Mama Kroko, ich habe bei dir sehr viel gelernt. Ich möchte bitte eine Entschuldigung. Deine Lucy.“	2	3,5	2	3	2	2	
3	„Liebe Mama Kroko, ich werde dich sehr vermissen, Mama Kroko. Bleib gut und gesund. Ich bedanke mich nochmal. Dein Mio ¹⁰ .“	3	3	4	3	2	3	
4	„Liebe Mama Kroko, ich habe dich verlassen, weil ich gehänselt wurde. Warum? Ja, weil ich nicht brüllen kann und weil ich nicht so schwimmen kann und weil ich nicht so aus- sehe wie sie. Viele Grüße, Lucy.“	4	4	4	4	4	3	

Insbesondere der erste Text demonstriert eine anfängliche Form des „knowledge-tellings“ (Bereiter/Scaradmalia 1987), bei der ohne Planung das, was dem Schreiber in den Sinn kommt, aufgeschrieben wird. Der Brief erinnert aufgrund seiner elliptischen Charakteristik stark an einen Telegrammstil, der sich vor dem Hintergrund des Schreibalters weniger als Stilmittel, sondern vielmehr als assoziatives Schreiben interpretieren lässt. Dadurch, dass die einzelnen Sätze nur rudimentär Verben bzw. Substantive enthalten und unverbunden (ohne Kohäsionsmittel)

¹⁰ Der Name des Schülers wurde anonymisiert.

aneinandergereiht sind, wird das Verständnis stark erschwert („Es ärgert. Entschuldigung“). Dies bewirkt letztendlich, dass dem Brief trotz der im Kern erfüllten Aufgabenstellung (z. B. Thematisierung des Abschieds) eine geringe Qualität zugeschrieben wird.

Ein weiterhin existentes Kohärenzproblem besteht bei Textbeispiel 2. Einzeln für sich genommen sind die beiden Sätze begreifbar und lassen keine sprachlichen Defizite erkennen, was gerade im Vergleich zu Textbeispiel 1 positiv auffällt. Jedoch bauen die beiden Aussagen nicht aufeinander auf und besitzen keinen logischen Zusammenhang. Aufgrund dieser fehlenden Kohärenz und Kohäsion sowie des nicht thematisierten Abschieds wird trotz der guten sprachlichen Leistung der Wert 2 vergeben.

Bei dem dritten und vierten Beispiel aus dem Textkorpus zeigt sich dagegen ein kohärenter Aufbau. Im Beispiel 3 wird ein konventioneller Verabschiedungskontext aufgegriffen, der vor allem die emotionale Ebene reflektiert. So werden Gefühlsbekundung („Ich werde dich sehr vermissen, Mama Kroko“), gute Wünsche („Bleib gut und gesund“) sowie ein nicht näher beschriebener Dank („Ich bedanke mich nochmal“) ausgesprochen. Im Gegensatz zu Beispiel 4 mit der höchsten Bewertung ist der Brief jedoch wenig detailreich und liefert keine Begründung für den Abschied, der zudem nur implizit („Ich werde dich sehr vermissen, Mama Kroko“) thematisiert wird, weswegen ihm eine durchschnittliche Textqualität zugeschrieben wird. Bemerkenswert in Text 4 ist neben der hohen sprachlichen wie kohärenzbezogenen Qualität insbesondere die Leserorientierung durch die Frage-Antwort-Sequenz („Warum? Ja, weil ich nicht brüllen kann [...]“). Aus diesen Gründen demonstriert dieses Beispiel trotz des fehlerhaften Verweismittels („und weil ich nicht so aussehe wie sie“) eine hohe Qualität.

Zusammenfassend lässt sich an dieser Auswahl von Beispielbriefen untermauern, dass Schreibanfänger Mitte der ersten Klasse unter bestimmten Voraussetzungen Texte unterschiedlicher Qualität verfassen können. Mit der Annahme einer linearen Ausdifferenzierung von Schreibfähigkeiten, wie sie in Stufenmodellen proklamiert wird, wären wahrscheinlich hauptsächlich Texte zu vereinbaren, wie sie in den Beispielen 1 und 2 vorliegen. Dagegen sprechen jedoch die empirischen Befunde, die im Rahmen der holistischen Auswertung eine mittlere Schreibfähigkeit ($M = 2.4$) belegen. Die zumindest im Ansatz in den Daten vorhandene Varianz, die auch an den Beispieltextrn sichtbar wird, spricht damit eher für eine zeitlich parallele Entwicklung der Schreibkompetenz.

8.3 Textlänge als Einflussfaktor

Die Annahme, dass die hohen Übereinstimmungen der Beurteiler auf die zeitintensive Schulung und hohe Anzahl an Übungstexten zurückzuführen sind, ist naheliegend. Die hohen

Übereinstimmungen können jedoch insbesondere vor dem Hintergrund des geringen Umfangs der Texte (durchschnittlich zwei Sätze pro Brief und 17,4 Wörter) auch skeptisch stimmen. Dementsprechend weisen Befunde im Rahmen von Nacherzählungen darauf hin, dass längere Texte weniger konsistent bewertet werden (Kötter/Grau 1965, vgl. auch Baurmann 2006), womit ein Einfluss der Textlänge auf die Bewertung deutlich wird.

Analysen zu ausgewählten Stichproben innerhalb des NaSch1-Korpus belegen zumindest im Ansatz, dass die Textlänge keinen messbaren Einfluss auf die Übereinstimmung der Rater untereinander besitzt (vgl. Tab. 8). Sowohl für kürzere (bis zu 17 Wörter) als auch für längere Texte (ab 17 Wörter) werden die Richtwerte des relativen Generalisierbarkeitskoeffizienten als Maß für die Übereinstimmung erfüllt. Im Vergleich zur Stichprobe mit allen Texten ($N = 500$) verringert sich der Kennwert zwar, bleibt aber in dem Rahmen, der für eine konsistente Einschätzung spricht.

Tab. 8: Gegenüberstellung der Reliabilitätskennwerte für verschiedene Textlängen-Stichproben

Ratingdimension	g_{relativ} für gesamte Stichprobe ($N = 500$)	g_{relativ} für Stichprobe bis einschließlich 17 Wörter ($N = 318$)	g_{relativ} für Stichprobe ab 17 Wörter ($N = 222$)
Holistisches Rating	.91	.85	.87
Textidee und -entfaltung	.99	.96	.95
Implizitheit	.99	.99	.97
Kohäsionsmittel	.99	.98	.99
Angemessenheit der Sprachmittel	.98	.86	.78
Wortschatzverwendung	.97	.89	.87

Diese Resultate lassen sich jedoch mit Vorsicht interpretieren, da die Unterschiede in der Textlänge vergleichsweise gering ausfallen. Zudem können auf Grundlage der Beurteilerkonsistenz keine Aussagen darüber getroffen werden, welche Rolle die Textlänge für die Einschätzung der Textqualität einnimmt.

Anhand der Mittelwerte des holistischen sowie der analytischen Urteile zeigt sich, dass den Texten mit einem geringeren Umfang eine niedrigere Qualität zugesprochen wird. Dies kennzeichnet sich darin, dass Texte mit einem geringeren Umfang (bis einschließlich 17 Wörter) nahezu um eine Ratingstufe schlechter bewertet werden. Der Mittelwert des Gesamteindrucks nimmt beispielsweise in der Stichprobe bis einschließlich 17 Wörter einen Wert von 2.0 ($SD = .67$) ein, während er ab 17

Wörter bei 3.0 ($SD = .81$) liegt¹¹. Derartig deutliche Unterschiede werden vor allem auch bei der Textidee und -entfaltung ($M = 2.1$; $SD = .78$ vs. $M = 3.3$; $SD = .76$), den Kohäsionsmitteln ($M = 1.8$; $SD = .83$ vs. $M = 2.9$; $SD = .94$) und dem Wortschatz ($M = 2.3$; $SD = .68$ vs. $M = 3.3$; $SD = .60$) sichtbar¹².

Diese Analysen verdeutlichen, dass die Untersuchungen zur Textqualität bei Schreibprodukten aus der ersten Klasse nicht unabhängig von der Textlänge zu betrachten sind. Es bleibt aber die Frage offen, ob Schüler mit geringer ausgebildeten Schreibfähigkeiten kürzere Texte verfassen und diese tatsächlich eine geringere Qualität aufweisen oder ob die Textlänge gewissermaßen als externer Faktor einfließt und somit kürzere Texte an sich (trotz vergleichbarer Qualität) schlechter bewertet werden.

9. Diskussion und Ausblick

Bereits Texte von Schreibanfängern können einer detaillierten inhaltlichen Analyse unterzogen werden und es lässt sich Varianz erkennen. Die Befunde verdeutlichen im Einklang mit einer explorativen Studie in zweiten Klassen (Bachmann/Becker-Mrotzek 2010), dass es Schreibanfängern gelingen kann, funktional-pragmatische Schreibfähigkeiten zu realisieren. Auch lässt sich nicht bestätigen, dass die Aufmerksamkeit und damit viele Ressourcen der Schreibanfänger ausschließlich an die Schreibungen, die entweder aus dem Gedächtnis abgerufen oder lautanalytisch abgeleitet werden müssen, gebunden sind (Ellis 1982), so dass es ihnen weniger gelingt, auf einer übergeordneten Ebene eine inhaltlich wie sprachlich kohärente Textwelt aufzubauen. Um dies fundiert nachweisen zu können, sind allerdings die unterrichtlichen Merkmale sowie die Lernvoraussetzungen der Schüler und andere Kontextmerkmale zu erheben.

Eine Begrenzung der vorliegenden Analysen besteht in der aus dem PERLE-Projekt stammenden selektiven Stichprobe (Lipowsky et al. 2013), weshalb die berichteten Befunde grundsätzlich nur eingeschränkt auf konventionelle Grundschulen übertragbar sind. Da sich aber bereits bei der vorliegenden relativ homogenen Stichprobe deutliche Unterschiede auf Klassenebene abzeichnen, dürfte das bei einer heterogeneren Stichprobe umso eher der Fall sein.

In Bezug auf die empirische Bestimmung von Textqualität lassen die berichteten Ergebnisse den Schluss zu, dass sich Schülertexte im Anfangsunterricht sowohl mit holistischen als auch analytischen Verfahren reliabel und objektiv auswerten lassen. Analysen zu ausgewählten Stichproben mit unterschiedlichen Textlängen (bis einschließlich 17 Wörter und ab 17 Wörter)

¹¹ Die Mittelwertsunterschiede sind signifikant ($p < .05$).

¹² Die Mittelwertsunterschiede sind signifikant ($p < .05$).

deuten auf eine Konsistenz der Raterurteile hin, was vor dem Hintergrund der insgesamt marginalen Unterschiede im Umfang der Texte allerdings mit Vorsicht interpretiert werden sollte. Vergleicht man die Qualitätseinschätzungen der beiden Stichproben miteinander, zeichnet sich ab, dass Texte mit geringerem Umfang (bis einschließlich 17 Wörter) als schwächer eingeschätzt werden. Dabei kann jedoch die Beziehung zwischen Textlänge und Textqualität statistisch nicht geklärt werden. Aussagen darüber, ob Schreibprodukte mit einem geringeren Umfang tatsächlich auf eine schwächere Schreibfähigkeit hinweisen oder ob sie unter dem Einfluss der Textlänge schlechter bewertet werden, können anhand der vorliegenden Daten nicht getroffen werden.

