

Bibliographischer Hinweis sowie Verlagsrechte bei den online-Versionen der DD-Beiträge:



**Halbjahresschrift für die Didaktik
der deutschen Sprache und
Literatur**

<http://www.didaktik-deutsch.de>
17. Jahrgang 2012 – ISSN 1431-4355
Schneider Verlag Hohengehren
GmbH

*Norbert Kruse, Anke Reichardt, Maik
Herrmann, Friederike Heinzl
& Frank Lipowsky*

**ZUR QUALITÄT VON
KINDERTEXTEN
Entwicklung eines
Bewertungsinstruments in der
Grundschule**

In: Didaktik Deutsch. Jg. 17. H. 32. S. 87-
110.

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. – Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden.

Norbert Kruse, Anke Reichardt, Maik Herrmann, Friederike Heinzl
& Frank Lipowsky

ZUR QUALITÄT VON KINDERTEXTEN

Entwicklung eines Bewertungsinstruments in der Grundschule

Abstract

In the project “Peer feedback within cooperative settings for text revision in German classes in primary schools” a rating system comprising 10 dimensions was developed to evaluate the quality of student story writing. The sample consisted of 132 texts written by 3rd grade students. First analyses show that the quality of student texts could be evaluated with an acceptable degree of correspondence among the raters. Secondly, the structure of the hypothesized model was tested by using a confirmatory factor analysis (CFA). These results indicated that the hypothesized model did not fit the data. Therefore the model was modified by eliminating three items. This modified model allows us to generate two valid factors (textuality – conventional/unconventional), which fit the model adequately. The factor *textuality conventional* comprises the items text coherency, textual implicitness, textual explicitness, vocabulary, the factor *textuality unconventional* comprises textual risk-taking, with regard to language and to content.

Zusammenfassung

Im Rahmen des Forschungsprojekts „Kooperative Schülerrückmeldungen bei der Textüberarbeitung im Deutschunterricht der Grundschule“ wurde ein Ratingsystem entwickelt, mit dessen Hilfe die Qualität von Schülertexten in 10 Dimensionen eingeschätzt wurde. Die herangezogene Stichprobe bestand aus 132 Texten von Schülern des 3. Schuljahres. Die Analysen zeigen in einem ersten Schritt, dass sich die Qualität der Texte auf den 10 unterschiedlichen Dimensionen mit einer hohen Beurteilerübereinstimmung einschätzen ließ. Mittels konfirmatorischer Faktorenanalyse wurde in einem zweiten Schritt die angenommene Modellstruktur überprüft. Die entsprechenden Ergebnisse zeigen, dass das ursprünglich angenommene Modell nicht zu den Daten passte. Daher wurde das Modell modifiziert, indem drei Items aus dem Modell entfernt wurden. Die erneute Überprüfung des reduzierten Modells mit den beiden Faktoren Textualität konventionell und Textualität unkonventionell ergab schließlich einen akzeptablen Modellfit. Der Faktor *Textualität konventionell* setzt sich aus den Items Kohärenz, Impliztheit und Expliztheit des Textes sowie Wortschatz zusammen aus, der Faktor *Textualität unkonventionell* aus den Items sprachliches Wagnis und inhaltliches Wagnis zusammen.

1 Einleitung

Im vorliegenden Beitrag wird ein Instrument zur Bewertung der Qualität von Schülertexten vorgestellt, das im Rahmen der DFG-Studie KoText entwickelt wurde.¹ In der Studie geht es um die Rückmeldungen von Schülerinnen und Schülern der 3. Klassenstufe zu Texten, die von ihren Mitschülern verfasst wurden und um die daraus resultierenden Textüberarbeitungen.

Die Erhebungen wurden in der Stadt und im Kreis Kassel von Januar bis Juni 2010 durchgeführt. Insgesamt nahmen 357 Schülerinnen und Schüler aus 16 Klassen teil, die aufgefordert waren, Geschichten zu Bildern zu schreiben. Das Schreiben zu Vorgaben (Dehn/ Merklinger/ Schüler 2011, S. 96ff.) wurde in einen funktionalen Rahmen eingebettet, es sollte ein Geschichtenbuch für eine andere 3. Klasse entstehen.

Nachfolgend wird ein Teilvorhaben aus diesem Projekt erläutert und diskutiert. Denn zur Bewertung der Qualität der Erst- und Endfassung der Texte musste ein Analyseinstrument entwickelt werden, das den besonderen Bedingungen des Schreibunterrichts in der Grundschule entsprach und die Qualität der Kindertexte zuverlässig messen sollte. Dabei wurde unter Rückgriff auf das Zürcher Textanalyseraster (ZTR) (vgl. Nussbaumer/Sieber 1994) für die Grundschule ein Instrument konstruiert, das sich an fachdidaktischen Ansprüchen von aufgabenorientierten und schreiberdifferenzierten Kriterien orientieren und zugleich den Anforderungen für Gütekriterien von Erhebungsinstrumenten genügen sollte.

2 Forschungsstand zur empirischen Bestimmung von Textqualität

Das derzeit feststellbare starke Interesse an der Bestimmung der Textqualität von Lernertexten² resultiert u.a. aus den internationalen Schulleistungsstudien und der Einführung von Bildungsstandards. Blatt/ Ramm/ Voss (2009) sehen in der „Beurteilung von Schülertexten“ eine herausfordernde und für „die Praxis“ relevante Forschungsaufgabe.³ Die dazu nötigen Kriterienkataloge seien, so heißt es in dem

1 Das DFG-Projekt hat den Titel „Kooperative Schülerrückmeldungen bei der Textüberarbeitung im Deutschunterricht der Grundschule (KoText)“. Leitung: Friederike Heinzl, Norbert Kruse, Frank Lipowsky, Förderkennzeichen HE 4562/2-1.

2 Wir sprechen in unserer Studie von Schülertexten, Kindertexten oder auch von Lernertexten. Obwohl der Begriff ‚Lernertext‘ häufig für den Zweit- und Mehrspracherwerb gebraucht wird, halten wir ihn auch für die schriftlichen Produkte von Grundschulern, die in ihrer Muttersprache schreiben, für sinnvoll. In die Schreibprodukte sind einerseits frühere Lernprozesse eingegangen, so dass Ergebnisse dieses Lernens dokumentiert sind, zweitens aber enthalten sie auch das Potenzial für zukünftige Lernprozesse. Der Ausdruck fasst mithin sinnvoll die Spannung von Produkt und Prozess des Schreibens.

3 Bislang ungenügend geklärt scheint in diesem Zusammenhang das Verhältnis von Textqualität und Schreibkompetenz – wie der uneinheitliche Gebrauch der Termini belegt

Beitrag, in ihrer linguistischen Variante „fachlich“ bestimmt, im Hinblick auf „empirische Beurteilungskriterien“ dagegen auf das „Maß an Übereinstimmung der Bewerter“ (ebd., S. 55) hin ausgerichtet. Für die Entwicklung eines Instruments zur Erhebung und Bestimmung der Qualität von Lernertexten geht es demnach um die Integration textlinguistischer und schreibdidaktischer, aber auch forschungsmethodischer Ansprüche an fachdidaktisch-empirische Unterrichtsforschung.

So ging es z. B. im Zuge des Projekts „Lernstand in Klasse 6“ (2008) darum, Vergleichsarbeiten mit einem hohen wissenschaftlichen Standard zu entwickeln (vgl. Blatt/ Ramm/ Voss 2009). Zur Erfassung der Textkompetenz sollte eine Schreibaufgabe (Textsorte Brief) mit dazugehöriger Kodieranweisung für die Hand der Lehrpersonen erarbeitet werden. Die Pilotierung, auf deren Werte hier Bezug genommen wird, erfolgte mit geschulten Deutschlehrkräften, die 431 Schülertexte kodierten. Jedoch zeigten sich Probleme einer streng kategorialen Beurteilung von Texten, die auch in der IGLU-E-Aufsatzstudie (Blatt/ Voss/ Matthießen 2005) deutlich wurden. Nach dem Ausschluss von Kategorien, die nicht reliabel zu bewerten waren, wurden – mit Verweis auf ‚gängige schulische Praktiken‘ – 16 dichotome Items zum Inhalt, dem Aufbau, der Sprache und der Einhaltung von Schreibkonventionen (Rechtschreibung, Grammatik, Schriftbild) erhoben (Blatt/ Ramm/ Voss 2009, S. 64). Eine Modellprüfung zur Frage, inwieweit Schreibkompetenz ggf. differenziell strukturiert ist (etwa hinsichtlich zweier getrennter Faktoren zu Inhalt und Aufbau einerseits und Sprache und Stil andererseits), zeigte entgegen den Annahmen eine lediglich einfaktorielle Datenstruktur aller Items (ebd., S. 75). Als Maß der Interraterreliabilität wurde Cohens Kappa berechnet (ein Wert < 50 wird als befriedigende und < 75 als sehr gute Übereinstimmung angenommen). Die Werte lagen zwischen .19 (Rechtschreibung/Grammatik) und .90 (Unterschrift vorhanden). Die Autoren schlussfolgern daher, dass „die angestrebten Übereinstimmungen von mindestens 0,7 [...] für die Kriterien der Schreibaufgabe nicht erreicht werden [können]“ (ebd., S. 71).

