

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>
14. Jahrgang 2009 – ISSN 1431-4355
Schneider Verlag Hohengehren
GmbH

Inge Blatt/Gesa Ramm/Andreas Voss

**MODELLIERUNG UND MESSUNG
DER TEXTKOMPETENZ IM
RAHMEN EINER
LERNSTANDSERHEBUNG IN
KLASSE 6 (2008)**

In: Didaktik Deutsch. Jg. 14. H. 26. S. 54-81.

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.

Inge Blatt/Gesa Ramm/Andreas Voss:

MODELLIERUNG UND MESSUNG DER TEXTKOMPETENZ IM RAHMEN EINER LERNSTANDSERHEBUNG IN KLASSE 6 (2008)

Einleitung: Erfassung der Textkompetenz zur Lernstandserhebung

Die Beurteilung von Schülertexten stellt eine herausfordernde Forschungsaufgabe dar, die eine hohe Bedeutung für die Praxis besitzt. Sie bleibt trotz einschlägiger Studien bis heute noch weitgehend ein Forschungsdesiderat (Nutz 2003). Es fehlen empirisch abgesicherte Ergebnisse auf die Fragen: Was bestimmt die Qualität von Texten und mit welchen Kriterien kann diese erfasst werden?

Die Diskussion um die Willkür der Aufsatznoten in den 70er Jahren führte zu einer Ablehnung der Noten und zu einer Abwendung der deutschdidaktischen Forschung von diesem Thema (Ivo 1982). In den Schulen wurde die herkömmliche Aufsatzkorrektur tradiert, die nach den Ergebnissen der Hamburger Aufsatzstudie - einer 1985 im Rahmen der international vergleichenden IEA-Studie „Study of Achievement in Written Composition“ durchgeführten Untersuchung - viel Lehrerarbeitszeit bindet und keine lernförderliche Wirkung hat (Hartmann/Blatt 1991).

Im Zuge einer sich wandelnden Aufsatzdidaktik, die dem Schreibprozess eine hohe Beachtung schenkt, wurde die Konkretisierung von Beurteilungskriterien wieder zu einem zentralen Anliegen (Baurmann/Dehn 2004).

Da die Überarbeitung im Mittelpunkt eines prozessorientierten Schreibunterrichts steht, geht es hier um die Frage, mit welchen Kriterien der Überarbeitungsbedarf eines Textes festgestellt und die Beurteilung unterschiedlicher Textfassungen vorgenommen werden. Vorliegende linguistisch fundierte Kriterienraster (z. B. Nussbaumer 1991) liefern zwar einen Rahmen dafür, für die Unterrichtspraxis gilt es jedoch Kriterien und Verfahren zu entwickeln, die eine verlässliche Beurteilung ermöglichen, die für die Lehrkräfte leistbar, für die Schüler nachvollziehbar und für den Lernprozess förderlich sind.

Mit der „empirischen Wende“ in der Bildungsforschung und den damit zusammenhängenden Lernstandserhebungen und empirischen Studien wird die Frage nach der Messung von Kompetenzen zu einem zentralen bildungspolitischen Anliegen. In diesem Zusammenhang wird die Beurteilung von Schülertexten zum Gegenstand interdisziplinärer Forschung. Schülertexte werden theoriebasiert und mit statistischen Methoden untersucht, um empirisch validierte Kriterien zur Erfassung der Textqualität zu gewinnen (z. B. Blatt/Voss/Matthießen 2005, Neumann 2007). Weiterhin werden in der Deutschdidaktik Kompetenzmodelle theoriebasiert entwickelt (Ossner 2006).

Für die interdisziplinäre Forschung stellt sich das Problem, Beurteilungskriterien zu entwickeln, die sowohl die Textqualität abbilden als auch zu übereinstimmenden Beurteilungen unterschiedlicher Bewerter führen. Während die linguistisch basierten Kriterienkataloge in erster Linie an ihrer fachlichen Fundierung gemessen werden, ist das Maß an Kodierübereinstimmung

der Bewerter ein zentrales Gütekriterium für empirisch abgesicherte Beurteilungskriterien. Dass die Entwicklung solcher Kriterien bislang nur bedingt gelungen ist, hängt mit der möglichen Vielfalt und Komplexität von Texten zusammen. Daher kann die Textkompetenz im Gegensatz zur Lesekompetenz auch nicht mit geschlossenen Aufgabenformaten getestet werden. Die Erfassung der Textkompetenz stellt deshalb eine besondere Herausforderung für die Konzipierung von Lernstandserhebungen dar, die zur Überprüfung der Bildungsstandards in allen Bundesländern durchgeführt werden.

In dem vorliegenden Beitrag wird dargestellt, wie die Textkompetenz im Rahmen des Länderkooperationsprojekts Lernstand 6 (2008) erhoben wurde. In dem Projekt kooperieren die Länder Hessen, Hamburg, Mecklenburg-Vorpommern, Schleswig-Holstein, Thüringen und Sachsen mit dem Ziel, Vergleichsarbeiten für den Einsatz in der sechsten Jahrgangsstufe kosteneffizient und mit einem hohen wissenschaftlichen Standard zu entwickeln. Dabei wird kein Ländervergleich angestrebt, sondern das Ziel verfolgt, durch an Bildungsstandards orientierten Aufgaben hoher Qualität innerhalb der Länder gute diagnostische Informationen zu Lernständen in den Fächern Deutsch, Mathematik und Englisch bereitzustellen. Die erste Erhebung fand im Februar 2008 statt. Die Organisation der Durchführung und Nachbereitung in den Ländern unterschied sich. In einigen Ländern waren die Tests verpflichtend und es wurden Rückmeldungen zu den Ergebnissen über Internetanwendungen zur Verfügung gestellt und Auswertungsveranstaltungen angeboten, in anderen wurde lediglich das Testinstrument zur freiwilligen Durchführung zur Verfügung gestellt.

Um Schreibaufgaben zur Erfassung der Textkompetenz zu konzipieren, erarbeitete eine Entwicklergruppe zwei Schreibaufgaben und die dazugehörigen Kodieranweisungen. Als theoretisches Rahmenmodell wurde das Schreibkompetenzmodell von Bereiter (1980; s. u. Punkt 1) herangezogen. Die Aufgaben wurden in der Pilotierung der Lernstandsaufgaben 2007 eingesetzt, an der insgesamt rund 4000 Schülerinnen und Schüler teilnahmen. Die beiden Schreibaufgaben wurden jeweils von rund 400 Schülerinnen und Schülern bearbeitet und von geschulten Kodierern ausgewertet. Die theoriebasierten Beurteilungskriterien der Schreibaufgaben wurden im Hinblick auf die Aufgabenschwierigkeit, die Kodiererübereinstimmung, die Passung zum Rasch-Modell und die Fairness der Aufgabenstellung empirisch überprüft.