Für die Schreibdidaktik lässt sich aus diesen ersten Befunden trotz des Vorbehalts der selektiven Stichprobe und der Problematik der Textlänge schließen, dass auch im Anfangsunterricht bei herausfordernden Schreibaufgaben Schüler auch funktional-pragmatische Schreibfähigkeiten (Kohärenz, Berücksichtigung des Adressaten) zeigen können, so dass auch komplexe Aufgaben neben sinnvollen, isolierten Übungen zum Erwerb orthografischer Schreibfähigkeiten ihre Berechtigung haben (vgl. auch Bachmann/Becker-Mrotzek 2010).

Literaturverzeichnis

- Augst, Gerhard/Disselhoff, Katrin/Henrich, Alexandra/Pohl, Thorsten/Völzing, Paul.-Ludwig. (2007): Text - Sorten - Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter. Frankfurt a. M.: Lang.
- Bachmann, Thomas/Ospelt-Geiger, Barbara/Ospelt, Kathrin/Vital, Nathalie (2007): Aufgaben mit Profil. Frühe Förderung funktional-pragmatischer Schreibfähigkeiten. Unveröffentlichter Schlussbericht zum internen Projekt der Pädagogischen Hochschule Zürich.
- Bachmann, Thomas/Becker-Mrotzek, Michael (2010): Schreibaufgaben situieren und profilieren. In: Pohl, Thorsten/Steinhoff, Torsten (Hrsg.): Textformen als Lernformen. Duisburg: Gilles & Francke. S. 191–1210.
- Bachmann, Thomas (2012): „Schriftliches Instruieren in der 2. Klasse - das geht ganz sicher nicht!“. Frühe Förderung hierarchiehöherer Schreibfähigkeiten auf der Unterstufe. http://www.forumlecture.ch/myUploadData/files/2012_2_Bachmann.pdf. Abgerufen am 25.02.2015.
- Baurmann, Jürgen (2006): Schreiben - Überarbeiten - Beurteilen. Ein Arbeitsbuch zur Schreibdidaktik. Seelze (Velber): Kallmeyer.
- Becker-Mrotzek, Michael (1997): Schreibentwicklung und Textproduktion. Opladen: Westdeutscher Verlag GmbH.

- Becker-Mrotzek, Michael/Schindler, Kirsten (2007): Schreibkompetenz modellieren. In: Becker-Mrotzek, Michael/Schindler, Kirsten (Hrsg.): Texte schreiben. Duisburg: Gilles & Francke Verlag. S. 7-26.
- Becker-Mrotzek, Michael/Schindler, Kirsten (2008): Schreibkompetenz modellieren, entwickeln und testen. In: Didaktik Deutsch. Sonderheft 2. S. 94-106.
- Becker-Mrotzek, Michael/Böttcher, Ingrid (2011): Schreibkompetenz entwickeln und beurteilen. Praxishandbuch für die Sekundarstufe I und II. Berlin: Cornelsen-Scriptor.
- Bereiter, Carl/Scardamalia, Marlene (1987): The psychology of written composition. Hillsdale NJ: Erlbaum.
- Blatt, Inge/Ramm, Gesa/Voss, Andreas (2009): Modellierung und Messung der Textkompetenz im Rahmen einer Lernstandserhebung in Klasse 6. In: Didaktik Deutsch. H. 26. S. 54-81.
- Böhme, Katrin/Bremerich-Vos, Albert/Robitzsch, Alexander (2009): Aspekte der Kodierung von Schreibaufgaben. In: Granzer, Dietlinde/Köller, Olaf/Bremerich-Vos, Albert/van den Heuvel-Panhuizen, Marja./Reiss, Kristina./Walther, Gerd (Hrsg.): Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule. Weinheim: Beltz. S. 290-329.
- Böttcher, Ingrid/Becker-Mrotzek, Michael (2003): Texte bearbeiten, bewerten und benoten. Schreibdidaktische Grundlagen und unterrichtspraktische Anregungen. Berlin: Cornelsen Scriptor.
- Böttcher, Ingrid (2010): Kreatives Schreiben. Grundlagen und Methoden; Beispiele, Vorschläge, Projekte; ab Jahrgangsstufe 2. Berlin: Cornelsen Scriptor.
- Bremerich-Vos, Albert/Possmayer, Miriam (2011): Zur Reliabilität eines Modells der Entwicklung von Textkompetenz im Grundschulalter. In: Didaktik Deutsch. H. 31. S. 30-49.
- Clausen, Marten/Reusser, Kurt/Klieme, Eckhard (2003): Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen. Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. In: Unterrichtswissenschaft. H. 31. S. 122-141.
- Doucet, Sharon Arms/Wilsdorf, Anne (2005): Lucy rettet Mama Krokodil. Hamburg: Oetinger.
- Ehlich, Konrad (1983): Text und sprachliches Handeln. Die Entstehung von Texten aus dem Bedürfnis nach Überlieferung. In: Assmann, Aleida/Assmann, Jan/Hardmeier, Christof (Hrsg.): Schrift und Gedächtnis. München: Fink. S. 24-43.
- Ellis, Andrew W. (1982): Spelling and writing (and reading and speaking). In: Ellis, Andrew W. (Ed.): Normality and pathology in cognitive functions. New York: Academic Press. pp. 113-146.
- Feilke, Helmut/Augst, Gerhard (1989): Zur Ontogenese der Schreibkompetenz. In: Antos, Gerd/Krings, Hans Peter (Hrsg.): Textproduktion. Ein interdisziplinärer Forschungsüberblick. Tübingen: M. Niemeyer. S. 297-327.

- Feilke, Helmuth (1995): Auf dem Weg zum Text. Die Entwicklung der Textkompetenz im Grundschulalter. In: Augst, Gerhard (Hrsg.): Frühes Schreiben. Studien zur Ontogenese der Literalität. Essen: Die Blaue Eule. S. 69-88.
- Feilke, Helmuth (1996): Die Entwicklung der Schreibfähigkeiten. In: Günther, Hartmut/Ludwig, Otto (Hrsg.): Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung. Berlin u. New York: Walter de Gruyter. S. 1178-1191.
- Fix, Martin (2008): Texte schreiben. Schreibprozesse im Deutschunterricht. Paderborn: Schöningh.
- Fritzsche, Joachim (1989): Schreibwerkstatt. Geschichten und Gedichte: Schreibaufgaben, -übungen, -spiele. Stuttgart: Klett.
- Grzesik, Jürgen/Fischer, Michael (1984): Was leisten Kriterien für die Aufsatzbeurteilung? Theoretische, empirische und praktische Aspekte des Gebrauchs von Kriterien und der Mehrfachbeurteilung nach globalem Ersteindruck. Opladen: Westdt. Verlag.
- Harsch, Claudia/Neumann, Astrid/Lehmann, Rainer H./Schröder, Konrad (2007): Schreibfähigkeit. In: Beck, Bärbel/Klieme, Eckhard (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messung; DESI-Studie (Deutsch-Englisch-Schülerleistungen-International). Weinheim: Beltz. S. 42-62.
- Knopp, Matthias/Jost, Jörg/Natchwei, Nicole/Becker-Mrotzek, Michael/Grabowski, Joachim (2012): Teilkomponenten von Schreibkompetenzen untersuchen. Bericht aus einem interdisziplinären empirischen Projekt. In: Bayrhuber, Horst/Harm, Ute/Muszynski, Bernhard/Ralle, Bernd/Rothgangel, Martin/Schön, Lutz-Helmut/Vollmer, Helmut J./Weigand, Hans-Georg (Hrsg.): Formate fachdidaktischer Forschung. Empirische Projekte - historische Analysen - theoretische Grundlegungen. Münster: Waxmann. S. 47-56.
- Kötter, Ludwig/Grau, Uwe (1965): Zur Bedingtheit der uneinheitlichen Benotung von Schüleraufsätzen (Nacherzählung). In: Zeitschrift für experimentelle und angewandte Psychologie. H. 12. S. 278-301.
- Kruse, Norbert/Reichardt, Anke/Herrmann, Maik/Heinzel, Friederike/Lipowsky, Frank (2012): Zur Qualität von Kindertexten. Entwicklung eines Bewertungsinstrumentes in der Grundschule. In: Didaktik Deutsch. H. 32. S. 87-110.
- Kürzinger, Anja/Lotz, Miriam/Gleich, Ann-Kathrin/Kempton, Iris (2013): Auswertung der Lucybriefe. Perspektivenübernahme und Schreibkompetenz. In: Lotz, Miriam/Lipowsky, Frank/Faust, Gabriele (Hrsg.): Technischer Bericht zu den PERLE-Videostudien (S. 25-296). In: Lipowsky, Frank/Faust, Gabriele (Hrsg.): Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 3. Bd. 23/3. Frankfurt a. M.: DIPF.
- Kürzinger, Anja/Pohlmann-Rother, Sanna (2014): Kriterienkatalog Textkorpus 1 – Ein Instrument zur Bestimmung von Textqualität in Klasse 1. Bamberg: University of Bamberg Press.

- Lipowsky, Frank/Faust, Gabriele/Kastens, Claudia (2013): Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren. Münster: Waxmann.
- Lotz, Miriam/Berner, Nicole E./Gabriel, Katrin. (2013): Auswertung der PERLE-Videostudien und Überblick über die Beobachtungsinstrumente. In: Lotz, Miriam/Lipowsky, Frank/Faust, Gabriele (Hrsg.): Technischer Bericht zu den PERLE-Videostudien (S. 83-104). In: Lipowsky, Frank/Faust, Gabriele (Hrsg.): Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 3. Bd. 23/3. Frankfurt a. M.: DIPF.
- May, Peter (2001): Lernförderlicher Unterricht. Untersuchung zur Wirksamkeit von Unterricht und Förderunterricht für den schriftsprachlichen Lernerfolg. Frankfurt a. M.: Peter Lang.
- Neumann, Astrid (2007): Briefe schreiben in Klasse 9 und 11. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen. Münster: Waxmann.
- Nussbaumer, Markus/Sieber, Peter (1994): Texte analysieren mit dem Zürcher Textanalyseraster. In: Sieber, Peter (Hrsg.): Sprachfähigkeiten – besser als ihr Ruf und nötiger denn je. Aarau: Sauerländer. S. 141-186.
- Nussbaumer, Markus/Sieber, Peter (1995): Über Textqualitäten reden lernen - z. B. anhand des „Zürcher Textanalyserasters“. In: Diskussion Deutsch. H. 141. S. 36-52.
- Persky, Hilary R./Daane, Mary C./Jin, Ying (2003): The nation's report card. Writing 2002 (NCES 2003-529). <http://nces.ed.gov/nationsreportcard/pdf/main2002/2003529.pdf>. Abgerufen am 25.02.2015.
- Pohl, Thorsten/Steinhoff, Torsten (2010): Textformen als Lernformen. In: Pohl, Thorsten/Steinhoff, Torsten (Hrsg.): Textformen als Lernformen. Duisburg: Gilles & Francke. S. 5-26.
- Schulte, Kerstin (2000): Kohärenz. Eine Untersuchung an Texten aus Klasse 1. Frankfurt a. M.: P. Lang.
- Steinhoff, Torsten (2010): Differenzierte Schülertextbeurteilung. Entwicklungs-, Prozess- und Situierungsdimension. In: Pohl, Thorsten/Steinhoff, Torsten (Hrsg.): Textformen als Lernformen. Duisburg: Gilles & Francke. S. 257-280.
- Steinig, Wolfgang/Betz, Dirk/Geider, Franz Josef/Herbold, Andreas (2009): Schreiben von Kindern im diachronen Vergleich. Texte von Viertklässlern aus den Jahren 1972 und 2002. Münster: Waxmann Verlag.
- Thonke, Franziska/Groß Ophoff, Jana/Hosenfeld, Ingmar/Isaac, Kevin (2008): Kriteriengestützte Erfassung von Schreibleistungen im Projekt VERA. In: Hofmann, Bernhard/Valtin, Renate (Hrsg.): Checkpoint Literacy. Tagungsband 2 zum 15. Europäischen Lesekongress in Berlin. Berlin: Deutsche Gesellschaft für Lesen und Schreiben. S. 28-35.