Eine weitere groß angelegte quantitative Untersuchung der Textkompetenz erfolgte im Rahmen der DESI-Studie (Neumann 2007). Die Aufsätze der 9.-Klässler aus vier verschiedenen Aufgabenstellungen zur Textsorte Brief wurden zusammen mit den Daten der Hamburger Aufsatzstudie aus Klasse 11 (LAU 11/ ULME 1) untersucht. Mittels 43-47 dichotomer Items wurden formale und inhaltliche Elemente der Briefe auf ihr Vorhandensein hin geprüft. Fünfstufige Ratingskalen erfassten zusätzlich die Qualität der Texte hinsichtlich Textaufbau, Stil, Wortwahl, Orthographie, Gramma-

(vgl. u. a. Blatt/ Voss/ Matthießen 2005, Blatt/ Ramm/ Voss 2009, Neumann/ Lehmann 2008). Das KoText-Qualitätsraster misst unseres Erachtens keine Schreibkompetenz, sondern lediglich die Qualität von (Text)Produkten, aus der allenfalls indirekt auf eine vorhandene oder nicht vorhandene Kompetenz geschlossen werden kann. Denn Schreibkompetenzen fallen in den Bereich der Potentialität, nicht der Aktualität (Steinhoff 2010, S. 268). Bereits Lehmann (1988, S. 349ff.) stellt fest, dass Leistungsunterschiede zwischen Schülern keineswegs die wichtigste Varianzkomponente aufweisen, etwa doppelt so gewichtig war die Leistungsschwankung über verschiedene Aufgaben hinweg. Aus der Bearbeitung von vier Aufsätzen ließ sich keine ‚generalisierte Schreibfähigkeit‘ ableiten.

tik und Satzkonstruktion. Über Strukturgleichungsanalysen zur Identifizierung von Teilfähigkeiten konnten textimmanente Beziehungen der inhaltlichen, formalen und sprachlichen Teildimensionen abgebildet werden: Es zeigte sich hier ein zweidimensionales Faktorenmodell, das einerseits aufgabenspezifische Aspekte der Semantik/Pragmatik (Textaufbau, Wortwahl, Stil sowie inhaltliche und formale Merkmale) bündelt. Andererseits zeigte sich ein aufgabenunabhängiger Faktor mit sprachsystematischen Merkmalen wie Rechtschreibung, Grammatik und Satzkonstruktion.⁴ Die Beurteilerübereinstimmung lag für die DESI-Daten bei den dichotomen Items bei 95,3% sowie bei 95,2% (mit max. einer Stufe Abweichung) für die Ratingskalen. Die Autorin geht daher von einer objektiven Erfassung der Datenqualität aus (Neumann 2007, S. 89).

Auch eine aktuelle Untersuchung (Bremerich-Vos/ Possmayer 2011) stellt die Frage nach den Möglichkeiten einer reliablen Bewertung von Schülertexten. In ihrem Beitrag geht es den Autoren um die Analyse, mit welcher Beobachterübereinstimmung die Zuordnung des von Augst et al. (2007) erhobenen Textkorpus zu den Kompetenzniveaus 1-4 erfolgen kann. Die Schülertexte aus der Grundschule sollten nochmals von mehreren Beurteilern unabhängig voneinander holistisch kodiert werden. Dabei zeigte sich, dass die Beurteilerübereinstimmung, insbesondere für die narrativen Texte, zunächst vergleichsweise gering ausfiel (Bremerich-Vos/ Possmayer 2011, S. 42). Erst nach einer Modifikation der Kompetenzmodelle durch eine weniger strenge Detaillierung der vier Niveaus (insbesondere hinsichtlich der graduellen Ausgestaltung der Punkte) erreichte die Interraterreliabilität annähernd die angestrebten Werte (ebd., S. 46). Das gewichtete Kappa lag für die narrativen Texte im Mittel bei .59 ($SD = .05$), für die argumentativen Texte bei .64 ($SD = .03$). Eine Analyse der Validität erfolgte in diesem Rahmen nicht.

Die bislang vorliegenden Untersuchungen verdeutlichen zugleich die aktuellen Herausforderungen. Die folgenden Ausführungen sollen nun zeigen, dass empirische Untersuchungen zur Messung der Textqualität einerseits Modifikationen eines schreibdidaktisch begründeten Textmodells erfordern, sich andererseits aber durchaus fachdidaktisch vertreten lassen, und dabei zugleich – unter bestimmten Voraussetzungen – ein hohes Maß an Interraterreliabilität erreichen können.

3 Modellierung und Dimensionen des Qualitätsrasters⁵

Das Qualitätsraster ist in Orientierung an dem vielfach diskutierten und für den schulischen Schreibunterricht adaptierten ‚Zürcher Textanalyseraster‘ (ZTR) (Nussbaumer/ Sieber 1994) entstanden, das erstmals schon 1991 von Nussbaumer

4 Die Datenstruktur steht damit in Analogie zu den Ergebnissen der Hamburger Aufsatzstudie (Hartmann/ Lehmann 1987), die einen aufgabenspezifischen Faktor zu Inhalt/ Aufbau sowie einen ‚Mechanics‘-Faktor (Orthographie/ Grammatik) identifizierte.

5 Aus Platzgründen kann an dieser Stelle das Raster nur in seinen wesentlichen Grundzügen skizziert werden. Das vollständige Manual kann als pdf-Datei abgerufen werden unter <http://www.uni-kassel.de/fb2/primardidaktik/mitarbeiter/kruse.html#forschung>

vorgestellt wurde. Für das KoText-Raster wurden ausgewählte Aspekte aus dem ZTR übernommen, aber auch Modifikationen vorgenommen. Das KoText-Raster besitzt einen ‚Null-Teil‘ mit fünf Items und ein weiteres, dem Rating gleichsam vorgeschaltetes Item, mit dem der Text holistisch bewertet wird. Zwei Basisdimensionen, jeweils mit sieben bzw. zwei Items, bilden dann den Kern des Bewertungsinstruments. Zum Raster gehört ein Manual im Umfang von über 50 Seiten, in dem der theoretische Rahmen zur Bewertung von Kindertexten dargelegt ist und jedes Item mit Bezug zum fachlichen Diskurs benannt, seine Grundidee beschrieben und die Indikatoren der Bewertung aufgeführt sind. Schließlich werden in einer vierstufigen Skala die Werte jeweils erläutert und jeder einzelne Wert mit Ankerbeispielen ausgestattet. Die einzelnen Teile seien nun im Folgenden vorgestellt und erläutert.

3.1 Zur Funktion des Null-Teils

Das Raster beginnt mit einem deskriptiven Null-Teil, der nicht in die Bewertung eingeht, aber dennoch wichtiger Bestandteil des Qualitätsrasters ist, weil er eine operative Funktion für den Prozess der Bewertung durch zwei unabhängige Rater übernimmt. Die Rater sollen vor der eigentlichen Bewertung die Besonderheiten des jeweiligen Textes in den Blick nehmen, um vorschnelle Verallgemeinerungen zu vermeiden. Indem sie aufgefordert sind, zunächst den Text anhand der Handlungsepisoden zu strukturieren, den Inhalt komprimiert zu paraphrasieren, die Themenentfaltung zu charakterisieren, das dominierende Genre zu bestimmen und das Textthema zu benennen, wird der jeweilige Text mit Bezug auf das Vorgefundene qualifiziert, nicht im Hinblick auf abstrakte und vom Schreibprozess losgelöste Textmuster. Auf diese Weise wird das Ratingverfahren sensibel für die jeweiligen Stärken und Schwächen auf inhaltlicher, sprachlicher und formaler Ebene. Die im Null-Teil erfolgte Charakterisierung der Besonderheiten des Textes sichert die Lernerorientierung des Bewertungsprozesses, weil hier der strukturelle Rahmen zur Einschätzung der nachfolgenden Items gebildet wird.

Der Null-Teil soll also u. a. bewirken, dass im Bewertungsprozess Zuordnungen nicht ausschließlich nach einer vorab festgelegten Merkmalsmatrix vorgenommen werden können. Die Übersicht der Tab. 1 zeigt, dass von den Ratern beispielsweise Interpretationen zum möglichen Textthema gefordert sind (Item 0.5). Es geht also um eine bewertende Rekonstruktion dessen, was sich in den Texten finden lässt, nicht um das, was sich nach Kriterien eines Idealtextes dort finden lassen sollte. Fachdidaktisch bedeutsam ist daran erstens, dass von vermeintlichen Defiziten eines Textes nicht auf die Ideen oder Schreibkompetenz des Schreibers geschlossen werden kann. Und zweitens folgt daraus eine kooperative Lesehaltung, die bei der Bewertung eingenommen werden muss, so dass das jeweils erworbene Können der Lernenden indirekt in die Beurteilung des Textes einfließt (vgl. dazu auch Fix 2006, S. 194ff.).

Rater-ID:		Text-ID:		0.1.Wörterzahl (token):	
0.2	Inhalt anhand der Handlungsepisoden wiedergeben	E:			
		HE1:			
		...			
		S:			
<i>hier mehrere Kreuze möglich:</i>					
0.3	Modalität der Themenentfaltung	Narrativ- erzählend			
		Explicativ-erklärend			
		Deskriptiv-beschreibend			
		Argumentativ-begründend			
<i>hier ein Kreuz für das dominierende Genre:</i>					
0.4	Genre	<i>fiktive Texte</i>	Liebesgeschichte		
			Kriminal- und Detektivgeschichte		
			Horror-, Schauer-, Gruselgeschichten		
			Abenteuergeschichte		
			Fabel		
			Sonstige fiktive Texte		
			Beschreibung, Bericht, Erklärung		
			Sonstige Sachtexte		
0.5	Textthema und Profilierung	Welches Thema behandelt der Text?			
		Durch welche Oppositionen wird das Thema profiliert?			

Tab. 1: Null-Teil des KoText-Analyserasters

3.2 Modifikationen des Zürcher Textanalyserasters

Das ZTR musste für das KoText-Raster aus mehreren Gründen modifiziert werden. Während das ZTR für eine qualitative Untersuchung konzipiert wurde, sollten die Texte des KoText-Korpus intervallskaliert geratet werden. Dazu galt es, jedem Indikator einen Wert 1 bis 4 zu hinterlegen.

Weitere Modifikationen waren nötig, da Texte von Grundschulern mit vornehmlich narrativer Themenentfaltung analysiert werden sollten. Erste Versuche der Analyse von Texten mit den aus dem Zürcher Analyseraster angepassten Kategorien hatten gezeigt, dass diese nicht vollständig in der Lage waren, die vorfindlichen Lernertexte mit ihren Qualitäten abzubilden. Induktive Analyseprozesse förderten lernertextspezifische Besonderheiten zutage, für die es weitere Kriterien zu entwickeln galt. Dazu wurden Einzeltexte sowohl ganzheitlich als auch sequenziell textlinguistisch untersucht. Gegenüber dem ZTR sind im Null-Teil (s.o.) die rein deskriptiven Items 0.2 Inhalt anhand der Handlungsepisoden, 0.4 Genre und 0.5 Textthema und seine Profilierung ergänzt sowie innerhalb der Basisdimensionen die Items 1.6 Angemes-

senheit der Überschrift und 1.7 Orientierung am Grundmuster fiktiver, narrativer Texte.