Neben den Schreibaufgaben wurden beim Lernstand 6 auch Aufgaben zum Leseverständnis sowie Sprachaufgaben zu Grammatik und Rechtschreibung entwickelt und erprobt.

In einer wissenschaftlichen Begleitforschung zur Entwicklung und Auswertung der Schreibaufgaben in Lernstand 6 wurden folgende fünf Forschungsfragen untersucht:

Übergeordnete Forschungsfrage:

1. Lassen sich theoretisch verankerte und unterrichtstaugliche Kriterien zur Erfassung von Textkompetenz entwickeln, die zeitökonomisch einsetzbar und als Lernbeobachtungsinstrument geeignet sind?
2. Lassen sich die theoriebasierten Kriterien auch empirisch absichern?
3. Handelt es sich bei der Textkompetenz um eine globale oder eine differentielle Kompetenz?
4. Lassen sich unterschiedliche Fähigkeitsniveaus ausweisen und aus didaktischer Sicht

sinnvoll interpretieren?

5. Bestehen Zusammenhänge zu anderen schriftsprachlichen Kompetenzen (Lese-, Grammatik-, Rechtschreibkompetenz)?
1. Forschungsstand zur theoretischen und empirischen Modellierung der Textkompetenz

Über die Textkompetenz sowie ihren Erwerb und Ausbau liegen unterschiedliche wissenschaftliche Auffassungen und Studienergebnisse vor, auf die an dieser Stelle kurz eingegangen wird.¹ Forschungsergebnisse aus unterschiedlichen Disziplinen eröffnen einen differenzierten Blick auf diesen Gegenstand.

Ergebnisse linguistischer Forschung liefern Raster zur Beschreibung von Texten, um die Schreibleistung festzustellen. Ein hochkomplexes Beobachtungssystem für die Diagnose der Schreibleistung legt Markus Nussbaumer (1991) vor. Es wurde auf der Grundlage der Analyse von Texten gewonnen, die von Schweizer Maturanden geschrieben wurden. Das Züricher Analyseraster hat sich für schulische Belange jedoch als zu komplex erwiesen. Deshalb wurden daraus vereinfachte Evaluationssysteme abgeleitet, um sie für die Untersuchung von Schülertexten in der Praxis handhabbar zu machen (Nussbaumer/Sieber 1994). Dabei reduzierten sich die Kriterien aber weitgehend auf gut nachprüfbar Aspekte wie Rechtschreibung, Grammatik und formale Aspekte.

Ein kognitionspsychologisch fundiertes und in der Deutschdidaktik vielfach rezipiertes Schreibentwicklungsmodell legt Carl Bereiter (1980) vor (vgl. Lexikon zur Deutschdidaktik 2006, S. 656 ff.). Darin weist er Teilkompetenzen einer voll ausgebildeten Schreibkompetenz aus. Schreibenlernen betrachtet er aus einer strukturalistischen Perspektive als Entwicklung und Integration distinktiver kognitiver Teilkompetenzen ohne feste Abfolge. Die Entwicklung von Schreibfähigkeiten verläuft nach Bereiter in unterscheidbaren Organisationsstadien, wobei die Teilkompetenzen in einer individuellen Reihenfolge entwickelt, integriert und ausdifferenziert werden. Dabei bauen die Lernenden mentale Modelle auf, die beim Hinzukommen neuer Teilkompetenzen oder bei ihrer Erweiterung neu organisiert werden. Schreibenlernen erscheint als ein lebenslanger Prozess, in dem die einzelnen Organisationsstadien immer höher entwickelt werden können.

Demgegenüber gehen Stufenmodelle zur Schreibentwicklung von einer sukzessiven und in der Reihenfolge nicht beliebigen Entwicklung von sprachlichen und kognitiven Fähigkeiten aus. Die Annahme einer Stufenabfolge ist im Anschluss an Piagets Entwicklungsmodell in der Pädagogik weit verbreitet.

Ein vielfach rezipiertes Stufenmodell liefern H. Feilke und G. Augst (1989) als Ergebnis einer empirischen Studie. Darin gehen sie – mit Bezug auf das Organon-Modell der Sprachfunktionen von Karl Bühler – von einer sukzessiven Ausdifferenzierung von Kommunikationsfunktionen beim Schreibenlernen aus. Danach rücken in der Schreibentwicklung die expressive, kogniti-

1 Einen Überblick über empirische Studien zur Erfassung der Textkompetenz liefern die beiden neueren empirischen Studien zur Text-Sorten-Kompetenz in der Grundschule (Augst et al. 2007, S. 28-29) und zur Modellierung und Messung von Textkompetenz auf der Grundlage der Daten aus den Studien DESI und LAU 11/Ulme (Neumann 2007; S. 53-63).

ve, textuelle und soziale Teildimension der Reihenfolge nach in den Aufmerksamkeitsfokus des Schreibers.

In einer qualitativen Longitudinalstudie mit 39 Grundschulkindern aus drei Parallelklassen untersuchen Augst u. a. (2007) die Entwicklung der Textkompetenz dieser Kinder von Klassenstufe 2 bis 4 mit Schreibaufgaben zu vier verschiedenen Textsorten: Erzählung, Bericht, Instruktion, Beschreibung, Argumentation (ebd., S. 42) an insgesamt 585 Schülertexten. Auf der Grundlage von qualitativen Analysen weisen sie vier Niveaustufen der Textqualität aus, die sich für die fünf Textsorten unterschiedlich verteilen. Während die höchste Niveaustufe bei den Textsorten Erzählung und Instruktion bei über 40 Prozent liegt, beträgt sie bei den Textsorten Argumentation und Bericht ca. 10 bzw. 5 Prozent und bei der Textsorte Beschreibung ca. 20 Prozent (ebd., S. 355). Die Autoren ziehen daraus den Schluss, dass sich die Textsortenkompetenz in einer Stufenabfolge aufbaut, wobei sich „die nächst höheren Stufen entwicklungslogisch aus der vorhergehenden ergeben und dass keine der Zwischenstufen übersprungen werden kann.“ (ebd., S. 351). Angesichts des geringen Umfangs der Stichprobe und deren homogener sozialer Herkunft (Kinder mit Deutsch als Muttersprache in einer ländlichen Mittelpunktschule) sowie der rein qualitativen Auswertung erscheinen diese Folgerungen allerdings sehr weitreichend.