Weinhold, Swantje (2000): Text als Herausforderung. Zur Textkompetenz am Schulanfang. Mit 296 Schülertexten aus Klasse1. Freiburg i. B.: Fillibach.

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: Buske.
- 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.

Aufgabenstellung¹ als zentrale Gelenkstelle im Schreibunterricht der ersten Klasse – Eine videogestützte Analyse²

Zusammenfassung

Im Beitrag wird der Frage nachgegangen, wie sich die Aufgabenstellung als zentrale Gelenkstelle im Schreibunterricht der ersten Klasse nach empirischen Gütekriterien erheben und charakterisieren lässt. Die Analysen stammen aus dem DFG-Projekt NaSch1 (Narrative Schreibkompetenz in Klasse 1) in dem die Aufgabenstellung im Fach Deutsch in 43 Unterrichtsvideos anhand eines Beobachtungssystems aus 25 fach- und allgemeindidaktischen Kriterien untersucht wird. Vor dem Hintergrund schreibdidaktischer Konzeptionen wird dargestellt, inwieweit sich die beiden Dimensionen „soziale Interaktion“ und „Zielorientierung“ theoretisch bilden und empirisch überprüfen lassen. Insgesamt illustrieren die deskriptiven Befunde, dass der Schreibunterricht allgemeindidaktischen Qualitätskriterien (z. B. Präzisionsgrad) weitgehend genügt, hinsichtlich mancher fachdidaktischer Aspekte (z. B. Lebensweltbezug) aber eine geringere Qualität aufweist.

Schlagworte: Schreibunterricht, Aufgabenstellung, Anfangsunterricht, Videostudie, Unterrichtsqualität

Abstract: This paper focuses on empirical quality criteria of task settings in first grade writing lessons and how these criteria can be measured and characterized. The analyses are taken from the DFG-project NaSch1 (Narrative writing skills in grade 1). NaSch1 examines task

¹ Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift publizierten Originalbeitrag und kann daher nicht zitiert werden. Der Originalbeitrag ist erschienen als: Kürzinger, A., Pohlmann-Rother, S. & Hess, M. (2017). Aufgabenstellung als zentrale Gelenkstelle im Schreibunterricht – Eine videogestützte Analyse. *Zeitschrift für Pädagogik*, 63 (5), 636–656.

² Leitung: Sanna Pohlmann-Rother & Gabriele Faust (†), DFG-Förderkennzeichen PO 1739/1-1. Die Videos wurden im Rahmen der PERLE-Studie (Leitung: Frank Lipowsky und Gabriele Faust; vgl. Lipowsky, Faust & Kastens, 2013) erhoben.

settings in the school subject “German” based on an observational system consisting of 25 subject specific and universal didactical criteria using 43 lesson videos. In view of didactic conceptions it is shown to what extent the two dimensions “social interaction” and “target orientation” can be theoretically described and empirically tested. The descriptive results indicate that writing lessons largely fulfil universal didactical quality criteria (e.g. precision). With respect to subject specific didactical criteria (e.g. integration of the student’s experience realm) writing lessons show a lower quality.

Keywords: Writing Classes, Assignment, Elementary Instruction, Video Study, Quality of Teaching

1. Einleitung

Mit den Bildungsstandards und der damit erforderlichen Orientierung an Kompetenzen erfährt die Debatte um Aufgaben im Unterricht eine neue Resonanz. Während für Mathematik bereits zentrale Befunde zur Aufgabenkonzeption bzw. -qualität vorliegen (vgl. z. B. Baumert, Kunter, Brunner, Krauss, Blum, 2004), ist die Aufgabenkultur im Schreibunterricht bislang kaum untersucht. Insbesondere fehlen empirisch gestützte Analysen zum Zusammenhang zwischen Schreibunterricht und Schreibentwicklung (vgl. Bachmann & Becker-Mrotzek, 2010; Pohl & Steinhoff, 2010).

Eine nicht-repräsentative Pilotstudie von Bachmann, Ospelt-Geiger, Ospelt und Vital (2007), bei welcher der Zusammenhang zwischen Aufgabenstellungen und der Realisierung bestimmter Schreibfähigkeiten von 56 Schreibanfängern/-innen in Klasse 2 analysiert wurde, berichtet erste Trends zur Relevanz von Schreibaufgaben. Ergebnissen der Studie zufolge wirken sich gut strukturierte und tief in soziale Interaktion eingebettete Aufgabenstellungen – also „Aufgaben mit Profil“ – positiv auf die frühe Realisierung anspruchsvoller Schreibfähigkeiten aus. Zudem lassen sich durch derartige Aufgabenstellungen ungünstige Startbedingungen wie etwa Bildungsferne, Geschlecht oder Deutsch als Zweitsprache zu Teilen kompensieren (vgl. Bachmann & Becker-Mrotzek, 2010).

Diese Befunde zeugen von der Bedeutung der Aufgabenkultur für den Schreibunterricht und der Notwendigkeit weiterer empirischer

Forschungen. Im Projekt NaSch1 wird daher erstmalig die Aufgabenstellung im schriftsprachlichen Anfangsunterricht videobasiert analysiert. Im Folgenden werden schreibdidaktische Konzeptionen vorgestellt, aus denen sich didaktisch relevante Aufgabenmerkmale ableiten lassen, die durch allgemeindidaktische Kriterien ergänzt werden.

2. Aufgabenstellung – Eine begriffliche Klärung

Eine „instruktive Systematisierung in Bezug auf Schreibaufgaben“ (Becker-Mrotzek, 2014, S. 492) wird von Baurmann und Pohl (2009) realisiert, die zwischen Schreibauftrag und Schreibarrangement bzw. Schreibsetting unterscheiden. Unter der Bezeichnung Schreibarrangement bzw. Schreibsetting werden situative Schreibfaktoren und Bedingungen verstanden: Schreib- und Lernziel, Schreibbedingungen (z. B. Schreibzeit, Textumfang), Wissensquellen und der Prozess-Status (z. B. Entwurf, Überarbeitung) (Baurmann & Pohl, 2009, S. 99-101). Davon abgegrenzt wird der Schreibauftrag, bei dem Thema, Material und Impuls, Adressat sowie Textfunktion formuliert werden.

Im vorliegenden Beitrag wird das Schreibsetting bzw. -arrangement zusammen mit dem Schreibauftrag übergreifend als Aufgabenstellung bezeichnet. In diesem Kontext wird die Aufgabestellung folglich nicht nur auf einen Schreibauftrag reduziert, sondern schließt auch die Hinführung auf das Schreiben bis zum Beginn der Schreibphase mit ein. Diese relativ weite Definition wird an dem Begriff der Aufgabenkultur verdeutlicht, der die Einbindung von Aufgaben in verschiedene unterrichtliche Kontexte meint (vgl. Bohl & Kleinknecht, 2009).

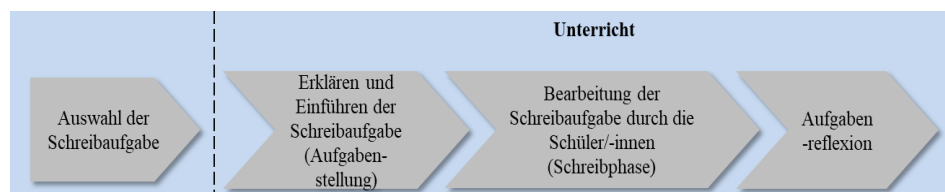


Abbildung 1: Einbindung von Aufgaben im Unterricht (in Anlehnung an Bohl & Kleinknecht, 2009, S. 331)

3. Aufgabenstellungen im schriftsprachlichen Anfangsunterricht

Aus didaktischer Perspektive kommt dem schriftsprachlichen Anfangsunterricht, also der ersten Begegnung der Kinder mit Lesen und Schreiben im institutionellen Kontext Schule, eine wichtige Bedeutung zu. Obwohl Schulanfänger/-innen durch den eigenen Sprachgebrauch, durch das Hören von Geschichten und durch Beobachtungen Erwachsener beim Lesen und Schreiben vielfältige und heterogene Vorerfahrungen mit Schrift und Schriftlichkeit haben, ist für sie das Schreiben- und Lesenlernen eine grundlegend neue Erfahrung. In ihrer sprachlichen Entwicklung befinden sie sich im Übergang von Mündlichkeit zur Schriftlichkeit, der bedeutsamen Einfluss auf ihr Sprachbewusstsein, Denken und Handeln nimmt (Weinhold, 2014, S. 143). In der gegenwärtigen schreibdidaktischen Diskussion besteht dahingehend Einigkeit, dass die Schüler/-innen Schreiben von Beginn an „als sprachlich-konzeptionellen, kommunikativen und kooperativen Prozess erlernen und als Ausdrucksmittel ihrer Gedanken in verschiedenen Funktionen nutzen“ sollen (Weinhold, 2014, S. 144). Mit dem Textschreiben in der Grundschule und damit auch mit der Gestaltung der Aufgabenstellung beim Textschreiben beschäftigen sich verschiedene literatur- und schreibdidaktische Ansätze. Die Konzeption „Schreiben als kulturelle Tätigkeit“ (Dehn, Merklinger & Schüler, 2011) sieht vor, dass Kinder beim Schreiben von Anfang an mit komplexen und inhaltlich bedeutsamen Inhalten konfrontiert werden – sowohl bei der Auswahl der Schreibvorgabe, als auch bei der Formulierung der Aufgabenstellung. Richter (2007), die die Grundschule als Ort der Anregung literar-ästhetischer Prozesse ansieht, hebt ebenfalls hervor, Kindern von Beginn an „komplexe literarische Welten aufzuschließen“ (S. 21), weshalb sie u.a. die Vereinfachung originaler Texte ablehnt. Ein weiterer Ansatz, der aus der grundschulpädagogisch orientierten Forschung zum kreativen Schreiben stammt, betont vorrangig das Freisetzen kreativer Prozesse und fokussiert weniger auf bestimmte Text- und Bildinhalte (vgl. Kohl & Ritter, 2010). Kompetenzorientierte Ansätze der Schreibforschung und -didaktik knüpfen dagegen an die Tradition der amerikanischen kognitiven Schreibforschung an und stellen die Formen und Funktionen des Schreibens sowie ihre Entwicklung in den Fokus (vgl. z. B. Bereiter, 1980; Augst, Disselhoff, Henrich, Pohl & Völzing, 2007).