Außerdem ist das KoText-Raster um das Item *Holistisches Urteil über Textqualität vor dem Gesamtcharakter des Textes* erweitert. Tab. 2 zeigt stellvertretend für die zehn Items – die im Gegensatz zu den Items des Null-Teils hoch-inferent geratet wurden – die Kurzfassung des Manuals, das dem Bewertungsprozess zugrunde lag.

Item	Grundidee (Kurzfassung)	Ratingformat
Holistisches Urteil über Textqualität vor dem Gesamtcharakter des Textes Vgl. Brinker (2010)	Das holistische Urteil soll global die Textqualität unter Berücksichtigung mehrerer Einzelaspekte erfassen. Es wird getroffen nach der Erfassung der textbestimmenden Charakteristika aus dem Null-Teil, da hier bereits sowohl positive als auch negative Auffälligkeiten im Text deutlich sichtbar werden. Das Item übernimmt damit die Rolle einer Generalvariable.	Wert 1 Wenig gelungen ist ein Text, wenn die Bestimmung des Textthemas schwer gefallen ist, die Textteile wenig verbunden nebeneinander stehen, die sprachliche Ausgestaltung problematisch und der Inhalt wenig ansprechend ist.
	<i>Indikatoren (Kurzfassung)</i> Das Item berücksichtigt den Kontext der Aufgabenstellung im Hinblick auf die a) Bestimmung des Textthemas b) Textstruktur (z. B. Kohärenz) c) sprachliche und d) inhaltliche Ausgestaltung. Der Kern des Textinhalts ist das Textthema. Der Begriff Textstruktur bezieht sich auf die Gesamtstruktur eines Textes, gebildet aus den Teiltexten und deren Beziehungen untereinander und Gefüge von Einheiten im Text, zwischen denen semantisch-syntaktische Relationen bestehen. Die sprachliche Ausgestaltung umfasst die Angemessenheit der Wortwahl, des Satzbaus und den Einsatz besonderer ästhetischer Mittel. Unter Textinhalt versteht man die Gesamtheit der Informationen eines Textes.	Wert 2 In Teilen gelungen ist ein Text, wenn ein Textthema bestimmt werden konnte, aber z.B. nicht dominant ist, die Textteile an einigen Stellen nicht miteinander verbunden sind, die sprachliche Ausgestaltung und der Inhalt in Teilen ansprechend ist. Wert 3 Überwiegend gelungen ist ein Text, wenn ein dominantes Textthema bestimmt werden konnte, die Textteile meist gut miteinander verbunden sind, die sprachliche Ausgestaltung überwiegend gelungen und der Inhalt überwiegend ansprechend ist. Wert 4 Vollständig gelungen ist ein Text, wenn ein dominantes Textthema problemlos und unmittelbar bestimmt werden konnte, die Textteile sehr gut miteinander verbunden wurden, die sprachliche Ausgestaltung sehr gelungen und der Inhalt besonders ansprechend ist.

Tab. 2: Item *Holistisches Urteil über Textqualität vor dem Gesamtcharakter des Textes*

Der gesamte „A-Teil“, der im ZTR die sprachsystematische und orthographische Richtigkeit bewertet, wurde für die Analysen bewusst ausgeklammert, da aus-

schließlich die Textualität auf der Ebene des Textes, nicht aber die syntaktische und graphematische Ebene mit der Einhaltung sprachsystematischer Normen im Zentrum des Interesses standen.

Da die Texte im Rahmen eines Projekts zur Untersuchung von kooperativer Textüberarbeitung geschrieben wurden, wurden sie für die Kinder vor der späteren Überarbeitung abgetippt sowie rechtschriftlich und syntaktisch korrigiert.⁶ Dadurch sollte die Aufmerksamkeit der sogenannten „Mitarbeiterkinder“ (vgl. Spitta 1992, S. 42) von der graphematischen Ebene auf die inhaltliche Ebene gelenkt werden.⁷ Außerdem konnten durch diese Maßnahme die Ausgangsbedingungen (hier in Form der Herstellung einer unproblematischen Lesbarkeit der Texte für die Gruppenmitglieder) weiter vereinheitlicht werden, so dass die Möglichkeit bestand, die Textualität der Texte im Hinblick auf ihre kommunikative Funktion zu fokussieren. Die Rater erhielten ebenfalls die gedruckten und korrigierten Erstfassungen der Kindertexte, um die Einflüsse auf die Beobachterobjektivität gering zu halten.

3.3 Textualität in Lernertexten

Die Anforderungen an einen zureichenden schreibdidaktischen Textbegriff, mit dem zugleich ein Modell zur Analyse der Ergebnisse der Textbewertung konstruiert werden kann, führen einerseits zu Integrationsproblemen im Hinblick auf textlinguistische und empirische Dimensionierungen von Textualität. Andererseits ergeben sich Probleme zwischen Textlinguistik und Schreibdidaktik. Im KoText-Modell ist versucht worden, die sich daraus ergebenden Integrationsprobleme mit einer *Konventionalitäts-* und einer *Unkonventionalitätsdimension* der Lernertexte zu perspektivieren.

Operiert wird mit dem von Ehlich entwickelte Textbegriff, demzufolge die Produktion und Rezeption von Texten auf kulturell gültigen Verfahren zur Bewältigung „zerdehnter Sprechsituationen“ (Ehlich 1983, 2005) beruht. Baurmann/ Pohl (2009, S. 97) fassen die aus der Idee der ‚zerdehnten Kommunikation‘ entstehenden Anforderungen an schriftliche Texte wie folgt zusammen: „Während sich ein (mündlicher) Sprecher auf Kontextinformationen stützen kann, muss der (schriftliche) Textproduzent solche für das Verstehen zentralen Informationen in seinem Text erst etablieren, also eine aus sich heraus verständliche ‚Textwelt‘ erschaffen“ (Ergänzungen in Klammern von den Verfassern des Beitrags).

6 Diese abgetippte Fassung der Kindertexte wurde mit dem Terminus ‚Erstfassung‘ versehen.

7 Aus der Schreibforschung ist bekannt, dass bei jüngeren Schreibenden die Aufmerksamkeit beim Überarbeiten lokal gebunden ist (vgl. dazu Langer 1986, Fitzgerald 1987, Feilke 1993).

3.3.1 Textualität – konventionell

Die Orientierung an funktionalen und kommunikativen Ansprüchen der Textbeschreibung führt beim KoText-Qualitätsraster dazu, die dieser Dimension zugeordneten Items als konventionell zu qualifizieren, weil mit ihnen das Normale und Übliche bei der Beschreibung der Texthaftigkeit von schriftlichen Texten betrachtet werden soll.⁸ Die Basisdimension 1 des Rasters haben wir deshalb *Textualität – konventionell* genannt, sie umfasst zum Bereich der kommunikativen und funktionalen Angemessenheit der Texte sieben Items. Mit den Items 1.1 bis 1.3 sind textuelle Merkmale erfasst, die den Text im Hinblick darauf bewerten, ob eine konsistente Textwelt entsteht und ein Textthema erkennbar ist. Das Item ‚Kohärenz‘ gilt dabei als textlinguistisch klassisches Kriterium für die Textualität eines Textes. Implizitheit und Explizitheit der Lernertexte zeigen, in welchem Maße ein Leser Korrekturleistungen zu erbringen hat, weil ein Text zu ‚dunkel‘ oder zu wenig nachvollziehbar ist oder positiv bewertet werden kann, weil nichts Unnötiges den Text belastet.

Item 1.1 Charakterisierung der Kohärenz:

Das Item bewertet den Text hinsichtlich seiner Verständlichkeit, vermittelt über die Kohärenz des Textes. „Kohärenz ist etwas, was der Rezipient eines Textes sich im Zug des Verstehens gedanklich aufbaut und wozu der Text [...] eine mehr oder minder gute Anleitung ist“ (Nussbaumer/ Sieber 1995, S. 20). Es handelt sich dabei um ein umfassendes Kohärenzkonzept (Brinker 2010, S. 17), das nach verschiedenen Aspekten (thematisch, grammatisch) differenziert wird.

Item 1.2 Charakterisierung der Implizitheit des Textes:

Dieses Item bewertet die Stärken und Mängel der textuellen Durchgestaltung von Inhalten, welche trotz dieser textuellen Mängel erkennbar und verstehbar bleiben. „Ein Text gibt mehr zu denken als zu lesen!“ (Nussbaumer/ Sieber 1994, S. 164). Ein guter Text setzt das voraus, was der Rezipient von sich aus ohne großen Aufwand und ohne Verunsicherung an der betreffenden Stelle ergänzen kann. Ein mangelhafter Text sagt zu wenig und ist damit zu knapp, zu implizit, ist sprunghaft oder ‚dunkel‘ (vgl. dazu Nussbaumer/ Sieber 1994).

8 Bei Adamzik (2004) findet sich eine Bestimmung von Textlinguistik, mit der ihr die Aufgabe zugewiesen wird, sich um den normalen Gebrauch der Möglichkeiten eines Systems in Texten zu kümmern. Zu dieser ‚Konventionalitätsrahmung‘ der Textlinguistik heißt es: „Wenn es also in der sog. Systemlinguistik darum geht, die Elemente einer Sprache und die Regeln für deren Verbindbarkeit zu eruieren, um die Frage zu beantworten: Was ist in dieser Sprache möglich?, ist es in der Textlinguistik als Linguistik des Sprachgebrauchs um die Frage zu tun: Wie werden die Möglichkeiten des Systems normalerweise, in einem einzelnen Text oder in einer Gruppe von Texten genutzt?“ (Adamzik 2004, S. 138f.).

Item 1.3 Charakterisierung der Explizitheit des Textes:

Von dem, was ein Text transportieren möchte, sagt er immer nur einen Teil. Entscheidend ist, dass explizit und deutlich ausgeführt wird, was der Rezipient nicht ohne Weiteres ergänzen kann. „Ein in dieser Hinsicht mangelhafter Text sagt [...] zu viel und ist damit zu explizit, zu geschwätzig, ist redundant, ohne Entwicklung und langweilig“ (Nussbaumer/ Sieber 1994, S. 164).