Gegen Stufenmodelle generell gibt es in der deutschdidaktischen Diskussion vor allem zwei Einwände: Zum einen, dass die Rolle des Unterrichts bei der Modellierung der Stufen unberücksichtigt bleibt, zum anderen, dass sie die Vorstellung einer naturgegebenen Entwicklung in einer festen Abfolge nahe legen (z. B. Bremerich-Vos 2004, Ossner 1995; Dehn 1990).

Eine groß angelegte quantitative Untersuchung zur Modellierung und Messung von Textkompetenz erfolgte im Rahmen der DESI-Studie.² Die dort erhobenen Daten von Neuntklässlern wurden zusammen mit den Aufsatzdaten aus der Hamburger Lernausgangslagenuntersuchung in Klassenstufe 11 an allgemeinbildenden und berufsbildenden Schulen (LAU 11/ULME I)³ ausgewertet (Neumann 2007). Die Qualität der insgesamt 23 000 Schülertexte wurde sowohl mit dichotomen Kriterien (Inhalt, Textsorte Brief) als auch über eine fünfstufige Ratingskala (Aufbau, Ausdruck, Stil, Rechtschreibung, Grammatik, Syntax, Globalurteil) erfasst (Neumann 2007, S. 72ff.).

Mit unterschiedlichen statistischen Verfahren (IRT-Skalierungsverfahren, multiple Regressionsanalysen, Strukturgleichungsmodelle) wurde u. a. untersucht, ob mit dichoto-

2 Es handelt sich um die repräsentative Studie Deutsch Englisch Schülerleistungen International, die im Auftrag der Kultusminister der Länder zu zwei Messzeitpunkten (September/Oktober 2003 und Mai/Juni 2004) an einer Stichprobe von ca. 11.000 Schülerinnen und Schülern durchgeführt wurde. Zwischen den beiden Messzeitpunkten lag die DESI-Videostudie. <http://www.dipf.de/desi/> (letzter Zugriff: 30.7.2008)

3 LAU steht für Lernausgangslagenuntersuchung, die in Hamburg in den Klassenstufen 5, 7, 9, 10 bzw. 11 und 13 als Längsschnittuntersuchung von 1996 - 2004 durchgeführt wurde. <http://www.hamburger-bildungsserver.de/welcome.phtml?untent=schulentwicklung/lau/welcome.htm> (letzter Zugriff: 30.7.2008). ULME steht für Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in Berufs- und Fachschulen. ULME I wurde 2002 in den Eingangsklassen dieser Schulen in Hamburg durchgeführt. <http://fhh.hamburg.de/stadt/Aktuell/pressemeldungen/2006/mai/12/2006-05-12-bbs-ulme-II.html> (letzter Zugriff: 30.7.2008)

men Kriterien oder Ratingskalen die Textqualität besser erfasst werden kann, ob die erfasste Textkompetenz ein ein- oder mehrdimensionales Konstrukt darstellt und welche Merkmale die Schwierigkeit eines Items bestimmen (auch im Folg. Neumann 2007, S. 94 ff., 199-203).

Die Kodierübereinstimmung bei der Auswertung in den Bereichen Inhalt und Textsorte Brief mit dichotomen Kriterien ist aufgrund der eindeutigen inhaltlichen Vorgaben und der klaren formalen Struktur der Schreibaufgaben sehr hoch und liegt zwischen 87,9% und 98,1%. Die Interraterkorrelationen bei den auf einer Fünferskala eingeschätzten Kategorien *Aufbau*, *Ausdruck*, *Stil*, *Rechtschreibung*, *Grammatik*, *Syntax*, *Globalurteil* liegen dagegen darunter, wobei die Daten aus LAU 11 und ULME I schlechtere Werte erzielten als diejenigen aus der DESI-Studie. Sie liegen für alle Daten zwischen 0.38 (Satzkonstruktion), 0.39 (Textaufbau) und 0.76 (Rechtschreibung) (Neumann 2007, S. 264ff.).

Eine Dimensionalitätsprüfung spricht für eine zweidimensionale Modellierung der Textkompetenz. Dabei werden die mit dichotomen Kriterien erfassten Bereiche *Inhalt und formaler Aufbau Brief* zur Dimension *Semantik/Pragmatik* zusammengefasst. Die *Sprachsystematik* als zweite Dimension umfasst den über die Ratingskala erhobenen Bereich von Sprache, Stil und orthographischer und grammatischer Korrektheit (Neumann 2007, S. 197ff.). Damit konnten Ergebnisse der Hamburger Aufsatzstudie, einer im Rahmen der IEA durchgeführten Studie mit Schülern der Klassenstufe 11, bestätigt werden, wonach die Bereiche *Inhalt/Aufbau* und *Orthographie/Grammatik* zwei unterschiedliche Teilkompetenzen darstellen (Hartmann/Lehmann 1987). Die in der DESI-Studie nachgewiesenen hohen Kovarianzen der Teilfähigkeiten und die statistisch abgesicherte eindimensionale Modellierung sprechen jedoch auch für eine übergreifende Dimension der Textkompetenz (Neumann 2007, S. 200).

Die Skalierungsergebnisse und die Berechnung der Regressionsanalysen geben Auskunft über schwierigkeitsbestimmende Merkmale der einzelnen Items (auch im Folg. Neumann 2007, S. 123ff.). Danach besteht in den dichotom erfassten Bereichen *Inhalt und formale Textsorte Brief* eine relativ geringe Differenzierung in der Aufgabenschwierigkeit. Bei den durch Ratings erfassten Bereichen *logischer Aufbau*, *Sprache/Stil* sowie *Rechtschreibung/Grammatik* zeigen sich Unterschiede in der Schwierigkeit bezogen auf die jeweiligen Anforderungen der Schreibaufgabe (persönlicher und offizieller Brief). Hieraus lässt sich ableiten, dass es den Schülern schwer fällt, sich in der jeweiligen Textfunktion angemessen schriftsprachlich auszudrücken.