Während das Formulieren von Schreibaufgaben recht allgemein als „...das genuine Geschäft der Didaktik und das der alltäglichen Konzeption von Unterricht“ (Dehn, 2009, S. 154) verstanden wird, rücken je nach didaktischer Konzeption unterschiedliche Aspekte der Aufgabenstellung in den Fokus. Im Ansatz „Schreiben als kulturelle Tätigkeit“ wird der Fokus „auf das Formulieren des Verhältnisses von Selbst und Welt“ (Dehn et al., 2011, S. 80) gerichtet, wobei Grundschul Kinder aus verschiedenen inhaltlichen Vorgaben wählen und für sich etwas thematisieren und formulieren. Dabei soll „die Kontextualisierung der inneren Sprache unmittelbar im Schreiben erfolg(en)“, weshalb es als „Entwertung des literalen Aspekts des Schreibens“ angesehen wird, wenn die Schreibanfänger vor dem Schreiben mündlich sagen, was sie zu Papier bringen wollen (Dehn et al., 2011, S. 100). In Abgrenzung zu diesem Ansatz orientiert sich die Aufgabenstellung im Projekt NaSch1 am kompetenzorientierten Ansatz, in dem es darum geht, die Lerner herauszufordern, bestimmte Textsorten zu realisieren und die Aufmerksamkeit auf bestimmte Textfunktionen zu richten (z. B. Berücksichtigung eines Adressaten bei der Textsorte Brief). Die mündliche Sammlung von Schreibideen vor dem Schreiben wird nicht als hinderlich, sondern insbesondere für leistungsschwächere Kinder als Strukturierungshilfe und daher als lernförderlich erachtet.

Im Sinne der kompetenzorientierten Schreibforschung und -didaktik bezeichnen Bachmann und Becker-Mrotzek (2010, S. 191) klar konturierte und profilierte Schreibaufgaben, die für die Lerner in einem deutlich erkennbaren und nachvollziehbaren Handlungszusammenhang stehen bzw. einen solchen abbilden als „Aufgaben mit Profil“, die als besonders bedeutsam für das Lernen erachtet werden. Bedingungen von „Aufgaben mit Profil“ bestehen in der identifizierbaren Funktion von Texten (Ziel und Adressat/-in sollen erkennbar sein), in Gelegenheiten für Schüler/-innen, sich das für das Schreiben nötige Weltwissen anzueignen, in der Möglichkeit, die Texte im Rahmen einer sozialen Interaktion zu verfassen und die Wirkung des Textes von den Schreibern testen zu lassen (Bachmann & Becker-Mrotzek, 2010, S. 195). Damit wird auf die Bedeutung der Schreibfunktion bzw. Zielorientierung und der sozialen Interaktion beim Verfassen eines Texts verwiesen, die auch in weiteren Konzeptionen zum Ausdruck gebracht werden (vgl. z. B. Abraham & Kupfer-Schreiner, 2007; Blömeke, Risse, Müller, Eichler, Schulz, 2006). Im vorliegenden Kontext wird unter Zielorientierung die

Definition der Kommunikationssituation verstanden, bei der der Zweck des Schreibens, die Autor/-innen, Rezipient/-innen und die Textart von der Lehrkraft expliziert werden (vgl. Schneuwly, 1996). Diese Informationen sind für die erfolgreiche Textproduktion der Lernenden zwingend notwendig, um sicherzustellen, dass die Kinder die Funktion der Schreibaufgabe nachvollziehen können: „Viele Schülerinnen und Schüler scheitern nicht deshalb an einer Schreibaufgabe, weil sie sie nicht lösen können, sondern weil sie nicht richtig verstanden haben, worum es bei der Aufgabe geht“ (Abraham & Kupfer-Schreiner, 2007, S. 32).

Anhand von sozialen Interaktionen kann Schreibanfänger/-innen die Gelegenheit geboten werden, miteinander zu kooperieren, zu zweit oder in Kleingruppen Schreibideen zu entwickeln und ihre Ideen aufeinander zu beziehen, was auch von Dehn (2007, S. 14) herausgestellt wird. Didaktisch-methodisch lässt sich dies durch gezielte Fragen durch die Lehrperson, Bilder oder (Hand-)Puppen literarischer Figuren sowie durch die Inszenierung von Rollenspielen gestalten. Durch den Austausch mit Anderen kann u. a. Vorwissen aktiviert werden und es können Ideen expliziert und kognitive Modelle bereitgestellt werden (vgl. z. B. Slavin, 1996; Dann, Diegritz & Rosenbusch, 1999; Blömeke et al., 2006). Die Bereitstellung von sozialen Interaktionen zielt auch darauf ab, die Trennung von Empfänger und Sender beim Schreiben, also die „Zerdehnung der Sprechsituation“ (Ehlich, 1983, S. 38) zu überwinden, die insbesondere jüngeren und ungeübten Schreiber/-innen Probleme bereitet. Statt eines sichtbaren Gesprächspartners, mit dem man in direkten Dialog tritt, besteht beim Schreiben die Notwendigkeit, sich einen Leser bzw. eine Leserin vorzustellen. Dadurch, dass sich die soziale Kognition noch entwickelt, bedeutet dies für die Schüler/-innen eine große Herausforderung (Bachmann & Becker-Mrotzek, 2010, S. 205).

Weitere Aspekte lernförderlicher Aufgaben sind beispielsweise der Lebensweltbezug bzw. die Authentizität (vgl. Maier, Bohl, Kleinknecht & Metz, 2013) sowie das Potenzial zur inneren Differenzierung (vgl. Blömeke et al., 2006). Im Rahmen eines handlungs- und produktionsorientierten Unterrichts lassen sich Texte und Bilder verwenden, die Probleme oder Erlebnisse aus der Lebenswelt der Schüler/-innen thematisieren. Das Problem von Anderssein und Ausgrenzung – wie es in dem Buch „Lucy rettet Mama Kroko“ (Doucet & Wilsdorf, 2005), das im vorliegenden Projekt als

Schreibvorlage diente, dargestellt wird – ist der Lebenswelt der Schulanfänger/-innen zugänglich. Die Repräsentation einer authentischen Situation sowie die Differenzierungsmöglichkeit sind eng mit der Schreibaufgabe verbunden, lassen sich aber auch gezielt im Unterricht vermitteln und explizit thematisieren. Insbesondere in der Diskussion über Unterrichtsqualität sind im Zusammenhang mit Motivation Konzepte wie Authentizität, Verknüpfungen mit dem Alltag sowie der Lebenswelt von enormer Bedeutung (vgl. Helmke, 2009).

Aus den oben genannten Konzeptionen und insbesondere aus der Auseinandersetzung mit der kompetenzorientierten Schreibdidaktik lassen sich Kriterien guter Schreibaufgaben und Aufgabenstellungen ableiten, wie z. B. der Lebensweltbezug, die mündliche Sammlung möglicher Schreibinhalte, sowie die Besprechung textsortenspezifischer Merkmale oder die Thematisierung der Schreibfunktion. Diese Merkmale können durch allgemeindidaktische Unterrichtsqualitätsaspekte ergänzt werden. Vor allem die Verständlichkeit und der Präzisionsgrad sind für die Qualität der Aufgabestellung bedeutsam – Kriterien, die häufig mit den Unterrichtsmerkmalen Klarheit und Strukturiertheit in Relation gesetzt werden (vgl. z. B. Helmke, 2009; Helmke, Helmke & Schrader, 2007). Verständlichkeit im vorliegenden Kontext bezieht sich auf eine den Schulanfängern/-innen angepasste, unkomplizierte Sprache ohne Wiederholungen und Fremdwörter. So werden beispielsweise Einfachheit, Kürze und Prägnanz, Ordnung und Gliederung sowie zusätzliche Stimulanz als sprachlich verständlich betrachtet (vgl. Langer, Schulz von Thun & Tausch, 1974; auch Lotz, 2016). Während die Verständlichkeit dementsprechend eher auf die akustische und sprachliche Komponente von Klarheit³ abzielt, wird anhand des Präzisionsgrades die inhaltliche Klarheit erfasst. Diese beiden Aspekte manifestieren sich auch in der Zielorientierung, indem sie auf die Notwendigkeit einer präzisen und verständlichen Sprache aufmerksam machen, in der die Informationen über den Zweck des Schreibens, die Autor/-innen, Rezipient/-innen und Textart offeriert werden sollen.

Neben der Verständlichkeit, Strukturiertheit und Klarheit des Unterrichts, der Passung von Unterrichtsangeboten und Lernvoraussetzungen (vgl. Helmke, 2004) wirkt sich vermutlich auch die Klassenführung (vgl. Helmke & Schrader, 2010) auf die Qualität der Aufgabenstellung aus. Die Implementierung der Schreibaufgabe nimmt oftmals eine längere

³ Helmke (2009) unterscheidet vier Komponenten von Klarheit: akustisch, sprachlich, inhaltlich und fachlich.

Zeit in Anspruch, ist in verschiedene thematische Phasen gegliedert (Einführung, Sammeln von Ideen, Schreibauftrag) und zum Teil mit dem Wechsel von Sozialformen verbunden, was eine sinnvolle Organisation und eine effektive Vorbereitung des Schreibarrangements erfordern.

Zusammenfassend lässt sich aus der bisherigen Darstellung ableiten, dass soziale Interaktionen und Zielorientierung zentrale Kriterien für Aufgabenstellungen verkörpern. Im Projekt NaSch1 wird daher unter Berücksichtigung weiterer allgemein- und schreibdidaktischer Merkmale auf die Umsetzung der Zielorientierung und sozialen Interaktion während der Aufgabenstellung fokussiert. Dabei wird eine detaillierte Operationalisierung dieser beiden Kriterien anvisiert, da sie verschiedene Aspekte beinhalten und zu ihrer Erfassung bislang eher vage Vorschläge vorliegen. Sowohl für die Zielorientierung als auch für die soziale Interaktion werden verschiedene Kriterien entwickelt, die zu je einer Dimension⁴ zusammengefasst werden sollen.

Für den vorliegenden Beitrag sind folgende Fragestellungen leitend:

- Wie lässt sich die Aufgabenstellung im schriftsprachlichen Anfangsunterricht anhand von allgemein- und schreibdidaktischen Kriterien erfassen und charakterisieren?
- Inwieweit lassen sich Dimensionen der sozialen Interaktion und Zielorientierung identifizieren?

4. Methodik

4.1 Stichprobe

Im Rahmen der PERLE-Videostudie wurden im März der Klassenstufe 1 in 37 Klassen jeweils ca. 90-minütige Unterrichtseinheiten im Fach Deutsch videografiert (vgl. Lotz & Corvacho del Toro, 2013). Die Stichprobe im Videomodul Deutsch setzt sich sowohl aus staatlichen als auch aus privaten Grundschulen, sogenannten BIP-Kreativitätsgrundschulen (vgl. Mehlhorn & Mehlhorn, 2003), zusammen. Durch die Besonderheit des überwiegend praktizierten Team-Teaching an den privaten Schulen, in dessen Kontext die beiden Lehrkräfte die

⁴ Die Verwendung des Begriffs „Dimension“ ist in diesem Kontext nicht methodisch (in Bezug auf Faktorenanalysen), sondern rein inhaltlich konnotiert. Unter einer Dimension wird demnach die Zusammenfassung verschiedener Einzelkriterien verstanden.

Klassen zeitweise in zwei Teilgruppen in unterschiedlichen Räumen aufteilen, ergeben sich insgesamt 49 Unterrichtsvideos.