Die Bewertungen zu den Items 1.4 – 1.6 untersuchen, ob das Kind für das, was ihm (kooperativ!) als Schreibabsicht unterstellt wird, „kontextsensitive“ Mittel gefunden hat. Es wird danach gefragt, ob die sprachlichen Mittel dem Text angemessen sind, bzw. dem Usus entsprechen und die Überschrift geeignet ist, auf den Text einzustimmen und zum Textthema hinzuführen. Der Wortschatz sollte dem Genre angepasst, die syntaktischen Mittel unauffällig sein.

Item 1.4 Charakterisierung des Wortschatzes:

Ein Text besitzt einen angemessenen Wortschatz, wenn die Wortwahl unauffällig ist und der Funktion des Textes dient. „Einerseits evoziert in der Entwicklung das Anstreben einer bestimmten Textstruktur bestimmte Formulierungsformative, andererseits eröffnet das Einbringen bestimmter Formulierungen die Möglichkeit, textsortenadäquate Teiltexthe auszdifferenzieren“ (Pohl 2007, S. 66). (Vgl. dazu auch Steinhoff 2011)

Item 1.5 Charakterisierung der sprachlichen Mittel:

Der Text soll danach beurteilt werden, ob die sprachlichen Mittel der Sache angemessen sind, für die sie stehen. Sind die sprachlichen Mittel „funktional besonders gut, in Ordnung oder aber mehr oder weniger dysfunktional eingesetzt worden (kontextsensitiv)?“ (Nussbaumer/ Sieber 1994, S. 166). Die Situationshilfen konzeptioneller Mündlichkeit wie Mimik, Gestik, situativer Kontext und Intonation werden durch spezifisches Vokabular und differenzierte Kohäsionsmittel ersetzt (vgl. Nussbaumer/ Sieber 1994, Koch/ Oesterreicher 1985).

Item 1.6 Angemessenheit der Überschrift

Der Titel informiert über den inhaltlichen Kern des Textes. Die Interpretation der Überschrift ist auf den nachfolgenden Text, genauer auf sein Thema, zu beziehen. Sie verweist auf den Textinhalt, weckt daher vor allem Erwartungen. Wer eine gute Überschrift formulieren will, muss kreativ sein. Und er kann nur kreativ sein, wenn er die möglichen konventionell ausgebildeten Muster kennt (Feilke 2010, S. 2f.). (Vgl. dazu auch Agricola 1979, Große 1976, Harweg 1984, Feilke 2010)

Indem in Item 1.7 nach der Angemessenheit der Textgestaltung im Hinblick auf die pragmatische Funktion des Textes gefragt wird, können die Vertextungsweisen, die bei der Orientierung an Textsorten, an Textstrukturmerkmalen bzw. Genres entstehen, nicht schematisch zum Kriterium für Schreibleistungen werden. Es wird darauf verzichtet, die narrative Qualität der Kindertexte etwa mit der „Story Grammar Scale“ von Harris und Graham (1992) zu bestimmen, indem das Vorhandensein der Strukturelemente erzählender Texte („Person, Ziel, Zeit, Ort, Handlung, Höhepunkt,

Ende“) erfasst wird. Eine solche Verwendung von Strukturelementen begreift diese lediglich als isolierte Merkmale der Texte. Erforderlich ist die Berücksichtigung des genetischen Zugangs von Lernenden zum Text (vgl. dazu auch Augst et al. 2007). Einzuschätzen sind die selbstständige Auswahl und die Kombination von verfügbaren sprachlichen Mitteln der Textgestaltung.

Item 1.7 Orientierung am Grundmuster fiktiver, narrativer Texte:

Die Verwendung formaler Schemata, Muster oder Textsortenspezifika lässt eine gelungene Übernahme von (in der erzählten Welt) vorgefundenen Strukturen erkennen. Fiktionale Texte zeichnen sich dadurch aus, dass zwischen Autor und Leser ein Fiktionalitätsvertrag eingegangen wird, der den Verstehensprozess des Textes steuert. Dieser Fiktionalitätsvertrag kann unter Umständen durch explizite, sprachliche Fiktionalitätssignale (narrative Marker) vom Autor beim Leser eingefordert werden⁹ (vgl. dazu Nickel-Bacon 2002, Augst 2010, Martinez/ Scheffel 2009).

3.3.2 Textualität – unkonventionell

In didaktischer Hinsicht ist die Ausrichtung an den textlinguistischen Normen und Modellen immer dann ein Problem, wenn die Erwerbsperspektive ins Spiel kommt, die es im Hinblick auf Lernprozesse mit der Annäherung an Normen, mit ihrer Durchbrechung, und mit der Subjektivität und Individualität des Schreibens zu tun hat (vgl. dazu Haueis 2006).¹⁰

Die Schreibdidaktik betont, dass die Fähigkeit zur Textgestaltung nicht nur eine Frage der auf Zukünftiges gerichteten Befolgung von Normen ist (die als Ergebnis von textlinguistischen Strukturanalysen formuliert wurden), sondern auch auf Erfahrungen mit der Gestaltung von Texten und mit Gestaltungsprozessen beruht. Aus textuellen Strukturbeschreibungen lassen sich aber nicht unbedingt Prozessbeobachtungen ableiten. Deshalb passt die analytische Sicht auf den Registergebrauch oder den Stil eines Textes nicht zu den vorfindlichen Formen der individuellen Gestaltung. In der Schreibdidaktik gibt es deshalb immer auch das Bestreben, den von der Textlinguistik bestimmten Diskurs zu verlassen und beispielsweise mit schrifttheo-

9 Narrative Grundmuster zeigen sich in der inhaltlichen Struktur eines Textes. Finden sich im Text ein Planbruch (= Bruch der Erwartung), eine Komplikation, Spannung oder eine Pointe? Narrative Marker sind versprachlichte Signale für einen fiktionalen Text: „Es war einmal“ etc.

10 In der Textlinguistik geht es auf diesem Feld um die Variation der Sprachmittel, um deren Neukombination in Abhängigkeit von der Textfunktion, um die Abweichung von konventionellen Textmustern etc. Dazu grenzt die Textlinguistik Kommunikationsbereiche ab und bestimmt mit unterschiedlichen Konzepten, etwa der Abweichungstilistik oder der Funktionalstilistik, Möglichkeiten der Befolgung oder Variation sprachlicher Normen (vgl. dazu Adamzik 2004, S. 68ff.). Dabei sind nicht nur überindividuelle Gestaltungsmuster im Blick, sondern durchaus auch individuelle Gestalt- und Sinnbildungsformen (vgl. z. B. Sandig/ Selting 1997).

retischen, literaturtheoretischen und psychologischen Konzepten die Subjektivität, Individualität und (sogar) Poetizität von Schreiblernprozessen zu modellieren.

In den – durchaus umstrittenen – Konzepten zum Kreativen Schreiben (vgl. z. B. pointiert Fix 2006, S. 118 f.) etwa argumentiert Spinner (2001), dass Schrift kulturgeschichtlich nicht allein das „logische, lineare, analysierende, abstrahierende Denken“ (ebd., S. 116) gefördert habe. Diese einseitige Darstellung übersehe nämlich, dass die ‚zerdehnte Kommunikation‘ (Ehlich) im Schreiben die Möglichkeit und den Raum zur Selbstreflexion und zur Individualität erst schaffe, denn der Schreibende sei für sich allein und habe kein direktes Gegenüber (vgl. Spinner 2001, S. 116f.). In der Schreibdidaktik ist versucht worden, dem Individuierungsaspekt nicht nur mit Konzepten des Kreativen Schreibens, sondern auch mit Modellen des Schreibens als „kultureller Tätigkeit“ (Dehn 1999, Dehn/ Merklinger/ Schüler 2011) oder als Form literarischen Lernens (Kruse 2003 u. 2011) Rechnung zu tragen.

Im KoText-Raster wurde diesem Aspekt der Textproduktion, in vorsichtiger Anlehnung an Kriterien aus dem Zürcher Textanalyseraster, Rechnung getragen. Die Basisdimension 2 (*Textualität – unkonventionell*) umfasst zwei Items zur Frage, inwieweit die Kinder sprachliche oder inhaltliche Wagnisse eingegangen sind.

Item 2.1 Maß des inhaltlichen Wagnisses:

Mit diesem Item werden inhaltliche Wagnisse, Momente des ‚Sich-auf-die-Äste-hinaus-Lassens‘ (Nussbaumer/ Sieber 1994, S. 168), gewürdigt. Inhaltliche Wagnisse sind gekennzeichnet durch eine mutige, ungewöhnliche inhaltliche Gestaltung. Die Handlung und Personen sind nicht klischeehaft, sondern fremd, originell, bizarr, grenzüberschreitend. Ist der Text „gesamthaft lohnend, vorwärtsbringend oder überflüssig?“ (Nussbaumer/ Sieber 1994, S. 168).

Item 2.2 Maß des sprachlich-formalen Wagnisses

Die sprachlich-formale Attraktivität erhöht den Eindruck von der Güte des Textes (Nussbaumer/ Sieber 1994, S. 167f.). Mit diesem Item werden besondere sprachliche Wagnisse ausgezeichnet, selbst dann, wenn sie letztlich missglücken und damit ggf. in einem anderen Item negativ bewertet werden. Sprachlich-formale Wagnisse sind u.a. gekennzeichnet durch ein Spiel mit Sprache, wenn z. B. Erdachtes und Vorgefundenes neu zusammengestellt wird. Positiv zu werten sind die Verwendung lyrischer und poetischer Elemente und unterstützender Intertexte, besondere Tonlagen, ein erkennbarer Rhythmus, Wendungen, Muster, Phrasen, Formeln in außergewöhnlichen Kombinationen, die eine besondere Funktion erfüllen (Nussbaumer/ Sieber 1994).

Die Wagnis-Items dienen der Bewertung unkonventioneller und auffälliger Textmerkmale. Diese Merkmale können oder sollen allerdings nicht von vornherein der „Belastung“ der „Literarität“ oder gar „Poetizität“ ausgesetzt werden. Sie werden vielmehr so verstanden, dass sie die sprachlich-formale und inhaltliche Attraktivität erhöhen und damit die Motivation, den Text zu lesen. Dabei geht es um ungewöhnliche Sprachgestaltungen, die selbst dann eine hohe Wertschätzung erfahren, wenn sie noch nicht in jedem Fall gelungen sind.