Im Hinblick auf die Kategorie *logischer Textaufbau* zeigen die Kovarianzen zwischen *Inhalt* und *Aufbau*, dass die Qualität des Textaufbaus ohne einen Bezug zum Inhalt nicht erfasst werden kann. Die Schwierigkeit für die Schüler besteht darin, den Inhalt in einer logischen Abfolge zu strukturieren.

Hier wird die Problematik sichtbar, die aus einer streng kategorialen Beurteilung von Texten resultiert und die auch in der IGLU-E Aufsatzstudie von 2001 deutlich gemacht wird (Blatt/Voss/Matthießen 2005). Dort bearbeiteten Viertklässler die Schreibaufgabe, aus dem Jahr 2051, wohin sie laut Aufgabenstellung durch eine Zeitmaschine befördert worden waren, einen Brief an eine gegenwärtige Freundin/einen gegenwärtigen Freund zu schreiben. Aus dieser Studie wurden 215 Kindertexte aus sechs Bundesländern mit dem Verfahren der

qualitativen Inhaltsanalyse und ergänzenden quantitativen Verfahren (Faktorenanalyse, LCA) untersucht. Das Ziel bestand darin, empirisch abgesicherte Kriterien zur Messung der Textqualität dieser Schreibaufgabe zu gewinnen und die Schreibleistung damit zu erfassen. Als Grundkategorien wurden die Kategorien *Inhalt*, *Aufbau*, *sprachliche Gestaltung* herangezogen. Die *schriftliche Darbietung* als vierte Kategorie betrifft Rechtschreibung und Grammatik. Deren Bewertung ist durch die Rechtschreib- und Grammatikregeln allgemein festgelegt. Eine Auswertung dieses Bereichs fand nur exemplarisch an ausgewählten Aufsätzen statt, da es in der Studie in erster Linie um Kriterienentwicklung und –überprüfung ging. Unter die Kategorie *sprachliche Gestaltung* fielen Ausdruck, Stil und Syntax.

Gängige Kriterien wie Stimmigkeit, Plausibilität, Verständlichkeit oder logischer Aufbau erwiesen sich als nicht operationalisierbar, da sie sich nicht einer Kategorie bzw. Textstelle zuordnen ließen, sondern sich auf das Zusammenspiel von Inhalt, Aufbau und Sprache beziehen. Dennoch konnten Kriterien gewonnen werden, deren durchschnittliche Interkoderreliabilität bei 0,69 lag (S. 115). Die Kriterienentwicklung erfolgte prozesshaft, indem theoretisch gewonnene Kriterien für die drei Kategorien aufgrund der Sichtung der Schülertexte ausdifferenziert und fortlaufend auf ihre Kodierübereinstimmung überprüft wurden. Mit diesen Kriterien konnten z. B. Ideenreichtum und Phantasie, die Übernahme einer anderen Perspektive sowie das Spektrum schriftsprachlicher Mittel erfasst werden (S.151f.). Durch eine explorative Faktorenanalyse konnten 14 zentrale Qualitätskriterien aus den Kategorien *Inhalt*, *Aufbau* und *Sprache* ermittelt werden (S. 126). Die Analyse der Daten mit dem Verfahren der Latent Class Analysis ergab drei unterschiedliche Typologien der Aufgabenbewältigung, die sich sowohl in der Schreib- als auch in der Leseleistung deutlich voneinander unterschieden (S. 126f.). Die Zuordnung zu den einzelnen Kategorien liefert somit auch eine Aussage über das Leistungsniveau.

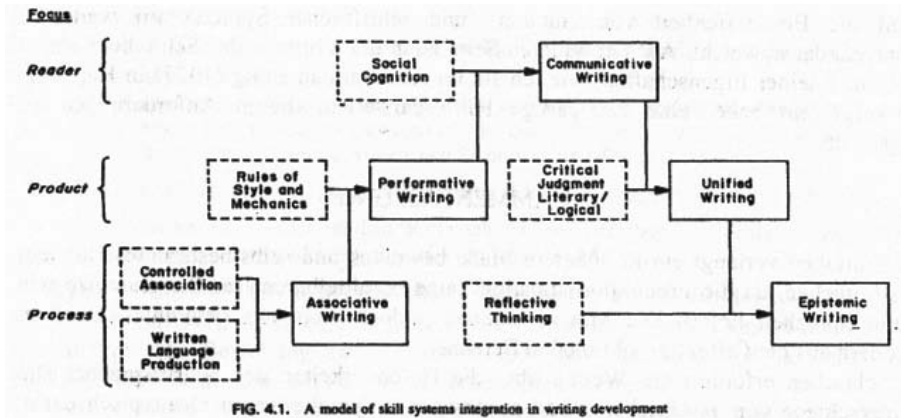
2. Theoretische Modellierung der Textkompetenz in Lernstand 6 (2008)

Für die Entwicklung von Schreibaufgaben und die Erfassung der Textkompetenz in Lernstand 6, einer Querschnitterhebung, wurde das Schreibentwicklungsmodell von Bereiter herangezogen, da es eine Modellierung erlaubt, die den Anforderungen und Zielen von Vergleichsarbeiten in vielen Punkten entgegenkommt.

Die in dem Modell ausgewiesenen Teilkompetenzen sind überschaubar und liefern zentrale Beurteilungskriterien für die Erfassung von Textqualität, die wiederum für die Rückmeldung der Ergebnisse und als Grundlage für anschließende Fördermaßnahmen aussagekräftig sind. Dazu trägt bei, dass sie sich den Kategorien *Inhalt*, *Aufbau*, *Sprache* und *Einhaltung von Schreibkonventionen* (Rechtschreibung/Grammatik/Schriftbild) zuordnen lassen, mit denen Texte strukturell beschrieben werden können und die auch in der Schulpraxis eingeführt sind. Die in dem Bereitermodell ausgewiesenen Teilkompetenzen können – mit Ausnahme der epistemischen Teilkompetenz – für die Beurteilung der Schülertexte operationalisiert werden. Ein weiterer Vorteil liegt darin, dass Bereiter mit den Kategorien „knowledge telling“ und „knowledge transforming“ die Möglichkeit eröffnet, bei den einzelnen Teilkompetenzen unterschiedliche Niveaus zu erfassen. Dies kommt der Anforderung entgegen, mit einer Aufgabenstellung ein breites Kompetenzspektrum abzudecken.