Von diesen 49 Videos werden für die schreibdidaktische Unterrichtsanalyse 43 Aufnahmen herangezogen⁵. Diese Stichprobe umfasst 516 Schüler/-innen, wovon 44 % männlich und 56% weiblich sind. 58,9 % der Schüler/-innen besuchten eine staatliche Schule, 41,1 % eine private BIP-Kreativitätsgrundschule. Zum Zeitpunkt der Einschulung betrug das Alter der Kinder im Mittel 6 Jahre und 7 Monate ($SD = 4.8$ Monate).

4.2 Schreibdidaktisches Setting

Das schreibdidaktische Setting ist durch zwei Vorgaben gekennzeichnet, mit denen die Lehrkräfte in den teilnehmenden Klassen beauftragt wurden. So sollte zum einen der Anfangsteil des fantastischen Bilderbuchs „Lucy rettet Mama Krok“ (Doucet & Wilsdorf, 2005) vorgestellt werden und zum anderen von den Schülern/-innen ein Brief aus der Perspektive der literarischen Hauptfigur Lucy an eine weitere literarische Figur (Krokodilmutter „Mama Krok“) verfasst werden. Charakteristisch für dieses Schreibarrangement ist damit ein handlungs- und produktionsorientiertes Verfahren, in welchem die Texte als Anschlussaufgabe an die rezipierte Geschichte konzipiert wurden. In diesem Kontext fordert die Schreibaufgabe die Kinder heraus, sich mit dem Bilderbuch gedanklich auseinanderzusetzen und auf Basis der Vorgabe eigene Schreibideen zu kreieren (vgl. Kürzinger & Pohlmann-Rother, 2015).

Die Orientierung an einen Adressaten bzw. eine Adressatin erfordert bei den Schüler/-innen u.a. Empathie und das Vermögen, sich sowohl in die literarische Figur (Lucy) als auch in die fiktive Adressatin (Mama Krok) hineinzusetzen. Diese Anforderungen an die Schreibanfänger/-innen lassen sich möglicherweise durch die situative Einbettung der Aufgabe in einen fiktiven Handlungsrahmen, den handlungs- und produktionsorientierten Literaturunterricht und durch die Verwendung der Textsorte Brief, die mit ihrem formalisierten Aufbau eine Orientierungshilfe darstellen kann, relativieren (vgl. Kürzinger & Pohlmann-Rother, 2015).

⁵ Sechs Videos scheiden aus, da teilweise in den BIP-Klassen die Aufgabenstellung von beiden Lehrkräften gemeinsam erklärt wurde und die Klassen lediglich für das Verfassen der Briefe geteilt wurden.

4.3 Modellierung des Instruments zur Bewertung der Aufgabenstellung

Ausgehend von den unter Kapitel 3 spezifizierten Annahmen wird im Projekt NaSch1 ein Modell zur Analyse der Qualität der Aufgabenstellung⁶ im Anfangsunterricht zugrunde gelegt, das didaktische und fachübergreifende Merkmale kombiniert und zwischen Schreibauftrag und Schreibarrangement (vgl. Baumann & Pohl, 2009) unterscheidet.

Um das schreibdidaktische Setting videogestützt zu erfassen, wurde ein Beobachtungssystem mit 25 niedrig, mittel und hoch inferenten Einzelkriterien konzipiert. Dabei wurde in Anlehnung an z. B. Hugener (2006) ein deduktiv-induktives Verfahren kombiniert, in welchem die Entwicklung der Kriterien sowohl auf Grundlage der Theorie zur Aufgabenstellung sowie vorhandener Instrumente als auch anhand von zufällig ausgewählten Teilstichproben der Unterrichtsvideos basiert. Für die deduktive Entwicklung wurden bestehende schreibdidaktische bzw. fachübergreifende Kriterien lernförderlicher Schreibsettings adaptiert (siehe Kap. 3). Zu den schreibdidaktischen Kriterien gehören beispielsweise die Dauer der Aufgabenstellung, Differenzierungsmaßnahmen, der Lebensweltbezug, soziale Interaktion, Zielorientierung, Schreibstrategien und die Vermittlung textsortenspezifischer Merkmale. Fachübergreifende Analysekriterien der Unterrichtsqualität bilden die vier Merkmale Organisation der Aufgabenstellung, Kohärenz und Fokussierung, der Präzisionsgrad sowie die Verständlichkeit des Schreibauftrags ab.

Alle Kriterien für die Analyse der Aufgabenstellung sind einzeln und unabhängig voneinander formuliert und unterscheiden sich je nach fachübergreifender und schreibdidaktischer Zielsetzung. Davon ausgenommen sind Einzelkriterien, die sich unter der Zielorientierung bzw. unter der sozialen Interaktion einordnen lassen. Die Zielorientierung setzt sich aus den Einzelkriterien Schreibfunktion, Präzision, Verständlichkeit sowie der Ideensammlung zum Abschiedsthema zusammen, womit schreibdidaktische und allgemeine Kriterien der Unterrichtsqualität kombiniert werden (Tabelle 1). Um zu erfassen, inwieweit die Aufgabenstellung in eine soziale Interaktion

⁶ Das vollständige Verzeichnis aller Kriterien sowie die präzisen Kodier- und Ratinganleitungen können dem Manual zur Auswertung der Schreibaufgabe entnommen werden (Kürzinger & Pohlmann-Rother, in Vorb.).

eingebettet ist, wird drei Leitfragen nachgegangen: Findet ein Austausch über Schreibideen statt? Ist eine soziale Interaktion im engeren Sinn erkennbar (z. B. szenisches Spiel)? Wird ein Wechsel der Sozialform (Partner- oder Gruppenarbeit) angeregt, der Möglichkeiten des Austausches bietet? In Tabelle 2 wird die Dimension „soziale Interaktion“ detailliert vorgestellt.

Tabelle 1: Dimension Zielorientierung

Item-bezeichnung	Grundidee	Ankerbeispiele	Inferenz/ Aufbau
Schreibfunktion Quelle: Eigenentwicklung	Inwiefern wird die Schreibfunktion von der Lehrkraft vermittelt? Bei der vorliegenden Schreibaufgabe besteht das Ziel darin, dass die Schüler als Lucy Mama Kroko über den Abschied informieren.	„L: Jetzt könnt ihr einen Brief an Mama Kroko schreiben, als Lucy die Krokodilfamilie verlässt. Mit dem Brief verabschiedet ihr euch von Mama Kroko, dann macht sie sich keine Sorgen.“ (Wert 3)	Mittel inferent; vierstufig, keine Besprechung (0) bis ausführliche Besprechung des Abschieds (3)
Präzisionsgrad des Schreibauftrags Quelle: Lotz (2016)	Inwieweit ist die Instruktion des Schreibauftrags klar formuliert und trägt zum Verständnis bei? Die Präzision wird bezüglich Textsorte Brief, Adressat, Sozialform und der zu nutzenden Materialien und Hilfsmittel erfasst.	„L: Wir haben jetzt viele Ideen gesammelt, weshalb Lucy wegschwimmt. Ich teile euch nun Arbeitsblätter aus und ihr schreibt an Mama Kroko“ (Wert 1)	Hoch inferent: vierstufig, sehr unpräzise (1) bis sehr präzise (4)
Verständlichkeit des Schreibauftrags Quelle: Lotz (2016)	Inwieweit wird der Arbeitsauftrag verständlich formuliert? Hebt die Lehrkraft wesentliche und besonders wichtige Aspekte durch Betonung hervor und verzichtet sie auf irrelevante und überflüssige Informationen?	„L: So, jetzt sollt ihr mit dem Brief beginnen. Ihr seid jetzt Lucy. An wen sollt ihr schreiben? S3: An Mama Kroko. L: Genau. Und wie beginnt und beendet man einen Brief? Erinnert ihr euch? S5: Liebe Mama Kroko. L: Ja, das ist ein Beispiel. Wie kann man noch beginnen?“	Hoch inferent: vierstufig von sehr unverständlich (1) bis sehr verständlich (4)

S11: Hallo Krokodile.

L: Genau. Und wie beendet ihr euren Brief?

S7: Deine Lucy.

L: Super. Habt ihr alles verstanden? Wer fasst noch einmal zusammen, was ihr zu tun habt?

(Wert 4)

Ideensammlung
zum Abschied

Inwieweit wird der Abschied Lucys von den Krokodilen während der Ideensammlung von den Schülern thematisiert?

„L: Was würdet ihr schreiben? S13: Liebe Mama, ich geh weg.“ (Wert 1)

Niedrig inferent; dichotom: keine Ideensammlung zum Abschied (0); Ideensammlung zum Abschied (1)

Quelle:
Eigenentwicklung

Tabelle 2: Dimension soziale Interaktion

Itembezeichnung	Grundidee	Ankerbeispiele	Inferenz/ Aufbau
Austausch über Schreibideen Quelle: Eigenentwicklung	Sammeln von konkreten Schreibideen von Schülern im Unterrichtsgespräch	„L: Was würdet ihr schreiben? S13: Liebe Mama, ich geh weg, weil mich alle hänseln.“ (Wert 1)	Niedrig inferent; dreistufig: keine Idee (0), eine Idee (1), mehr als eine Idee (2)
Soziale Interaktion im engeren Sinn Quelle: Eigenentwicklung	Bereitstellen von Gelegenheiten zur sozialen Interaktion	„L: Bevor ihr den Brief alleine schreibt, tauscht ihr euch bitte mit eurem Sitznachbarn aus. Besprecht miteinander, was ihr schreiben könntet. Dann schreibt jeder für sich den Brief.“ (Wert 1)	Niedrig inferent; dichotom: keine soziale Interaktion (0), Vorkommen einer sozialen Interaktion (1)
Wechsel der Sozialform Quelle: Eigenentwicklung	Anweisung zur Partner- oder Gruppenarbeit	„S1 und S2 können zusammen den Brief schreiben.“ (Wert 1)	Mittel inferent; dichotom: keine Anweisung zur Schreibkooperation (0), Anweisung zu einer Schreibkooperation (1)

Alle 25 Kriterien der Aufgabenstellung sind mit der entsprechenden Operationalisierung sowie den Rating- und Kodierregeln in einem Manual festgehalten. Für die Auswertung der Aufgabenstellung wurden zwei Beobachterinnen (Studentinnen aus erziehungswissenschaftlichen Disziplinen) in mehrtägigen Trainings geschult, die auf ein gemeinsames theoretisches Verständnis des Beobachtungssystems abzielten. Im Anschluss an die Schulung erfolgte die Analyse des Schreibsettings anhand eines Beobachtungsbogens.

5. Ergebnisdarstellung

5.1 Beurteilerübereinstimmungen und Reliabilitäten

Um Aussagen über die Güte der Kodierung bzw. des Ratings treffen zu können, wurde das Analyseinstrument auf die Übereinstimmung der Raterinnen- bzw. Kodiererinnenurteile überprüft. Die Güte der Einschätzung wird dabei anhand von Kennwerten abgebildet, die sich an denen des PERLE Projekts orientieren (Lotz, Berner & Gabriel, 2013): Sie betragen bei der prozentualen Übereinstimmung ($P\ddot{U}$) $\geq 85.00\%$, bei Cohens Kappa (κ) $\geq .70$ und beim relativen G-Koeffizienten (g) $\geq .70$.