4 Stichprobe und Methode

4.1 Stichprobe und Datengrundlage für die vorliegende Studie

Die hier analysierten Schülertexte wurden von 132 Kindern (66 Mädchen und 66 Jungen) verfasst.¹¹ Im Rahmen des Gesamtprojekts verfasste jedes Kind eine Erstfassung seines Textes, überarbeitete diese mithilfe seiner Mitschülerinnen und -schüler und schrieb schließlich die Endfassung der Texte. In dieser Teilstudie wird aus Platzgründen nur die Analyse der Erstfassungen der Texte dargestellt.¹² Auch auf eine Darstellung der offenen Aufgabenstellung und deren Implikationen muss an dieser Stelle verzichtet werden.¹³

4.2 Raterschulung und Beurteilung der Kindertexte

Die Einschätzungen der 132 Kindertexte wurden hoch-inferent vorgenommen; sie erfordern also ein vergleichsweise hohes Maß an Schlussfolgerungen (Lipowsky/ Rakoczy 2006, Hugener et al. 2006). Die Bewertungen der zehn Items reichten jeweils von 1 bis 4, wobei die 1 einer vergleichsweise geringen, die 4 einer hohen Qualitätsstufe entsprach (Lediglich die Wagnis-Items 2.1 und 2.2 weichen von dieser Struktur ab, vgl. Kap. 5.2). Die Stufen der Ratingskala wurden für jedes Item spezifisch definiert.

Um sicherzustellen, dass die Einschätzungen der zehn Items weitgehend unabhängig sind von Merkmalen und Präferenzen des Beurteilers, wurden die Bewertungen der Textqualität von zwei Ratern vorgenommen. Beide Rater studieren das Fach Deutsch für das Lehramt an Grundschulen und wurden aus einem Seminar zur Bewertung und Beurteilung von Schülertexten rekrutiert. Neben der Bereitstellung ei-

11 Diese Teilstichprobe des Projekts KoText unterscheidet sich in Bezug auf die Intelligenz, erfasst mit dem CFT 1 (vgl. Cattell/ Weiß/ Osterland 1997), den Wortschatz, erfasst mit einem Untertest des WWT 6-10 (vgl. Glück 2007) und in Bezug auf das Textverständnis, erfasst mit dem ELFE-Test (vgl. Lenhard/ Schneider 2006) nicht von den übrigen 225 Schülerinnen und Schülern der Gesamtstichprobe (vgl. Einleitung).

12 Weiterführende Analysen zeigen, dass auch die Qualitätseinschätzungen der Endfassungen der Schülertexte eine vergleichbare Faktorenstruktur aufweisen (vgl. Heinzel/ Kruse/ Lipowsky, 2011).

13 Bislang wird Schreibkompetenz in empirischen Studien überwiegend mittels geschlossener Aufgabenstellungen erhoben: „Um die Validität zu maximieren, wird man in Testsituationen immer wieder solche Aufgaben einsetzen, die nur eine spezifische Teilkompetenz erfassen sollen“ (Granter/ Böhme/ Köller 2008, S. 18). Diese Orientierung an der Bestimmung von Teilkompetenzen führt zur Klassifizierung von Fähigkeitskomplexen und Wissensformen, die zur Herstellung von Texten nötig sind (vgl. dazu z. B. Bremerich-Vos 2008), die dann an die Texte als Resultate von Leistungsanforderungen angelegt werden können. Das KoText-Raster will durch eine offene Aufgabenstellung und den Null-Teil, also über die Rahmung des Bewertungsverfahrens, eine Lernerorientierung sichern.

nes ausführlichen Manuals wurden beide Rater vorab intensiv trainiert. Ein hochinferentes Rating, das maßgeblich interpretativen Regeln folgt, erfordert die Erarbeitung einer gemeinsamen Textverständnis- und Bewertungsgrundlage. Das Training¹⁴ erstreckte sich insgesamt über einen Zeitraum von vier Wochen (ca. 25 Stunden Schulung + 25 Std. Eigenstudium).

Nach einer ersten Schulungsphase und einer ersten Pilotierung wurden einzelne, noch unsichere Items im Manual spezifiziert und nachgeschult. Erst nach einer erfolgreichen zweiten Pilotierung – wieder mit Texten, die nicht aus der Stichprobe der untersuchten Kinder, aber aus dem gleichen Projekt stammten¹⁵ – erfolgte das maßgebliche Rating. Die Interraterreliabilität wurde dabei fortlaufend kontrolliert. Die Beurteilung und Bewertung eines Textes benötigte je nach Textlänge 10-20 min.

4.3 Analysemethode

Die Überprüfung der Faktorenstruktur der Qualitätseinschätzungen fand mittels einer konfirmatorischen Faktorenanalyse mit der Statistiksoftware AMOS statt. Zur Parameterschätzung kam die klassische ML-Diskrepanzfunktion (maximum-likelihood) zum Einsatz. Dieses Schätzverfahren setzt zwar grundsätzlich (multivariat-)normalverteilte Daten voraus, gilt aber auch bei moderater Verletzung dieser Annahme noch als robust (vgl. Jahn 2007, S. 12ff.). Gegen die Verwendung alternativer Schätzverfahren, wie beispielsweise ADF (asymptotically-distribution-free) oder ULS (unweighted-least-squares), die keine Voraussetzungen gegenüber der Verteilung der Daten machen, sprechen gewichtige Nachteile: Zum einen liefern diese Verfahren im Vergleich zu ML weniger Gütemaße, die zur besseren Beurteilung des geschätzten Modells notwendig sind, zum anderen sind teils wesentlich größere Stichproben (z. B. $N \geq 5000$ bei ADF) erforderlich (vgl. ebd.).

Zur Beurteilung der Anpassungsgüte eines theoretischen Modells an die empirischen Daten dient eine Reihe von sogenannten Fit-Indizes. Eine zufriedenstellende Modellgüte setzt folgende Werte voraus: Der Quotient (χ^2/df) sollte $\leq .2.5$ sein, der χ^2 -Wert sollte nicht signifikant sein. Da der Chi-Quadrat-Wert stichprobenabhängig ist, sollten weitere Indizes zur Beurteilung der Modellgüte herangezogen werden. Der Tucker Lewis Index (TLI), der Comparative Fit Index (CFI), der Normed Fit Index (NFI) und der Root Mean Square Error Of Approximation (RMSEA) gelten als rela-

14 Die Schulung folgte dabei den Empfehlungen Langers und Schulz von Thuns (2007, S. 137f.). Sie umfasste ein a) Diskriminationstraining: Erkennen von Skalenausprägungsunterschieden, indem mehrere Objekte vorgelegt werden, die in eine Rangfolge zu bringen sind. b) Herstellungstraining: Ein Text mit bekannter Skalenausprägung wird vorgegeben. Die Rater sollen überlegen, wie der Text aussehen würde, wenn er eine andere, vorgeschriebene Ausprägung erhielte. Und schließlich ein c) Konzeptunterscheidungstraining: Elemente eines Textes müssen einer entsprechenden Kategorie zugeteilt werden.

15 Für die eigentliche Untersuchung wurden Gruppen nach bestimmten Kriterien der sozialen Kohäsion zusammengesetzt (vgl. Ludwig/ Herrmann/ Lipowsky/ Heinzl/ Kruse/ Eckermann (eingereicht)).

tiv unabhängig vom Stichprobenumfang. Der TLI, der NFI und der CFI sollten jeweils Werte $\geq .90$ annehmen, der RMSEA sollte $\leq .08$ sein (vgl. Weiber/ Mühlhaus 2010).

5 Ergebnisse

5.1 Interraterreliabilität

Als Maß für die Interraterreliabilität wurde der relative Generalisierbarkeitskoeffizient herangezogen. Dem Koeffizienten liegt ein varianzanalytisches Vorgehen zugrunde, das auf der Generalisierbarkeitstheorie (vgl. Brennan 2001, Cronbach et al. 1972, Cronbach/ Nageswari/ Gleser 1963) basiert. Die erforderlichen Berechnungen wurden mit dem Programm für Generalisierbarkeitsstudien Version 2.0 D von Ysewijn (1997) durchgeführt. Der Koeffizient kann Werte zwischen 0 und +1 annehmen. Je höher der jeweilige Wert liegt, desto höher ist auch die relative Übereinstimmung der beiden Rater. Für das Gesamtkorpus der Stichprobentexte ergaben sich Werte zwischen $\rho^2 = .87$ (für das Item 1.4 ‚Charakterisierung des Wortschatzes‘) und $\rho^2 = .97$ (für das Item 1.7 ‚Orientierung am Grundmuster fiktiver, narrativer Texte‘). Die Werte können somit als sehr zufriedenstellend betrachtet werden.

5.2 Textqualität

Da alle Schülertexte von jeweils zwei Ratern beurteilt wurden und ein hoher Grad an Übereinstimmung in den Urteilen gegeben war, konnten beide Urteile durch Bildung eines Mittelwerts zusammengefasst werden.

Item	N	Range	M	SD
holistisches Urteil	132	1-4	2.78	.71
Kohärenz			2.98	.79
Implizitheit			2.95	.79
Explizitheit			2.94	.78
Wortschatz			3.39	.68
sprachliche Mittel			2.85	.61
Überschrift			2.07	.71
narratives Grundmuster			2.38	.76
inhaltliches Wagnis			1.97	.89
sprachliches Wagnis			2.11	.87

Tab.3: Deskriptive Statistiken zu den eingeschätzten Textqualitätsitems der Erstfassungen

Tabelle 3 gibt eine Übersicht über die von den Kindern in ihren Erstfassungen erreichten Werte. Wie die Tabelle zeigt, wurden die Texte in den meisten Items mit einem Wert, der über dem theoretischen Mittelwert von 2.5 liegt, beurteilt.