Carl Bereiter (1980:82ff.) weist in seinem Modell sechs Teilkompetenzen aus, die er zu fünf Schreibentwicklungsstadien („stages“) zusammenfasst (Abbildung 1).

Abbildung 1: Schreibentwicklungsmodell von Bereiter



1. Assoziatives Schreiben (Ideen entwickeln, aufschreiben)
2. Performatives Schreiben (Schreibkonventionen einhalten)
3. Leserbezogenes Schreiben (die Perspektive des Lesers einnehmen)
4. Textgestaltendes Schreiben (Sprach/Stilmittel einsetzen und Textmuster nutzen, die dem Thema, der Intention und dem Adressaten angemessen sind)
5. Epistemisches Schreiben (Schreiben zur Gedankenentwicklung und Reflexion nutzen).

Bereiter unterteilt sie danach, ob der Fokus auf dem Prozess (1, 5), dem Produkt (2, 4) oder dem/ der Leser/in (3) liegt.

Die „stages“ beschreiben Teilkompetenzen einer voll ausgebildeten Schreibkompetenz. Die Kategorien Inhalt, Aufbau, Sprache und Einhaltung von Sprachkonventionen können damit wie folgt konkretisiert und ausdifferenziert werden:

Der Inhalt eines Textes kann darüber Auskunft geben, in welcher Weise der Schreiber Ideen zur Schreibaufgabe entwickeln kann. Es wird beurteilt, ob Inhalte nur in einem assoziativen Bezug zueinander stehen (assoziatives Schreiben) oder ob sie leserbezogen ausgewählt sind, indem sie z. B. einen logischen Zusammenhang, anschauliche Details oder überraschende Einfälle aufweisen, die das Leserinteresse wecken können (leserbezogenes Schreiben). Der Leserbezug wiederum kann auf einem unterschiedlich hohen Niveau liegen.

Der Aufbau eines Textes kann Auskunft über die Teilkompetenzen *assoziatives Schreiben* und *textgestaltendes Schreiben* geben. Es kann überprüft werden, ob die Inhalte assoziativ aneinandergereiht (assoziatives Schreiben) oder ein einfaches bzw. komplexeres Aufbauschema aufweisen (einfache und erweiterte *Textgestaltung*).

Mit den im Text eingesetzten sprachlichen Mitteln kann die Teilkompetenz des *textgestaltenden Schreibens* untersucht werden. Es geht dabei um textkohäsive Mittel, um sprachliche Angemessenheit in Bezug auf Inhalt und Adressaten, um Ausdrucksgenauigkeit und –vielfalt. Die dazu erforderlichen Beurteilungskriterien lassen auch Aussagen über die Teilkompetenz *Leserbezug* zu.

Die Teilkompetenz „performatives Schreiben“ wird mit der Kategorie *Einhaltung von Sprachkonventionen* erfasst. Hierbei wird untersucht, inwieweit die Einhaltung von Sprachkonventionen (Rechtschreibung, Grammatik) die Lesbarkeit des Textes gewährleistet. Zusätzlich bestimmt das Schriftbild als eine motorische Teilkompetenz die Lesbarkeit.

Die Teilkompetenz epistemisches Schreiben kann durch Tests nur sehr bedingt erfasst werden, wie z. B. durch die Untersuchungen von Textüberarbeitungen. Ihre Ermittlung setzt vielmehr eine Beobachtung des Schreibprozesses voraus.

3. Entwicklung der Schreibaufgaben und der Kodieranweisungen

3.1. Beschreibung der Aufgaben

Unterschiedliche Schreibaufgaben stellen Schreiber bzw. Schreiberinnen auch vor unterschiedliche Anforderungen, so dass mit einem Aufgabentyp nicht die *Schreibkompetenz an sich* getestet werden kann, sondern nur die diesem Aufgabentyp entsprechenden Anforderungen. Jede Schreibaufgabe kann aber so gestellt werden, dass sie eine Bearbeitung auf unterschiedlichem Niveau ermöglicht („knowledge telling“, „knowledge transforming“ nach Bereiter).

Für die Entwicklung der Schreibaufgaben für Lernstand 6 (2008) wurde vor allem aus zwei Gründen die Textsorte „persönlicher Brief“ gewählt.

Zum einen hat die schriftliche Kommunikation einen lebensweltlichen Bezug für die Schülerinnen und Schüler. Sie wird von den Jugendlichen heute überwiegend auf elektronischem Wege geführt, wobei E-Mails und vor allem SMS allerdings von der herkömmlichen Briefform in vielfacher Weise abweichen. Aber auch das Briefeschreiben ist weiterhin präsent. Es werden Briefe oder Karten an Familienmitglieder oder Freunde geschickt und auch das Pflegen von Brieffreundschaften ist, nach den Kontaktbörsen in Jugendzeitschriften zu schließen, noch immer aktuell. Weiterhin ist der Brief eine formalisierte Textsorte, die sich daher gut für eine kriteriengestützte Erfassung eignet.

Von der Aufgabenentwicklergruppe⁴ im Rahmen von Lernstand 6 (2008) wurden mehrere Schreibaufgaben entwickelt und im eigenen Unterricht erprobt. Auf der Grundlage der dabei gewonnenen Erfahrungen wurden zwei Aufgaben für die Pilotierung ausgewählt. Es handelt sich um einen Brief an einen kranken Mitschüler und um eine Antwort auf eine Anzeige „Brieffreund/Brieffreundin gesucht“. Von diesen beiden Aufgaben wurde aufgrund der Pilotierungsergebnisse eine Aufgabe für den Haupttest im Frühjahr 2008 ausgewählt, und

4 Die Aufgabenentwicklergruppe umfasst jeweils zwei bis vier Deutschlehrkräfte sowie Fachberaterinnen der beteiligten Länder (Hamburg, Hessen, Mecklenburg-Vorpommern, Sachsen, Schleswig-Holstein und Thüringen), Testkonstruktoren und Fachdidaktikerinnen.

zwar die Aufgabe *Brieffreundschaft*.