Für alle 25 Kriterien liegen Kennwerte zu den Reliabilitäten bzw. Übereinstimmungen über eine Stichprobe von 42 Videos vor, die restlichen sieben Videos wurden zu Übungszwecken herangezogen und gehen daher nicht in die Reliabilität ein⁷. Darüber hinaus wurden auch Übereinstimmungen der Raterinnen bzw. Kodiererinnen mit der Manualentwicklerin für jedes dritte Video der Stichprobe berechnet, um zu gewährleisten, dass die Beurteiler die Texte tatsächlich auf Basis des Manuals und des gemeinsam entwickelten Verständnisses eingeschätzt haben.

Insgesamt zeigen die Resultate der Beobachterübereinstimmung zufriedenstellende Kennwerte. So liegt der relative G-Koeffizient für die vierstufigen, hoch inferenten Items (Imaginationsförderung, Verständlichkeit des Schreibauftrags, Präzisionsgrad des

⁷ Die sieben Übungsvideos wurden ein weiteres Mal analysiert und gehen damit in die Auswertungsstichprobe ein, werden aber nicht zur Berechnung der Beobachterübereinstimmung herangezogen. Die Stichprobe zur Überprüfung der Beobachterübereinstimmung ist damit unabhängig von der Auswertungsstichprobe. Diese bestand aus 49 Videos, von denen sich nach der Analyse herausgestellt hat, dass nur 43 Videos verwendet werden können.

Schreibauftrags, Ablauf und Organisation des schreibdidaktischen Settings sowie Kohärenz und Fokussierung) mit Werten von .87 bis .98 deutlich über dem Richtwert und verweist auf eine hohe Reliabilität zwischen den Raterinnen. Bei den restlichen niedrig und mittel inferenten Items werden ebenfalls Übereinstimmungen über den Mindestkennwerten erreicht. Die prozentuale Übereinstimmung zwischen den beiden Raterinnen umfasst Kennwerte zwischen 88.09 % und 100 %, während die Werte für Cohens Kappa zwischen $\kappa = .84$ und 1.0 liegen.

5.2 Deskriptive Ergebnisse zur Aufgabenstellung im Anfangsunterricht

Ausgewählte allgemein- und schreibdidaktische Einzelkriterien

Dauer

Als interessanten Aspekt ist zunächst der zeitliche Umfang der Aufgabenstellung zu nennen, der nicht vorgegeben, sondern von den Lehrkräften frei gewählt wurde. Insgesamt beträgt die Dauer der Aufgabenstellung im Mittel 8.3 Minuten ($SD = 0.70$; $N = 43$), wobei sich Abweichungen innerhalb der Stichprobe erkennen lassen. So umfasst die kürzeste Aufgabenstellung 2.2 Minuten, die längste 22.1 Minuten. Diese Zeitangaben können jedoch nicht ohne weitere Kontextvariablen interpretiert werden. Es kann davon ausgegangen werden, dass beispielsweise die Intensität der Bilderbuchrezeption einen entscheidenden Einfluss auf die zeitliche Phase der Aufgabenstellung nimmt. Sofern bereits während der Besprechung des Buchs Zusammenhänge erläutert und verschiedene Themen angesprochen wurden (z. B. Abschied; Lebensweltbezug), wird die Aufgabenstellung eher auf den reinen Schreibauftrag beschränkt sein und relativ kurz ausfallen.⁸

Differenzierung

Ähnlich wie der kaum vorhandene Wechsel der Sozialform (z. B. kooperatives Schreiben) wird im vorliegenden Schreibunterricht auch selten differenziert. Insgesamt werden quantitative Differenzierungsmöglichkeiten (Unterschiede im Briefumfang) nur

⁸ Die Bilderbuchrezeption wurde im Projekt PERLE ausgewertet (Kruse, Gabriel & Faust, 2013) und wird in weiteren Analysen berücksichtigt.

einmal, qualitative Differenzierungen (z. B. inhaltliche Modifikation: Einzelne Schüler/-innen verfassen keinen Brief, sondern schreiben Lernwörter von der Tafel ab) fünfmal von den Lehrkräften realisiert.

Lebensweltbezug

Auch der Lebensweltbezug, bei dem auf die Bedeutung eines Briefs als alltagsbezogenes, allgemeines Informations- bzw. Kommunikationsmittel verwiesen wird, lässt sich selten vorfinden ($N = 8$). Der Transfer der Schreibaufgabe auf Alltagsereignisse bzw. Hinweise der Lehrpersonen auf künftige Anwendungsmöglichkeiten bleiben also zu über 80 % unberücksichtigt.

Imaginationsförderung

Ebenfalls wird die Imaginationsförderung während der Aufgabenstellung weniger gut umgesetzt ($M = 2.1$; $SD = 1.10$). Eine hohe Imaginationsförderung zeigt sich beispielsweise daran, dass die Lehrkraft einen Bezug zur Ausgangssituation herstellt (Lucy verlässt die Krokodilfamilie), Kreativitäts- und Vorstellungsvermögen anregende Maßnahmen (etwa eine Fantasiereise) realisiert, Impulse in Form von offenen Fragen gibt und die Schüler/-innen zu Gesprächen bzw. zu einem Nachdenken über das Bilderbuch aktiviert. Seine Relevanz erhält dieses Kriterium auch dadurch, dass es der Vertiefung der Informations- und Wissensgenerierung entspricht, das einen Aspekt für lernförderliche Schreibaufgaben darstellt.

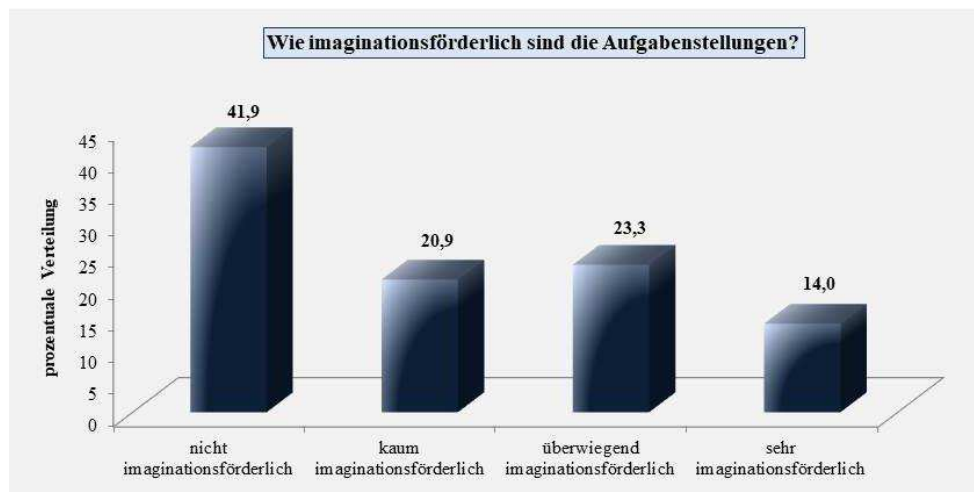


Abbildung 2: Die Imaginationsförderung in der Stichprobe

Schreibstrategien

Als positiv hervorzuheben ist hingegen die Anregung zum prozessorientierten Schreiben, indem die Schüler/-innen schreibstrategische Unterstützung erhalten. In 21 Klassen geben Lehrkräfte Hinweise, Instruktionen und Ratschläge, die das Prozedere des Schreibens beinhalten. So appellieren sie beispielsweise an die Kinder, sich erst Gedanken über den Brief zu machen, bevor sie mit dem Schreiben beginnen, Anlauttabellen als Hilfsmittel heranzuziehen oder ihre Ideen vorab auf einem gesonderten Blatt zu notieren.

Textsortenspezifische Merkmale

Auch die Erklärung textsortenspezifischer Merkmale wird relativ häufig realisiert. Lediglich in neun Klassen werden keine Briefmerkmale (z. B. Anrede- und Schlussformeln) besprochen. In sechs weiteren Klassen weisen die Lehrpersonen die Schüler/-innen darauf hin, auf Briefmerkmale zu achten, ohne sie jedoch zu erläutern. In den übrigen Klassen ($N = 28$) wird das textsortenspezifische Wissen dagegen ausführlich besprochen.

Fachübergreifende Qualitätseinschätzung

Die bisherigen Befunde, die sich auf fachspezifische Merkmale beziehen und den Unterricht in erster Linie beschreiben, lassen sich von den Ergebnissen zu den überwiegend fachübergreifenden, vierstufigen Kriterien unterscheiden, die sich mit der Qualität des Schreibunterrichts auseinandersetzen.

So wurden neben der Verständlichkeit und dem Präzisionsgrad, die im Zusammenhang mit der Dimension „Zielorientierung“ bereits erwähnt wurden, auch die Organisation des schreibdidaktischen Settings untersucht, die sich als überwiegend sinnvoll und im Wesentlichen effektiv vorbereitet charakterisieren lässt ($M = 2.9$; $SD = 1.01$). In einer Klasse, in der die Organisation mit dem höchsten Wert 4 eingeschätzt wird, sind die einzelnen Phasen des schreibdidaktischen Settings gut aufeinander abgestimmt und verlaufen weitgehend reibungslos und störungsfrei. Die Lehrkraft steuert das schreibdidaktische Setting und behält es unter Kontrolle.

Die Aufgabenstellung in der Stichprobe lässt sich auch als durchschnittlich kohärent und fokussiert bezeichnen ($M = 2.8$; $SD = 0.75$). Anhand dieses Items wird erfasst, inwieweit die

Aufgabenstellung einen roten Faden aufweist und sich auf die wesentlichen schreibbezogenen Aspekte konzentriert.



Abbildung 3: Die Organisation und Fokussiertheit in der Stichprobe

Dimensionen Zielorientierung und soziale Interaktion

Dimension „Zielorientierung“

Die zentrale Dimension „Zielorientierung“ wurde insgesamt durchschnittlich gut in der Stichprobe realisiert. So wird die Schreibfunktion in mehr als der Hälfte der Lerngruppen (62.80 %) durch die Lehrkraft thematisiert und in einem Fünftel der Klassen (20.93 %) ausführlich besprochen. In etwas weniger als der Hälfte der Stichprobe (44.18 %) wird das Abschiedsthema während der Ideensammlung von den Schüler/-innen zur Sprache gebracht.

Zudem wird, wie die nachfolgende Abbildung zeigt, der Schreibauftrag durchschnittlich präzise gestellt ($M = 2.7$; $SD = 0.94$) und für die Erstklässler/-innen verständlich kommuniziert ($M = 2.4$, $SD = 0.91$). Der Präzisionsgrad des Schreibauftrags trägt insbesondere zum Verständnis von Arbeitsaufträgen und zur Klarheit der Unterrichtsgestaltung bei, sodass den Schreibanfänger/-innen die Definition der Kommunikationssituation hinreichend bekannt ist. Das Schreibziel, also einen Abschiedsbrief zu verfassen, sowie Kenntnisse zu Adressat/-in, Absender/-in und Textsorte werden von den Lehrpersonen

dementsprechend relativ exakt erläutert. Als verständlich gelten Schreibaufträge, wenn der Arbeitsauftrag beispielsweise einfach formuliert ist und er eine geringe, dem Wortschatz von Erstklässler/-innen angemessene sprachlogische Komplexität aufweist.