Zu beachten ist, dass die Mittelwerte der einzelnen Items nicht unmittelbar miteinander verglichen werden können, denn die jeweiligen Ausprägungen des Ratingformats der einzelnen Kategorien haben teilweise unterschiedliche Bedeutungen. Die Items ‚sprachliches Wagnis‘ und ‚inhaltliches Wagnis‘ erzielten beispielsweise geringe Werte, so dass man auf Grund der Mittelwerte annehmen könnte, dass die Qualität der Texte im Bereich des Faktors Unkonventionalität gering ausgeprägt ist. Hier ist auf die Konstruktion der Wertebestimmung bei der Skalierung zu verweisen. Die geringen Werte der Items ‚sprachliches Wagnis‘ ($M = 2.11$; $SD = .87$) und ‚inhaltliches Wagnis‘ ($M = 1.97$; $SD = .87$) sind darauf zurückzuführen, dass diese beiden Kriterien besondere sprachliche bzw. inhaltliche Gestaltungen auszeichnen wollen. Der Wert 1 verhält sich neutral und wird, im Gegensatz zu den anderen Items, nicht als Negativwertung vergeben.

5.3 Validität des Messverfahrens

Mit der Übernahme von neun der insgesamt 15 Items des ZTR war auch dem KoText-Analyseraster ein lernerorientiertes, kognitiv-pragmatisches Textmodell zu Grunde gelegt, unter Annahme zweier Faktoren: einem ‚Konventionalitätsfaktor‘ (mit sieben Items, s. o.) und einem ‚Unkonventionalitätsfaktor‘ (mit zwei Items, s. o.). Unberücksichtigt blieb bei der Modellbildung das Item ‚holistisches Urteil‘, weil es einen anderen Auflösungsgrad besitzt als die anderen Qualitätsitems. Unter diesen Voraussetzungen wurde gefragt, ob das in dieser Studie angenommene zweifaktorielle Textmodell zur Abbildung der Qualität von Lernertexten geeignet ist.

Die Analyse des angenommenen Zweifaktorenmodells ergab jedoch, dass unter Beibehaltung aller neun Items sämtliche Gütemaße (vgl. Tab. 4) außerhalb der in der einschlägigen Literatur (z. B. Weiber/ Mühlhaus 2010, S. 159ff.) geforderten Wertebereiche lagen. Damit bildete dieses Modell die empirische Realität aus statistischer Sicht nicht angemessen ab.

Modell	RMSEA	TLI	CFI	NFI	χ^2	df	χ^2/df	p
Erstfassungen	.135	.837	.882	.843	87.69	26	3.37	.000

Tab. 4: Gütemaße für das Faktorenmodell zur Textqualität

Die nicht ausreichende Empirieadäquatheit führte zu einer Modifikation des Textmodells, bei der die Items ‚Überschrift‘, ‚narratives Grundmuster‘ und ‚sprachliche Mittel‘ aus dem Modell entfernt wurden. Das modifizierte Strukturgleichungsmodell mit den beiden Faktoren ‚Textualität – konventionell‘ (Faktor 1) und ‚Textualität – unkonventionell‘ (Faktor 2) gestaltet sich jetzt wie folgt:

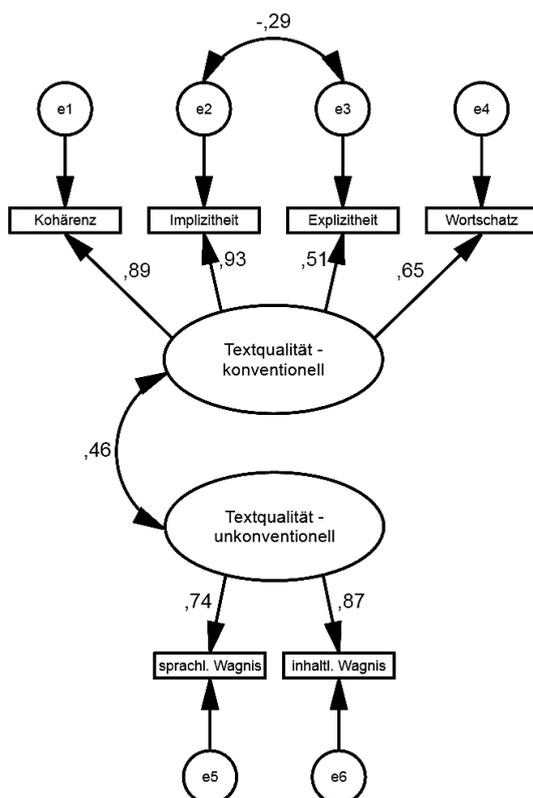


Abb. 1: Modifiziertes Faktorenmodell zur Textqualität (Erstfassungen) mit standardisierten Faktorladungen nach erfolgter ML-Schätzung.

Für diese Modifikation waren folgende Erwägungen ausschlaggebend: Die schlechte Passung des Items ‚sprachliche Mittel‘ kann auf die intendierte Standardisierung der Erhebungs- und Auswertungssituation zurückgeführt werden. Die Kindertexte wurden, wie oben beschrieben, syntaktisch und orthographisch korrigiert. Dies führte in gewisser Weise zu einer Art Nivellierung, denn schlechte Texte erfuhren dadurch in dieser Dimension eine Aufwertung, gute Texte blieben unverändert. Der Blick auf die deskriptive Statistik der einzelnen Items (s. Tab. 3) kann diese Vermutung erhärten, denn die Varianz für die sprachlichen Mittel ist mit $SD = .61$ von allen zehn Items am geringsten.

Auch das Item ‚narratives Grundmuster‘ wurde aus dem Modell entfernt, da nicht alle Kinder narrative Texte verfassten. Angestrebt war bei der Konstruktion des Rasters, dass in den anderen Items prinzipiell auch für Sachtexte eine hohe Bewertung möglich sein sollte. Sachtexte konnten im Item ‚narrative Grundmuster‘ jedoch nur niedrige Werte erhalten, so dass hier eine schlechte Passung von Rasterkonstruktion und Auswertungsmodell gegeben ist.

Für den Ausschluss der ‚Überschrift‘ aus dem Modell sprachen u.a. empirische Aspekte. Die Ladung der Dimension ‚Angemessenheit der Überschrift‘ auf dem Konventionalitätsfaktor war vergleichsweise gering (.28). Im Ratingprozess wurden fehlende Überschriften und unangemessene Überschriften gleichermaßen mit dem geringsten Wert (1) bewertet. Während aber unangemessene Überschriften als ein Indikator für geringe Qualität der sprachlichen Äußerungen auf der globalen Ebene des Textes betrachtet werden können, muss dies für fehlende Überschriften nicht zwingend gelten. Möglicherweise haben einige Schülerinnen und Schüler die Überschrift auch nur vergessen. Somit repräsentierte der Wert ‚1‘ mit unangemessenen und fehlenden Überschriften sehr heterogene Urteile.

Aus den dargestellten Überlegungen resultiert also das modifizierte Strukturgleichungsmodell (Abb. 1) mit den beiden Faktoren ‚Textualität – konventionell‘ (Faktor 1) und ‚Textualität – unkonventionell‘ (Faktor 2).

Der latente Faktor ‚Textqualität – konventionell‘ repräsentiert dabei die Kohärenz, die Implizitheit, die Explizitheit und den Wortschatz. Die Faktorladungen liegen zwischen $\lambda=.51$ und $\lambda=.93$. Der zweite latente Faktor ‚Textqualität – unkonventionell‘ wird durch die Indikatoren sprachliches und inhaltliches Wagnis operationalisiert, die Faktorladungen betragen hier $\lambda=.74$ und $\lambda=.87$. Beide Textqualitätsfaktoren hängen messfehlerbereinigt zu $r=.46$ zusammen, sind also hinreichend unterschiedlich. Die für die Erstfassungen ermittelten Fit-Indizes zeigen, dass das neue, modifizierte Modell zur Textqualität insgesamt als der empirischen Realität gut angepasst eingestuft werden kann (vgl. Weiber/ Mühlhaus 2010, S. 159ff.):

Modell	RMSEA	TLI	CFI	NFI	χ^2	df	χ^2/df	p
Erstfassungen	.059	.980	.991	.971	10.14	7	1.45	.181

Tab. 5: Gütemaße für das modifizierte Faktorenmodell zur Textqualität (Erstfassungen)

Die Tabelle zeigt, dass alle zur Beurteilung des Gesamtfits des Modells herangezogenen Gütemaße die üblicherweise herangezogenen Cut-Off-Kriterien erfüllen. (RMSEA ≤ 0.08 , TLI ≥ 0.90 , NFI ≥ 0.90 , CFI ≥ 0.90 , $\chi^2/df \leq 3.0$; vgl. Weiber/ Mühlhaus 2010, S. 176).

6 Diskussion

Die Ergebnisse zeigen, dass ein quantitativ ausgerichtetes Bewertungsverfahren reliabel die Textqualität erfassen kann und sich dabei mit einer lernerorientierten Sichtweise unter bestimmten Bedingungen verbinden lässt. Ein angemessenes Textmodell konnte allerdings nur dadurch erreicht werden, dass Items aus dem ursprünglichen Modell, das dem Bewertungsprozess unterlegt war, herausgenommen wurden. Das modifizierte Textmodell muss ohne die bedeutenden Items ‚Charakterisierung der sprachlichen Mittel‘, ‚Orientierung am Grundmuster fiktiver, narrativer

Texte‘ und ‚Angemessenheit der Überschrift‘ auskommen. Dafür gab es zwar gute Gründe (s. o.), die aber zu Lasten der Qualitätsperspektive des Rasters gehen. So ist zweifellos die narrative Struktur eines Textes ein wesentliches Kriterium für die Qualität von Lernertexten. Allerdings ist auch bekannt, dass von außen angelegte abstrakte Ordnungsschemata sich in der Grundschule als wenig flexibel in Bezug auf die Ordnungsschemata erweisen, welche die Kinder auf Grund der Aufgabenstellung tatsächlich verwendet haben (vgl. dazu auch Dehn/ Merklinger/ Schüler 2011). Hinzu kommt, dass es für das wegfallende Item „Überschrift“ auch ‚schreiblernerische‘ Gründe geben kann, denn die geringen Ladungen im ursprünglichen Faktorenmodell lassen darauf schließen, dass die Kinder mit den literalen Besonderheiten einer Überschrift Schwierigkeiten haben. Überschriften – darauf hat Feilke (2010) verwiesen – gibt es im Mündlichen nicht. Auf Grund ihrer besonderen Anforderungen an die Typographie und an ihre sprachliche Markiertheit sind sie oft grammatisch irregulär. Für die nötigen literalen Prozeduren konnten die Kinder möglicherweise im Schreibunterricht noch keine Schemata ausbilden.