Bei der Gestaltung der Schreibaufgaben wurde nicht nur die Frage der Kompetenzmessung bedacht, sondern es wurde auch darauf geachtet, dass die Aufgaben zum Schreiben herausfordern und geeignet sind, die Schülerinnen und Schüler im Schreibprozess zu unterstützen. Dazu sollte die Aufgabenformulierung beitragen:

Aufgabe: Einen Brief an einen kranken Mitschüler schreiben

Stell dir vor, dass sich ein Mitschüler kurz vor Ende der Sommerferien in Spanien ein Bein gebrochen hat. Er soll beim Spielen mit einem Hund schwer gestürzt sein. Da der Bruch sehr kompliziert ist, muss er sechs Wochen in Spanien im Krankenhaus bleiben. Er ist ganz alleine dort, da seine Eltern zurück mussten.

Überlege

- Was möchtest du ihn fragen?
- Wie könntest du ihn aufmuntern?
- Welche Neuigkeiten möchtest du ihm berichten?

Schreibe den Brief an den Mitschüler so, dass er ihn gut lesen und verstehen kann.

Aufgabe: Einen Brief auf eine Anzeige für eine Brieffreundschaft schreiben

Du liest in der Jugendzeitschrift *Geolino* folgende Anzeige für eine Brieffreundschaft:

Hallo! Ich heiße Rabea und bin elf Jahre alt. Ich suche Brieffreunde, egal ob Mädchen oder Junge. Meine Hobbys sind Wellenreiten, Schwimmen, Einrad fahren, Fußball spielen und Lesen. Ich wohne seit fünf Jahren in Honolulu auf Hawaii und ziehe in zwei Jahren zurück nach Deutschland

Du möchtest Rabea unbedingt als Brieffreundin. Du vermutest aber, dass sie sehr viele Briefe erhält. Schreibe ihr einen Brief und lass dir etwas besonders Gutes einfallen, damit sie dir zurückschreibt.

Schreibe so, dass Rabea deinen Brief gut lesen und verstehen kann.

In beiden Schreibaufgaben geht es um den Schülerinnen und Schülern vertraute Themen. Ein besonderer Schreibanreiz sollte beim *Brief an den kranken Mitschüler* davon ausgehen, dass er an einen im Urlaub in Spanien verunglückten Klassenkameraden gerichtet ist, der dort im Krankenhaus liegt.

Bei der Aufgabe *Brieffreundschaft* sollte ein Anreiz dadurch geschaffen werden, dass die Annonce von einem mit ihrer Familie in Honolulu lebenden deutschen Mädchen mit ungewöhnlichen Hobbys aufgegeben wird. Dieses etwas exotische Ambiente kann Schreibenden auch zu ungewöhnlicheren Lösungen anregen. Von der Tatsache, dass das Mädchen in absehbarer Zeit wieder nach Deutschland zurückkommt, kann ebenfalls ein Schreibanreiz ausgehen.

Beide Schreibaufgaben bestehen neben der Arbeitsanweisung aus einem kurzen Text. Im *Brief an den kranken Mitschüler* werden Angaben zum Unfall und zum Krankenhausaufenthalt geliefert, die Aufgabe *Brieffreundschaft* enthält den Annoncentext. Diese inhaltlichen

Vorgaben sollen die Schülerinnen und Schüler im Schreibprozess insofern entlasten, als sie das Generieren und Strukturieren von Schreibideen anstoßen können. Damit wird einer möglichen „Angst vor dem leeren Blatt“ entgegengewirkt, da allein durch das Aufgreifen der Vorgaben ein Brief geschrieben werden kann. Die Schreiber bzw. Schreiberinnen können die Entlastung auf inhaltlicher Ebene aber auch für die sprachliche Gestaltung oder für die Herstellung eines Leserbezuges nutzen.

3.2. Entwicklung der Kodieranweisungen

Für die Entwicklung der Beurteilungskriterien (Items) wurde das Verfahren der qualitativen Inhaltsanalyse eingesetzt und durch statistische Verfahren der Itemprüfung ergänzt. Die qualitative Vorgehensweise und die inhaltliche Ausgestaltung werden in diesem Abschnitt, die statistischen Verfahren in Punkt 4 dargestellt.

Die Kodieranweisungen wurden in mehreren Schritten entwickelt.

Zunächst wurde für die Textsorte Brief ein Grundkodierschema für die Beurteilung, basierend auf dem Modell Bereiters, erstellt. Es beinhaltet folgende Aspekte:

- I. Inhalt:
 - 1) Aufgreifen der Vorgaben und eigene Angaben (assoziatives bzw. elementares leserbezogenes Schreiben)
 - 2) Ausführungen zu den Vorgaben und eigenen Angaben (entfaltetes leserbezogenes Schreiben)
- II. Aufbau:
Briefform (elementares textgestaltendes Schreiben⁵)
- III. Sprache/Stil:
Verständlichkeit, Anschaulichkeit, Textkohäsion (entfaltetes textgestaltendes Schreiben)
- IV. Schreibkonventionen:
Normgerechtes und lesbares Schreiben (performatives Schreiben)

Dieses Grundmuster wurde anschließend für die jeweilige Schreibaufgabe konkretisiert. Zu diesem Zweck wurde aufgrund der in der Schreibaufgabe gestellten Anforderungen eine größere Anzahl dichotomer⁶ Kriterien (ja/nein) für die Analyse der Schülertexte entwickelt. Da erwartungsgemäß nicht alle Kriterien der empirischen Überprüfung standhalten würden, wurden für jede Schreibaufgabe über 30 Items entwickelt, obwohl die endgültige Kodieranweisung höchstens 20 Items enthalten sollte. Die Beschränkung der Itemanzahl ist zeitökonomisch, und damit zugleich pädagogisch begründet, da das Verhältnis von Arbeitsaufwand und Ertrag darüber entscheidet, ob Lehrkräfte neue Verfahren in ihren

5 Dass der Brief einen festgelegten formalen Aufbau hat, schließt eine Zuordnung der Kategorie *Aufbau* zu *assoziatives Schreiben* aus.

6 Dichotome Kriterien bieten gegenüber abgestuften Bewertungen den Vorteil höherer Kodierübereinstimmungen.

Unterricht integrieren. Dieser pragmatische Zugang ergibt sich auch vor dem Hintergrund der Lernstandserhebung, da der umfangreiche Datenkorpus der Schülertexte von Lehrkräften auszuwerten ist. Für jeden Schülertext muss dabei eine kriteriengestützte qualitative Analyse auf der Grundlage der Kodieranweisung vorgenommen werden.