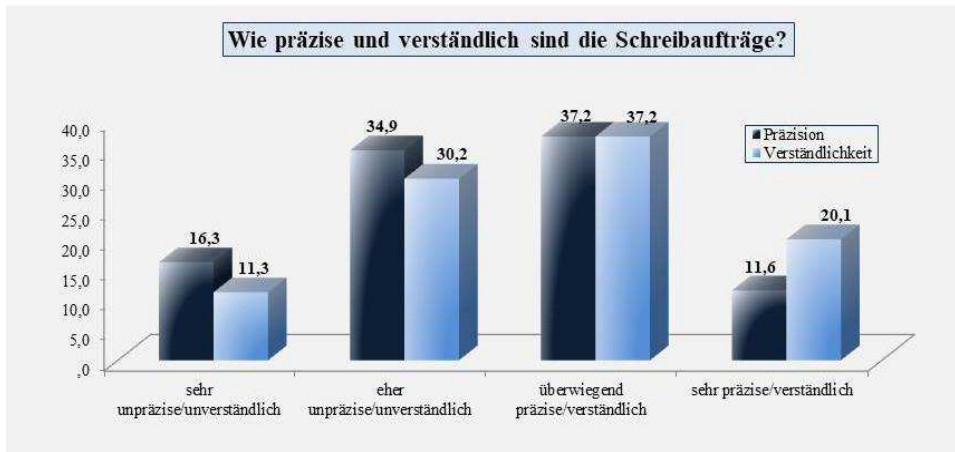


Abbildung 4: Die Verteilung des Präzisionsgrads und der Verständlichkeit der Schreibaufträge in der Stichprobe

Erwartungsgemäß korreliert der Präzisionsgrad mit der Verständlichkeit. So zeigt die Spearman-Rang-Korrelation einen signifikanten Zusammenhang zwischen den beiden Items ($r = .67, p < 0.01$). Auch für die weiteren Items der Dimension „Zielorientierung“ wurden Zusammenhänge untereinander berechnet, womit sich nachweisen lässt, inwieweit die theoretisch gebildeten Dimensionen auch einer empirischen Überprüfung standhalten. So korrelieren ebenfalls die beiden Varianten der Besprechung des Abschiedsthemas mit diesen allgemeinen Merkmalen der Unterrichtsqualität. Die Schreibfunktion weist dabei einen positiven Zusammenhang mit dem Präzisionsgrad mit $r = .32$ ($p < 0.05$) und mit der Verständlichkeit des Schreibauftrags mit $r = .39$ ($p < 0.01$) auf, die Ideensammlung zum Abschiedsthema korreliert mit dem Präzisionsgrad mit $r = .37$ (< 0.05) und mit der Verständlichkeit des Schreibauftrags mit $r = .42$ ($p < 0.05$).

Jedoch besteht zwischen den beiden Versionen der Besprechung des Abschiedsthemas (zum einen als Schreibfunktion, die von den Lehrkräften expliziert wird und zum anderen als Ideensammlung durch die Schüler/-innen) kein Zusammenhang. Dies erscheint zunächst

überraschend, könnte aber dadurch erklärt werden, dass lediglich in 16 Klassen das Abschiedsthema sowohl von der Lehrkraft als auch von den Kindern während der Ideensammlung besprochen wird. In den übrigen Klassen wird dagegen ausschließlich eine der Varianten im Unterricht bevorzugt.

Dimension „Soziale Interaktion“

Die zweite Dimension, die innerhalb der Stichprobe erfasst wurde, stellt die soziale Interaktion dar, die aus den drei Items „Austausch über Schreibideen“, „soziale Interaktion im engeren Sinn“ sowie „Wechsel der Sozialform“ besteht. Das Kriterium „Wechsel der Sozialform“ belegt, dass lediglich in drei Klassen Partner- und Gruppenarbeit stattfinden. Auch eine soziale Interaktion im engeren Sinn wird nur in einer einzigen Klasse umgesetzt. Schreibideen dagegen wurden in mehr als der Hälfte der Klassen ausgetauscht ($N = 26$), wobei am häufigsten (32.6 %) mehrere Ideen im Klassengespräch von den Schüler/-innen vorgetragen werden.

Im Gegensatz zur Zielorientierung muss aufgrund der geringen Fallzahlen auf eine Überprüfung der Korrelationen zwischen den einzelnen Kriterien der sozialen Interaktion verzichtet werden.

Resümee

Zusammenfassend lässt sich feststellen, dass die Dimension „Zielorientierung“ in den Klassen relativ häufig von den Lehrkräften umgesetzt wird, während die Aufgabenstellungen weniger in soziale Interaktionen eingebunden sind. Insgesamt scheinen eher die allgemeinen Aspekte von Unterrichtsqualität (Präzision, Verständlichkeit, Organisation sowie Kohärenz und Fokussierung) realisiert zu werden. Bei den schreibdidaktischen Kriterien des kompetenzorientierten Ansatzes, an dem sich die Auswertung der vorliegenden Aufgabenstellung orientiert (vgl. Kap. 3), zeichnet sich ein kontrastierendes Bild ab: während der Lebensweltbezug, die Differenzierung sowie Schreibkooperationen kaum von den Lehrkräften realisiert werden, treten die Besprechung von Schreibstrategien, die Sammlung von Schreibideen sowie die Erklärung textsortenspezifischer Merkmale häufiger auf.

6. Zusammenfassung, Diskussion und Ausblick

Im vorliegenden Beitrag wurde dargelegt, wie sich eine Aufgabenstellung im schriftsprachlichen Anfangsunterricht videogestützt analysieren lässt. Dazu wurde ein Beobachtungsinstrument entwickelt und nach empirischen Gütekriterien überprüft. Die Entwicklung der insgesamt 25 Kriterien fußt dabei auf schreibdidaktischen und fachübergreifenden Konzeptionen, wobei sich das Instrument insbesondere an der kompetenzorientierten Schreibdidaktik und an „Aufgaben mit Profil“ (Bachmann & Becker-Mrotzek, 2010) orientiert. In diesen Konzeptionen werden unter anderem die Zielorientierung und die Einbettung der Aufgabenstellung in eine soziale Interaktion als schreibförderliche Merkmale hervorgehoben. Die kompetenzorientierte Schreibdidaktik verweist zudem auf die Bedeutung des Lebensweltbezugs, der Sammlung von Schreibideen, der Besprechung textsortenspezifischer Merkmale und von Differenzierungsmöglichkeiten. Um die Sicht- und Tiefenstruktur des Unterrichts abzubilden und damit potentielle Einflussfaktoren auf die Textqualität beleuchten zu können, wurden diese Kriterien guter Aufgabenstellungen um allgemeindidaktische Unterrichtsqualitätsaspekte wie Fokussierung und Organisation ergänzt. Das Auswertungsinstrument umfasst neben 19 Einzelkriterien die beiden Dimensionen Zielorientierung und soziale Interaktion.

Es zeigt sich zunächst, dass die „Zielorientierung“ häufig umgesetzt wird und die einzelnen Items dieser Dimension teilweise miteinander korrelieren, womit sich die Fragestellung 2 bis zu einem gewissen Grad bestätigen lässt. Im Gegensatz dazu werden in der Stichprobe selten Gelegenheiten zur sozialen Interaktion geboten (mit Ausnahme der Sammlung von Schreibideen im Unterrichtsgespräch). Ein Nachweis über Korrelationen der verschiedenen Items ist somit nicht möglich.

Insgesamt setzen sich in der Stichprobe kontrastierende Befunde fort: Während allgemeine Kriterien der Unterrichtsqualität wie Präzision und Verständlichkeit, Kohärenz und Fokussierung sowie die Organisation des Schreibarrangements eine durchschnittlich ausgeprägte Qualität aufweisen, scheinen kompetenz- und schreibdidaktisch orientierte Kriterien teilweise vernachlässigt zu werden. So finden in den untersuchten Aufgabenstellungen kaum Differenzierungsmaßnahmen, Gelegenheiten zur sozialen Interaktion oder eine Betonung des

Lebensweltbezugs statt. Auch die Imaginationsförderung wird eher unterdurchschnittlich umgesetzt.⁹

Andere kompetenzorientierte Merkmale werden hingegen häufiger von den Lehrkräften priorisiert. So werden in fast der Hälfte der Klassen Schreibstrategien thematisiert und textsortenspezifisches Wissen durch die Lehrpersonen vermittelt. Im anglo-amerikanischen Sprachraum wurden beide Merkmale im Zusammenhang mit Textqualität bereits relativ gut erforscht und gelten als förderlich für die Textqualität (Graham, McKeown, Kihara & Harris, 2012).

Aus diesen deskriptiven Befunden eine allgemeine Relevanz bestimmter schreibdidaktischer Merkmale abzuleiten, wäre jedoch vorschnell. So hängt auch von der Zielsetzung des Unterrichts ab, welchen Merkmalen ein Vorzug eingeräumt wird: Steht etwa die Textqualität wie im angewandten kompetenzorientierten Ansatz im Vordergrund, werden eher Ideen gesammelt als wenn „die Kontextualisierung der inneren Sprache unmittelbar im Schreiben“ erfolgen soll (Dehn et al., 2011, S.100; vgl. Kap. 3). Aber auch das vorliegende schreibdidaktische Setting könnte sich auf die Vernachlässigung gewisser Kriterien auswirken. Es erscheint beispielsweise plausibel, dass ein Grund für die eher dürftige Vergegenwärtigung des Bilderbuchgeschehens die im Unterricht vorausgegangene Bilderbuchrezeption darstellt. In manchen Klassen wurde das Vorlesen des Bilderbuchs sehr ausführlich gestaltet und beispielsweise mit szenischen Spielen, in der die Kinder in die Rollen der Protagonisten schlüpfen, kombiniert. Möglicherweise wurde in derartigen Klassen mit einer breit gefächerten Bilderbuchrezeption auf eine detaillierte Vergegenwärtigung verzichtet, ohne dass dies sich negativ auf die Imaginationsfähigkeit auswirkt. Auch die wenig offerierten Gelegenheiten zur sozialen Interaktion stehen eventuell mit der vorausgegangenen Bilderbuchrezeption in einem ähnlichen Zusammenhang: Wenn bereits bei der Vorstellung des Bilderbuchs intensive Interaktionen stattfanden, die zu einem tieferen Verständnis der Geschichte beitragen, werden diese während der Aufgabenstellung nicht wiederholt. Auch der meist fehlende Lebensweltbezug, der jedoch auch durch das kindgerechte Thema des Bilderbuchs sowie der Schreibaufgabe zum Teil selbsterklärend ist, könnte bereits während der Bilderbuchrezeption thematisiert worden sein. Die separate Betrachtung

⁹ Die Umsetzung des Items ($M = 2.1$) ist als eher unterdurchschnittlich zu betrachten, wenn man sich am fiktiven Skalenmittelpunkt ($M = 2.5$) orientiert.

einer einzelnen Unterrichtsphase erscheint im Kontext der Einbettung in einen handlungs- und produktionsorientierten Literaturunterricht kritisch. In weiteren Analysen werden daher die Bilderbuchrezeption und zusätzliche umfangreiche Kontextdaten zu den Schüler/-innen, Eltern und Lehrern berücksichtigt. Auf diese Weise lässt sich feststellen, ob soziale Interaktionen in der Stichprobe tatsächlich unterrepräsentiert sind, oder ob sie im Hinblick auf andere Unterrichtsphasen kompensiert wurden.