Dennoch bleibt festzuhalten, dass mit den im modifizierten Faktorenmodell für den ersten Faktor (‚Textualität – konventionell‘) verbliebenen Items ‚Kohärenz‘, ‚Implizitheit‘, ‚Explizitheit‘ und ‚Wortschatz‘ die intendierte Erfassung der Textqualität hinsichtlich kommunikativer und funktionaler Anforderungen in Teilen gelungen ist. Die Items des Faktors bewerten, inwieweit es den Kindern geglückt ist, im Medium der Schriftlichkeit einen ‚Text für Andere‘ zu entwerfen. Sie durchlaufen im Schreiben kognitive, emotionale sowie imaginative Prozesse und müssen entscheiden, welche textuellen Mittel beim Schreiben angemessen sind. Um einen kohärenten Text zu erzeugen, muss das Geschehen einer Handlungslogik folgen; der Text darf nicht zu brüchig oder sprunghaft sein. Das Schreibprodukt wird danach beurteilt, inwieweit die Voraussetzungen erfüllt sind, eine konsistente Textwelt entstehen zu lassen, und das Textthema erkennbar ist.

Der zweite Faktor ‚Textualität – unkonventionell‘ untersucht, in welchem Maße der Rezipient darüber hinaus motiviert wird, den Text zu lesen oder ihm zuzuhören. Dazu gehört die Frage, inwieweit inhaltliche und sprachliche Wagnisse eingegangen wurden, indem neue und ungewöhnliche Ideen und Sprachmuster entwickelt und ausprobiert oder befremdliche Sujets verwendet wurden. Doch auch wenn der Faktor 2 empirisch bestätigt werden konnte, bleiben Fragen zur Tragfähigkeit bestehen. Denn zweifellos ist in der Perspektive der Unkonventionalität Schreiben mehr als das Eingehen von Wagnissen, es zieltauch auf die Artikulation ästhetischer Erfahrungen (Dehn 1999, S. 77). Nach Jakobson (1972) dominiert in poetischen Texten die poetische Funktion der Sprache im Vergleich zur referentiellen Funktion. Poetische Elemente verweisen nicht eindeutig auf etwas, sondern lassen sich in mehrfacher Weise interpretieren. „Der Vorrang der poetischen Funktion vor der referentiellen löscht die Referenz nicht aus, sondern macht sie doppeldeutig“ (ebd., S. 127). Gerade diese Doppeldeutigkeit stellt jedoch für die empirische Forschung eine große Herausforderung dar, weshalb auch im KoText-Raster die Operationalisierung unbe-

stimmter, poetischer Textmerkmale nicht geleistet werden konnte.¹⁶ Da der Faktor ‚Textualität – unkonventionell‘ bislang nur über zwei Items verfügt, wäre perspektivisch eine Erweiterung um darüber hinaus gehende Aspekte der Unkonventionalität von Texten wünschenswert.

So bleibt festzuhalten, dass die Aufforderung von Bremerich-Vos/ Possmayer (2011, S. 46), weiterführende Studien hinsichtlich der *Validität* von Schülertext-Beurteilungsrastern durchzuführen, nach Erkenntnissen der vorliegenden Studie gerechtfertigt ist. Wenngleich das hier vorgestellte zweifaktorielle Textmodell bereits sechs aussagekräftige Items zur Abbildung von Textqualität enthält, bleiben doch Unklarheiten über die Reichweite bestehen.

Nachfolgende Analysen sollten prüfen, ob und in welcher Konstellation die drei entfernten Items in das Textmodell wieder integriert werden können, wenn a) die Originaltexte (ohne die rechtschriftlichen und v. a. grammatischen Korrekturen) neu geratet werden und b) ausschließlich narrative Texte mit vorhandenen Überschriften in die Berechnungen einfließen. Auf diesem Wege könnten zumindest Aussagen über die Validität eines Analyserasters zur Bestimmung der *narrativen* Textqualität untermauert werden. Auch eine Prüfung, inwieweit ein Rating der Leistungen im Bereich Rechtschreibung und Sprachgebrauch (anhand der Originale) möglicherweise zu einem einzelnen/ weiteren Faktor (vgl. Ergebnisse von Hartmann/ Lehmann 1987, Neumann 2007) führt, sollte Inhalt weiterer Untersuchungen sein.

Ogleich die in dieser Studie zu beurteilenden zehn Items einen vergleichsweise hohen Interpretationsspielraum zuließen, fielen die Interraterreliabilitäten im vorliegenden Projekt sehr zufriedenstellend aus. Wir führen dies u. a. auf den vorgelagerten Null-Teil des Rasters, der ein steuerndes Element im Bewertungsprozess darstellt, auf die ausführliche Formulierung der jeweiligen Grundidee, die Definition der vier Stufen des Antwortformats sowie auf die umfangreiche und intensive Schulung der Rater und deren Fachexpertise zurück. Die Schulung der Rater mit Erläuterung der Fehlermöglichkeiten, regelmäßigen Rückmeldungen und gemeinsamer Diskussion über die abgegebenen Raterurteile anhand von Übungsbeispielen scheint unverzichtbar. Dieses Vorgehen bewirkt zugleich, dass Dimensionen bereits im Entwicklungsprozess „objektiviert“ und auf ihre empirische Güte geprüft werden. Es stellt sicher, dass ein gemeinsames Text- und Beurteilungsverständnis hinsichtlich der theoretischen Grundlagen entwickelt wird. Mit der herausgehobenen Bedeutung des – nicht in die Bewertung einfließenden – Null-Teils für das Erreichen tragfähiger und reliabler Ergebnisse muss zugleich dessen präzise Profilierung einhergehen, da er in hohem Maße geeignet scheint, den Bewertungsprozess zielgerichtet zu kontrollieren.

16 Vgl. für die Textverstehenskompetenz dazu aber das Projekt „Literarästhetische Urteilskompetenz“ Frederking (2008a) und Frederking et al. (2008b)

Literatur

- Adamzik, K. (2004). *Textlinguistik. Eine einführende Darstellung*. Tübingen: Niemeyer.
- Agricola, E. (1979). *Textstruktur – Textanalyse – Informationskern*. Leipzig: VEB Verlag Enzyklopädie.
- Augst, G. (2010). Zur Ontogenese der Erzählungskompetenz in der Primar- und Sekundarstufe. In: Pohl, T./ Steinhoff, T. (Hg.): *Textformen als Lernformen*. Duisburg: Gilles & Francke. S. 63-96.
- Augst, G./ Disselhoff, K./ Henrich, S./ Pohl, T./ Völzing, P. (2007). *Text-Sorten-Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter*. Frankfurt/M. et al.: Lang.
- Baurmann, J./ Pohl, T. (2009). Schreiben – Texte verfassen. In: Bremerich-Vos, A./ Granzer, D./ Behrens, U./ Köller, O. (Hrsg.): *Bildungsstandards für die Grundschule: Deutsch konkret*. Berlin: Cornelsen. S. 75-103.
- Brinker, K. (2010). *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden*. 7. Auflage. Berlin: Erich Schmidt Verlag.
- Blatt, I./ Ramm, G./ Voss, A. (2009). Modellierung und Messung der Textkompetenz im Rahmen einer Lernstandserhebung in Klasse 6. In: *Didaktik Deutsch*, H. 26, S. 54-81.
- Blatt, I./ Voss, A./ Matthießen, I. (2005). Kinder schreiben Briefe aus der Zukunft. Qualitative Analysen von Kindertexten und fachdidaktische Diskussion. In: Bos, W./ Lankes, E./ Prenzel, M./ Schwippert, K./ Valtin, R./ Walther, G. (Hg.): *IGLU. Vertiefende Analyse zu Leseverständnis; Rahmenbedingungen und Zusatzstudien*. Münster: Waxmann. S. 109-158.
- Bremerich-Vos, A. (2008). Benjamin S. Bloom (und andere) revisited. In: Bremerich-Vos, A./ Granzer, D./ Köller, O. (Hrsg.). *Lernstandsbestimmung im Fach Deutsch. Gute Aufgaben für den Unterricht*. Weinheim und Basel: Beltz. S. 29-49.
- Bremerich-Vos, A./ Possmayer, M. (2011). Zur Reliabilität eines Modells der Entwicklung von Textkompetenz im Grundschulalter. In: *Didaktik Deutsch* H. 31, S. 30-49.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Cattell, R. B./ Weiß, R. H./ Osterland, J. (1997). *Grundintelligenztest – CFT. Skala 1*. Göttingen: Westermann Hogrefe.
- Cronbach, L. J./ Gleser, G. C./ Nanda, H./ Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- Cronbach, L.J./ Nageswari, R./ Gleser, G.C. (1963). Theory of generalizability: A liberation of reliability theory. In: *The British Journal of Statistical Psychology* 16, S. 137-163.
- Dehn, M. (1999). *Texte und Kontexte. Schreiben als kulturelle Tätigkeit in der Grundschule*. Berlin: Volk und Wissen.
- Dehn, M./ Merklinger, D./ Schüler, L. (2011). *Texte und Kontexte. Schreiben als kulturelle Tätigkeit in der Grundschule*. Seelze: Klett Kallmeyer.
- Ehlich, K. (1983). Text und sprachliches Handeln. Die Entstehung von Texten aus dem Bedürfnis nach Überlieferung. In: Assmann, A./ Assmann J./ Christof Hardmeier, C. (Hrsg.): *Schrift und Gedächtnis. Beiträge zur Archäologie der literarischen Kommunikation*. München: Fink. S. 24-43.
- Ehlich, K. (2005). Sind Bilder Texte? In: *Der Deutschunterricht* 57, H. 4, S. 51-60.