Die theoretisch entwickelte Kodieranweisung wurde in einem ersten Schritt an ausgewählten Schülertexten erprobt, die im Kontext der Aufgabenentwicklung von Schülerinnen und Schülern der Klassenstufe sechs zu den beiden Schreibaufgaben *Krankenbrief* und *Brieffreundschaft* geschrieben wurden. Eine Auswahl dieser Texte wurde von vier bis sechs erfahrenen Kodiererinnen (Deutschlehrkräften) mit Hilfe dieser Kriterien beurteilt. Die Kodierergebnisse wurden auf Übereinstimmungen bzw. Abweichungen hin untersucht. Neben der kriterienbezogenen nahmen die Kodiererinnen auch eine holistische Einschätzung der Texte vor. Auf der Grundlage der daraus gewonnenen Erkenntnisse wurden die Kriterien überarbeitet.

Die überarbeiteten Kodierschemata für die beiden Schreibaufgaben *Brief an einen kranken Mitschüler* und *Brieffreundschaft* enthielten 31 bzw. 34 Kriterien.⁷ Damit wurden die rund 400 Schülertexte pro Schreibaufgabe aus der Pilotierung kodiert. Jeweils einhundert Texte wurden zusätzlich von einem zweiten und dritten Bewerter beurteilt. Die Ergebnisse der statistischen Auswertung dieser Daten (s. Punkt 5) wurden genutzt, um die Kodieranweisungen für den Einsatz in der Hauptuntersuchung zu überarbeiten. Die überarbeiteten Kodieranweisungen sind für die Deutschlehrerinnen und -lehrer bestimmt, um die im Rahmen von Lernstand 6 (2008) geschriebenen Schülertexte damit zu beurteilen.

3.3. Kodierschema Brieffreundschaft

Da die Schreibaufgabe *Brieffreundschaft* nach den Pilotierungsergebnissen, die in Punkt 5 dargestellt werden, gute Testeigenschaften aufweist, wurde sie für die Hauptuntersuchung ausgewählt. Das dazugehörige Kodierschema wird im Folgenden inhaltlich beschrieben, wobei auch die Entwicklung einbezogen wird.

Das in der Pilotierung eingesetzte Kodierschema für die Schreibaufgabe *Brieffreundschaft* besteht aus 34 Kriterien (siehe Anhang 1).⁸ Es enthält neben den einzelnen Kriterien (Items), die von den Kodiererinnen mit Ja oder Nein zu erfassen sind, auch Beurteilungshinweise. Dafür wurden Beispiele aus den Schülertexten, die während der Aufgabenentwicklung geschrieben wurden, ausgewählt.

Nach den Ergebnissen der empirischen Überprüfung der Pilotierungsdaten wurden 16 aus den ursprünglich 34 Items ausgewählt, die den angelegten Gütekriterien entsprachen. Sie bilden das Kodierschema für die Hauptuntersuchung (s. Anhang 3). Wie in Punkt 5 aufgezeigt wird, wurden bei der Auswahl der Items die einzelnen Gütekriterien gegeneinander abgewogen (s. S. 69f). An oberster Stelle sollte gewährleistet werden, dass mit dem Kodierschema das theoretisch verankerte Kompetenzkonstrukt auch gemessen werden kann. Dass dies der Fall ist, wird im Folgenden aufgezeigt (Abbildung 2).

7 Für die Schreibaufgabe *Brieffreundschaft* befindet sich dieses für die Pilotierung überarbeitete Kodierschema im Anhang (1).

8 Der Anhang ist unter der Internetadresse <http://www.didaktik-deutsch.de/> abzurufen.

Kriterien für die Kodieranleitung	
Elementarer Leserbezug (leserbezogenes Schreiben)	I. Inhalt 1) Bezugnahme zu Angaben der Brieffartnerin a) Wohnort b) Hobbys
Entfalteter Leserbezug (leserbezogenes Schreiben)	2) Ausführungen zu den eigenen Angaben a) Erläuterung/Kommentierung/Beispiele zur eigenen Person, Wohnort, Hobbys, Eigenschaften b) Etwas Außergewöhnliches aus dem eigenen Leben wird berichtet, um das Interesse an der Brieffreundschaft zu wecken 3) Ausführungen zur Bezugnahme auf die Brieffartnerin a) Besonderes Interesse an der Person: Fragen zu den Angaben
Elementare Textgestaltung (textgestaltendes Schreiben in formaler Hinsicht)	II. Aufbau (Textsorte Brief) a) Anrede/Begrüßungsformel b) Hauptteil (Angaben und Ausführungen zum Thema) c) Abschied (Verabschiedung bzw. Grußformel) d) Unterschrift
Entfaltete Textgestaltung (textgestaltendes Schreiben)	III. Sprache/Stil 1) Zur Erläuterung/Kommentierung /Begründung a) Verben, die das eigene Denken und Fühlen explizit ausdrücken b) Wertung durch Adjektive c) Nachgestellte Erklärung und Erläuterung
	2) Sprachliche Komplexität a) Satzgefüge (Haupt-Nebensatz-Verbindungen) b) Konjunktionen (komplexe Beziehung) c) Anschluss mit Relativ- oder Fragepronomen
Schreibkonventionen (performatives Schreiben)	IV. Rechtschreibung, Grammatik, Schrift a) Beurteilung von Rechtschreibung und Grammatik im Hinblick auf die Lesbarkeit des Textes b) Beurteilung des Schriftbildes im Hinblick auf die Lesbarkeit des Textes

Abbildung 2: Kriterien zur Beurteilung der Schreibaufgabe *Brieffreundschaft* als Grundlage für die Hauptuntersuchung

Als 17. Item kam die Beurteilung des Schriftbildes im Hinblick auf die Lesbarkeit des Textes im Bereich *Schreibkonventionen* hinzu.

Mit Hilfe der Kodieranweisung werden folgende Textqualitäten auf der Grundlage des Schreibentwicklungsmodells von Bereiter erfasst:

Um die Schreibaufgabe zu lösen, muss ein inhaltlicher Bezug zu den Angaben des Mädchens hergestellt werden, das die Annonce aufgegeben hat. Weiterhin müssen elementare Angaben zur eigenen Person gemacht werden. Die ursprünglich enthaltenen Items „Nennung von

Name, Alter und Wohnort“ waren laut statistischer Überprüfung viel zu leicht. Demgegenüber erwiesen sich Erläuterungen dazu als Qualitätskriterien. Sie ermöglichen der Briefpartnerin, sich eine Vorstellung von der ihr unbekanntem Schreiberin und ihren Lebensumständen zu machen. Ein Leserbezug wird weiterhin an Ausführungen zu den Angaben der Briefpartnerin deutlich. Dadurch, dass z. B. Nachfragen dazu gestellt oder Kommentare dazu geliefert werden, zeigt die Schreiberin/der Schreiber ihr/ sein Interesse an der Adressatin. Dies erhöht die Chance, als Brieffreundin/Brieffreund ausgewählt zu werden. Damit ist eine Vorgabe der Aufgabenstellung erfüllt.