Während also die Realisierung bestimmter und die Vernachlässigung anderer schreibdidaktischer Merkmale wahrscheinlich durch das didaktische Setting bedingt sind, deutet sich insgesamt an, dass Lehrkräfte der Zielorientierung und den fachübergreifenden Kriterien der Unterrichtsqualität tendenziell etwas mehr Gewicht einräumen. Mit diesen allgemeinen Aspekten der Unterrichtsqualität und der Zielorientierung, die ebenfalls u.a. aus den beiden Unterrichtsqualitätskriterien Präzision und Verständlichkeit besteht, wird ein Verweis auf die Tiefenstruktur von Unterricht deutlich. Im Gegensatz zu Sichtstrukturen, also leicht zu beobachtende Methoden und Organisationsmerkmale des Unterrichts, lässt sich an Tiefenstrukturen, etwa der Klassenführung, die Qualität von Lehr- und Lernprozessen bestimmen (Kunter & Trautwein, 2013). In der vorliegenden Stichprobe bilden Zielorientierung, Fokussierung und Organisation Tiefenstrukturen ab, aber auch die kompetenzorientierten Merkmale Ideensammlung, Schreibstrategien und Vermittlung von Textsortenwissen können daran anknüpfen. Inwieweit sich diese Tiefenstrukturen als besonders lernwirksam erweisen, wird in weiteren Analysen überprüft, in denen Zusammenhänge zwischen den einzelnen Kriterien und der Textqualität untersucht werden.

Der Fokus der Lehrkräfte auf fachübergreifenden Kriterien der Unterrichtsqualität könnte unter Umständen auch einem Defizit an fachdidaktischem Wissen bzw. dem fehlenden Bewusstsein über seine Relevanz geschuldet sein. Möglicherweise fehlt den Lehrpersonen das Wissen darüber, welche schreibdidaktischen Kriterien lernförderlich sind und auf welche Weise sie im Schreibsetting integriert werden können¹⁰. Ergebnisse der COACTIV-Studie belegen, dass das fachdidaktische Wissen unter Lehrkräften variiert, was sich auch im Praxishandeln

¹⁰ Anhand der vorliegenden Daten können keine Analysen zum fachdidaktischen Wissen der Lehrkräfte in der Studie vorgenommen werden.

niederschlägt (vgl. Kunter et al., 2011). Weitere Studien sollten sich daher mit dem fachdidaktischen Wissen von Deutschlehrern und ihrem Einfluss auf die Gestaltung des Schreibunterrichts auseinandersetzen. Mit Blick auf die Lehrerausbildung und -fortbildung wird deutlich, dass Lehrkräfte in höherem Maße für eine fachdidaktische Gestaltung des Unterrichts sensibilisiert und ihre professionelle Kompetenz dahingehend ausgebildet werden sollte. Ein höheres fachdidaktisches Wissen könnte ein ausgewogenes Zusammenspiel zwischen Schreibdidaktik und fachübergreifenden Merkmalen der Unterrichtsqualität im Schreibunterricht fördern.

Insgesamt gilt es zu beachten, dass die Ergebnisse vor dem Hintergrund des beschriebenen Forschungsdesigns zu interpretieren sind. Potenzielle Einschränkungen resultieren beispielsweise aus der relativ kleinen Stichprobe und der einmalig stattgefundenen Videografie. So erheben die Befunde, die lediglich eine einzige Deutschstunde abbilden, keinen Anspruch auf Repräsentativität.

Resümierend lässt sich aus dem vorliegenden Beitrag ableiten, wie ein Beobachtungssystem für die Analyse von Aufgabenstellungen im Anfangsunterricht entwickelt und angewandt werden kann. Für die Schreibdidaktik konnte in dieser Studie erstmals empirisch aufgezeigt werden, wie Lehrkräfte Aufgabenstellungen in Klasse 1 gestalten, welche Schwerpunkte sie setzen und welchen Merkmalen sie weniger Gewicht verleihen.

Literatur

- Abraham, U., & Kupfer-Schreiner, C. (Hrsg.) (2007). *Schreibaufgaben. Für die Klassen 1 bis 4*. Berlin: Cornelsen.
- 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.
- Bachmann, T., & Becker-Mrotzek, M. (2010). Schreibaufgaben situieren und profilieren. In T. Pohl & T. Steinhoff (Hrsg.), *Textformen als Lernformen* (Kölner Beiträge zur Sprachdidaktik, Reihe A, Bd. 7, S. 191–209). Duisburg: Gilles & Francke.
- Bachmann, T. Ospelt-Geiger, B. Ospelt, K. & Vital, N. (2007). *Aufgaben mit Profil. Frühe Förderung funktional-pragmatischer*

- Schreibfähigkeiten*. Unveröffentlichter Schlussbericht zum internen Projekt der Pädagogischen Hochschule Zürich. Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. H. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & Ulrich Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 314–354). Münster: Waxmann.
- Baumann, 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 Verlag Scriptor.
- Becker-Mrotzek, M. (2014). Bildungsstandards und Schreibaufgaben. In H. Feilke & T. Pohl (Hrsg.), *Schriftlicher Sprachgebrauch – Texte verfassen* (Deutschunterricht in Theorie und Praxis, Bd. 4, S. 481–500). Baltmannsweiler: Schneider.
- Bereiter, C. (1980). Development in Writing. In L. Gregg & E. Steinberg (Hrsg.), *Cognitive Processes in Writing* (S. 73-93). Hillsdale: Erlbaum.
- Blömeke, S., Risse, J., Müller, C., Eichler, D., & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht. Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft*, 34(4), 330–357.
- Bohl, T., & Kleinknecht, M. (2009). Aufgabenkultur. In S. Blömeke (Hrsg.), *Handbuch Schule. Theorie - Organisation - Entwicklung* (S. 331–334). Bad Heilbrunn: Klinkhardt.
- Dann, H.-D., Diegritz, T., & Rosenbusch, H.S. (Hrsg.) (1999). *Gruppenunterricht im Schulalltag. Realität und Chancen* (Erlanger Forschungen, Bd. 90). Erlangen: Universitätsbund Erlangen-Nürnberg.
- Dehn, M. (2007). *Kinder & Lesen und Schreiben. Was Erwachsene wissen sollten*. Seelze: Kallmeyer & Klett.
- Dehn, M. (2009). Zur Funktion der Aufgabe für den Schülertext. In B. Hoffmann & R. Valtin (Hrsg.), *Projekte, Positionen, Perspektiven. 40 Jahre DGLS* (S. 145-175). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.

- Dehn, M., Merklinger, D., & Schüler, L. (2011). *Texte und Kontexte. Schreiben als kulturelle Tätigkeit in der Grundschule*. Seelze: Kallmeyer & Klett.
- Doucet, S. A., & Wilsdorf, A. (2005). *Lucy rettet Mama Krokodil*. Hamburg: Oetinger.
- 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.
- Graham, S., McKeown, D., Kihara, S., & Harris, K. R. (2012). A Meta-Analysis of Writing Instruction for Students in the Elementary Grades. *Journal of Educational Psychology*, 104(4), 879–896.
- Helmke, A. (2004). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern*. Seelze: Klett Kallmeyer.
- Helmke, A., Helmke, T., & Schrader, F.-W. (2007). Qualität von Unterricht: Aktuelle Tendenzen und Herausforderungen im Hinblick auf die Evaluation und Entwicklung von Schule und Unterricht. *Pädagogische Rundschau*, 61(5), 527–543.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyer Klett.
- Helmke, A., & Schrader, F.-W. (2010). Hochschuldidaktik. In D.H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl., S. 273–279). Weinheim: Beltz.
- Hugener, I. (2006). Überblick über die Beobachtungsinstrumente. In I. Hugener, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“* (Teil 3, Videoanalysen, Bd. 15, S. 45–54). Frankfurt a.M.: GFPP/DIPF.
- Kohl, E.-M., & Ritter, M. (2010). *Schreibszenarien. Wege zum kreativen Schreiben in der Grundschule*. Seelze: Kallmeyer Klett.
- Kruse, I., Gabriel, K., & Faust, G. (2013). Hochinferentes Rating: Qualität angeleiteter Bilderbuchrezeption. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien* (S. 219–253). Frankfurt a.M.: GFPP.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von*

Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV.
Münster: Waxmann.

- Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts.* Paderborn: Schöningh.
- Kürzinger, A., & Pohlmann-Rother, S. (2015). Möglichkeiten einer objektiven und reliablen Bestimmung von Textqualität im Anfangsunterricht – Methodisches Vorgehen und deskriptive Befunde aus dem Projekt NaSch1. *Didaktik Deutsch*, 20(38), 60-79.
- Kürzinger, A., & Pohlmann-Rother, S. (in Vorb.). Modellierung eines Instruments zur Auswertung der Aufgabenstellung im Schreibunterricht von Klasse 1. In M. Hess & F. Lipowsky (Hrsg.), *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern (PERLE)“*. Frankfurt a.M.: GFFP.
- Kürzinger, A., Pohlmann-Rother, S., Hess, M., & Lipowsky, F. (eingereicht.). Zur Bedeutung der Aufgabenstellung für die Textqualität im Anfangsunterricht – Wie wirken sich Schreibfunktion, Verständlichkeit und Präzision auf die Schreibleistungen in der ersten Klasse aus? *Unterrichtswissenschaft*
- Langer, I., Schulz von Thun, F. & Tausch, R. (1974). *Verständlichkeit. In Schule, Verwaltung, Politik und Wissenschaft.* München: Reinhardt.
- 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., & Corvacho del Toro, I. (2013). Die Videostudie im Fach Deutsch: „Lucy rettet Mama Kroko“. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien* (S. 29–36). Frankfurt a.M.: GFFP.
- Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule. Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr.* Wiesbaden: Springer VS.
- Lotz, M., Berner, N. E., & Gabriel, K. (2013). Auswertung der PERLE-Videostudien und Überblick über die Beobachtungsinstrumente. In M. Lotz, F. Lipowsky & G. Faust (Hrsg.), *Technischer Bericht zu den PERLE-Videostudien* (S. 83– 103). Frankfurt a.M.: GFFP.
- Maier, U., Bohl, T., Kleinknecht, M., & Metz, K. (2013). Allgemeindidaktische Kategorien für die Analyse von Aufgaben. In M. Kleinknecht, T. Bohl, U. Maier & K. Metz (Hrsg.), *Lern- und*

- Leistungsaufgaben im Unterricht. Fächerübergreifende Kriterien zur Auswahl und Analyse* (S. 9–45). Bad Heilbrunn: Klinkhardt.
- Melhorn, G. & Melhorn, H.-G. (2003). Kreativitätspädagogik – Entwicklung eines Konzepts in Theorie und Praxis. *Bildung und Erziehung*, 56 (1), 23-45.
- 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.
- Richter, K. (2007). *Kinderliteratur im Literaturunterricht der Grundschule. Befunde – Konzepte – Modelle*. Baltmannsweiler: Schneider.
- Schneuwly, B. (1996). Der Nutzen psychologischer Schreibforschung für die Didaktik des Schreibens. In H. Feilke (Hrsg.), *Schreiben im Umbruch. Schreibforschung und schulisches Schreiben* (S. 29–39). Stuttgart: Klett.
- Slavin, R.E. (1996). *Education for All*. Lisse: Swets & Zeitlinger.
- Weinhold, S. (2014). Schreiben in der Grundschule. In H. Feilke & T. Pohl (Hrsg.), *Schriftlicher Sprachgebrauch. Texte verfassen* (S. 143-158). Baltmannsweiler: Schneider.