- Feilke, H. (1993). Schreibentwicklungsforschung. Ein kurzer Überblick unter besonderer Berücksichtigung der Entwicklung prozessorientierter Schreibfähigkeiten. In: Diskussion Deutsch, H. 129, S. 17-34.
- Feilke, H. (2010). „Aller guten Dinge sind drei“ – Überlegungen zu Textroutinen & literalen Prozeduren. In: Fest-Platte für Gerd Fritz. Hg. und betreut von Iris Bons, Thomas Gloning und Dennis Kaltwasser. Gießen 17.05.2010.URL: http://www.festschrift-gerd-fritz.de/files/feilke_2010_literale-prozeduren-und-text-routinen.pdf.
- Fitzgerald, J. (1987). Research of revision in writing. In: Review of Educational Research, 57(4). S. 481-506.
- Fix, M. (2006). Texte schreiben. Schreibprozesse im Deutschunterricht. Paderborn: Schöningh.
- Frederking, V. (2008a). Literarästhetische Urteilskompetenz. Probleme und Möglichkeiten einer Operationalisierung ästhetischer Kompetenz. In: Frederking, V./ Hu, A./ Schreiber, W./ Zwergel, A./ Methfessel, E. (Hg.): Schwer operationalisierbare Kompetenzen. Herausforderungen empirischer Fachdidaktik. Baltmannsweiler: Schneider-Hohengehren, S. 36-64.
- Frederking, V./ Meier, C./ Stanat, P./ Dickhäuser, O. (2008b). Ein Modell literarästhetischer Urteilskompetenz. In: Didaktik Deutsch 25, S. 11-31.
- Glück (2007). Wortschatz- und Wortfindungstest für 6- bis 10-Jährige. WWT 6–10. München: Elsevier, Urban & Fischer.
- Granzer, D./ Böhme, K./ Köller, O. (2008). Kompetenzmodelle und Aufgabenentwicklung für die standardisierte Leistungsmessung im Fach Deutsch. In: Bremerich-Vos, A./ Granzer, D./ Köller, O. (Hrsg.): Lernstandsbestimmung im Fach Deutsch. Weinheim: Beltz. S. 10-28.
- Große, E.U. (1976). Text und Kommunikation. Eine linguistische Einführung in die Funktionen von Texten. Stuttgart: Kohlhammer.
- Hartmann, W./ Lehmann, R.H. (1987). The Hamburg Study of Achievement in Written Composing. National Report for the IEA International Study of Achievement in Written Composing. Part I: Methods and Findings. Hamburg: University.
- Harweg, R. (1984). Initialsätze und Überschriften. Bemerkungen zur Struktur von Textanfängen. In: Poetica (19). S. 63-90.
- Harris, K./ Graham, S. (1992). Helping young writers master the craft: Strategy instruction and self-regulation in the writing process. Cambridge: Brookline Books.
- Haueis, E. (2006). Formen schriftlicher Texte. In: Bredel, U./ Günther, H./ Klotz, P./ Ossner, J./ Siebert-Ott, G. (Hg.): Didaktik der deutschen Sprache. Band I. Paderborn: Schöningh. S. 224-236.
- Heinzel, F./ Kruse, N./ Lipowsky, F. (2011). Arbeitsbericht an die DFG zum Projekt „Kooperative Schülerrückmeldungen im Deutschunterricht der Grundschule“. Universität Kassel.
- Hugener, I./ Rakoczy, K./ Pauli, C./ Reusser, K. (2006). Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lernprozesse. In: Rahm, S./ Mammes, I./ Schratz, M. (Hrsg.): *Schulpädagogische Forschung* (Band 1). *Unterrichtsforschung - Perspektiven innovativer Ansätze*. Innsbruck: Studienverlag. S. 41-53.

- Jahn, S. (2007). Strukturgleichungsmodellierung mit LISREL, AMOS und SmartPLS. Eine Einführung. Wirtschaftswissenschaftliches Diskussionspapier (WWDP 86/07), Fakultät für Wirtschaftswissenschaften, Technische Universität Chemnitz.
- Jakobson, R. (1972). Linguistik und Poetik. In: Ihwe, J. (Hrsg.) Literaturwissenschaft und Linguistik. Ergebnisse und Perspektiven. Frankfurt/M.: Athenäum. S. 142-178.
- Koch, P./ Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: Romanistisches Jahrbuch, 36, S. 15-43.
- Kruse, N. (2003). Literarisches Lernen in der Grundschule – Textschreiben als Herausforderung literarischen Lernens bei Kindern. In: Brinkmann, E./ Osburg, C./ Kruse, N. (Hrsg.). Kinder schreiben und lesen. Beobachten – Verstehen – Lehren, Freiburg: Fillibach, S. 233-254.
- Kruse, N. (2011). Zur Poetik in Kindertexten in schulischen Lernprozessen. Ein Vorschlag zur Erweiterung des sprachlichen Rahmens. In Kohl, E.-M./ Ritter, M. (Hrsg.): Die Stimmen der Kinder. Kindertexte in Forschungsperspektiven. Baltmannsweiler: Schneider-Hohengehren, S. 73-82.
- Langer, J. A. (1986). Children reading and writing: Structures and strategies. Norwood, NJ: Ablex.
- Langer I./ Schulz von Thun, F. (2007). Messung komplexer Merkmale in Psychologie und Pädagogik - Ratingverfahren. Waxmann: Münster.
- Lehmann, R.H. (1988). Reliabilität und Generalisierbarkeit der Aufsatzbeurteilung im Rahmen des Hamburger Beitrags zur internationalen Aufsatzstudie der IEA. In: Empirische Pädagogik. Zeitschrift zur Theorie und Praxis erziehungswissenschaftlicher Forschung. Heft 2 (4), S. 349-365.
- Lenhard, W./ Schneider, W. (2006). ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler. Göttingen: Hogrefe.
- Lipowsky, F./ Rakoczy, K. (2006). Videoanalysen in der Unterrichtsforschung. Workshop auf der DGfE-Sommerschule in Ludwigsfelde. 05./06.9.2006.
- Ludwig, M./ Herrmann, M./ Lipowsky, F./ Heinzel, F./ Kruse, N./ Eckermann, T. (eingereicht). Textrevisionen im Grundschulunterricht – Welchen Einfluss haben die Lernumgebung und die soziale Kohäsion der Gruppe?
- Martinez, M./Scheffel, M. (2009). Einführung in die Erzähltheorie. 8. Auflage. München: Beck.
- Neumann, A. (2007). Briefe schreiben in Klasse 9 und 11. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen. Münster/New York/München/Berlin: Waxmann-Verlag.
- Neumann, A./ Lehmann, R. H. (2008). Schreibkompetenz. In: DESI-Konsortium (Hg): Unterricht und Kompetenzerwerb in Deutsch und Englisch. Weinheim/Basel: Beltz.
- Nickel-Bacon, I./ Groeben, N./ Schreier, M. (2000). Fiktionssignale pragmatisch. Ein medienübergreifendes Modell zur Unterscheidung von Fiktion(en) und Realität(en). In: Poetica H. 3-4, S. 267-299.
- Nussbaumer, M. (1991). Was Texte sind und wie sie sein sollen. Ansätze zu einer sprachwissenschaftlichen Begründung eines Kriterienrasters zur Beurteilung von schriftlichen Schülertexten. Tübingen: Niemeyer.
- Nussbaumer, M./ Sieber, P. (1994). Texte analysieren mit dem Zürcher Textanalyseraster. In: Sieber, P. (Hg.): Sprachfähigkeiten – Besser als ihr Ruf und nötiger denn je! Aarau: Sauerländer. S. 141-186.

- Nussbaumer, M./ Sieber, P. (1995). Was sich in Abituriententexten zeigt. Ergebnisse aus dem Zürcher „Sprachfähigkeiten“-Projekt. In: Diskussion Deutsch, H. 141, S. 15-24.
- Pohl, T. (2007). Emotionalität im frühen Schreiben – Von emotionaler Involviertheit zu emotionaler Involvierung. In: Becker-Mrotzek, M./ Schindler, K. (Hg.): Text schreiben. Duisburg: Gilles & Francke. S. 63-80.
- Sandig, B./ Selting, M. (1997). Sprech- und Gesprächsstile. Berlin; New York: de Gruyter.
- Spinner, K. (2001). Kreativer Deutschunterricht: Identität – Imagination – Kognition, Seelze: Kallmeyer.
- Spitta, G. (1992). Schreibkonferenzen in Klasse 3 und 4. Ein Weg vom spontanen Schreiben zum selbstbewussten Verfassen von Texten. Frankfurt/M.: Cornelsen Scriptor.
- Steinhoff, T. (2010). Differenzierte Schülertextbeurteilung. Entwicklungs-, Prozess- und Situierungsdimension. In: Pohl, T./ Steinhoff, T. (Hrsg.). Textformen als Lernformen. Duisburg: Gilles & Francke. S. 257-280.
- Steinhoff, T. (2011). Textwortschatzerwerb im Sprachvergleich. Lexikalisches Lernen deutscher und italienischer Grundschüler am Beispiel schriftlicher Zimmerbeschreibungen. In: Foschi A./ Hepp, M./ Neuland, E./ Dalmas, M. (Hrsg.): Text und Stil im Kulturvergleich. Pisaner Fachtagung 2009 zu interkulturellen Wegen Germanistischer Kooperation. München: Iudicium. S. 192-214. (gemeinsam mit Daniela Sorrentino)
- Weiber, R./ Mühlhaus, D. (2010). Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS. Springer: Heidelberg.
- Ysewijn, P. (1997). GT -Programm für Generalisierbarkeitsstudien. Neuchâtel: Institut de recherche et de documentation pédagogique. Verfügbar unter: <http://www.irdp.ch/methodo/generali.htm> [Stand: 01.08.2011]

Anschrift der Verfasser und Verfasserinnen:

Prof. Dr. Norbert Kruse, Universität Kassel, Institut für Germanistik, Kurt-Wolters-Str. 5, 34125 Kassel.

norbert.kruse@uni-kassel.de

Prof. Dr. Friederike Heinzl, Prof. Dr. Frank Lipowsky, Maik Herrman, Universität Kassel, Institut für Erziehungswissenschaft, Nora-Platiel-Str. 1, 34109 Kassel.

heinzl@uni-kassel.de, lipowsky@uni-kassel.de, maik.herrmann@uni-kassel.de

Anke Reichardt, Goethe-Universität Frankfurt am Main, Institut für Psycholinguistik und Didaktik der deutschen Sprache, Grüneburgplatz 1, 60323 Frankfurt am Main
Reichardt@em.uni-frankfurt.de