Der Aufbau wird ausschließlich aufgrund formaler Kriterien beurteilt, die die Textsorte Brief charakterisieren.

Um Kriterien für die sprachliche Gestaltung der Texte zu gewinnen, wurden zwei Aspekte berücksichtigt: Wie wurden erläuternde, kommentierende und begründende Inhalte sprachlich umgesetzt? Wie komplex ist die Sprachstruktur des Textes?

Beim ersten Aspekt geht es um die Verständlichkeit und Anschaulichkeit des Textes. Dazu tragen Verben und Adjektive bei, die der Adressatin signalisieren, was die Brieffschreiberin/der Brieffschreiber denkt und fühlt und wie sie/er Dinge bzw. Personen beurteilt. Nachgestellte Erklärungen zeigen, dass sich der Schreibende in die Adressatin hineinversetzt und auf diese Weise versucht, ihr einen genannten Sachverhalt zu verdeutlichen.

Der Aspekt der *sprachlichen Komplexität* bezieht sich auf die Textkohäsion. Werden Satzgefüge verwandt, Sätze und Absätze durch unterschiedliche Konjunktionen verknüpft (kausal, final, adversativ) und nicht nur durch „und“ aneinandergereiht sowie Teilsätze durch Relativ- oder Fragepronomen angeschlossen, so entsteht eine zusammenhängende Textoberflächenstruktur.

Die Einhaltung von Sprachkonventionen hat in einem Text den Sinn, dem Leser bzw. der Leserin das Lesen und Verstehen des Textes zu erleichtern bzw. überhaupt erst zu ermöglichen. Zum einen werden daher in den Schülertexten Rechtschreibung und Grammatik betrachtet und zum anderen das Schriftbild. Beide Bereiche werden unter dem Aspekt gewertet, inwieweit sie die Lesbarkeit des Briefes gewährleisten.

Rechtschreibung und Grammatik wurden in der Pilotierung zunächst holistisch eingeschätzt. Dies führte jedoch zu keiner zufriedenstellenden Kodierübereinstimmung. Daher wurde anschließend eine alternative Erfassung durch einen Prozent-Richtig-Koeffizienten erprobt. Er führte zu besseren Übereinstimmungen und genaueren Ergebnissen, aber er erwies sich für den Haupttest als zu zeitaufwendig. Daher wurde für den Haupttest die holistische Einschätzung in einer präzisierten Variante gewählt: Rechtschreibung und Grammatik wurden gesondert eingeschätzt, und zwar im Hinblick darauf, ob sie in den Texten überwiegend richtig gebraucht werden.

4. Datengrundlage und eingesetzte Verfahren

In den Punkten 1 bis 3 wurde gezeigt, welche Kriterien zur Textbeurteilung auf der Grundlage des Schreibprozessmodells und des im Aufsatzunterricht tradierten Schemas *Inhalt, Aufbau, Sprache* und *Schreibkonventionen* entwickelt wurden und welche Kriterien empirisch abgesichert werden konnten.

Im Folgenden wird dargelegt, auf welcher Datengrundlage und mit welchen statistischen Verfahren diese empirische Absicherung und die weiteren in der Einleitung entwickelten Forschungsfragen 2 - 5 untersucht wurden (Tabelle 1).

Die übergeordnete Forschungsfrage 1, inwieweit sich theoretisch verankerte und unterrichtstaugliche Kriterien zur Erfassung von Textkompetenz entwickeln lassen, die zeitökonomisch einsetzbar und als Lernbeobachtungsinstrument geeignet sind, wird auf der Grundlage der Auswertungsergebnisse zu den Forschungsfragen 2-5 vorgenommen.

Datengrundlage und Forschungsfragen (F 2-F5)	Statistische Verfahren
Datengrundlage: 431 Schülertexte	
Forschungsfragen: F 2. Lassen sich die theoriebasierten Kriterien auch empirisch absichern?	Die Aufgaben wurden mit Hilfe mehrerer Kriterien mit unterschiedlichen Verfahren untersucht: a) Angemessene Schwierigkeit: Berechnung der Lösungshäufigkeiten b) Auswertungsobjektivität: Berechnung der Übereinstimmung der Kodierungen durch 3 Kodierer mit dem Maß Cohens Kappa c) Fairness: Berechnung der Lösungsunterschiede nach Gruppen durch das <i>Differential Itemfunctioning</i> mit ConQuest für Geschlecht, Migrationsstatus und Land (Nord-Süd; Ost-West) d) Passung zum Rasch-Modell: Bestimmung der Aufgabenschwierigkeit auf der Basis der Raschskalierung. Kriterien für die Itemauswahl sind ein WMNSQ zwischen 0.8 und 1.2 (vgl. Adams, 2002, S. 105)
F 3. Handelt es sich bei der Textkompetenz um eine globale oder eine differentielle Kompetenz?	<ul style="list-style-type: none"> • Modellvergleiche (ein-, mehrdimensional) mit psychometrischen Verfahren (Raschskalierung mit ConQuest)
F 4. Lassen sich unterschiedliche Fähigkeitsniveaus ausweisen und aus didaktischer Sicht sinnvoll interpretieren?	<ul style="list-style-type: none"> • Raschskalierung zur Verortung der Itemschwierigkeiten auf der Skala und Interpretation der Differenzen nach schreibtheoretischen Kriterien
F 5. Bestehen Zusammenhänge zu anderen schriftsprachlichen Kompetenzen (Lese-, Grammatik-, Rechtschreibkompetenz)?	<ul style="list-style-type: none"> • getrennte Rasch-Skalierung der Aufgaben <i>Sprachgebrauch, Lesen, Texten</i> (getrennt nach Bereichen I-III und IV) mit ConQuest. • Korrelationsanalysen

Tabelle 1: Datengrundlage, Forschungsfragen und Forschungsdesign

Um die Gütekriterien aller im Rahmen von Lernstand 6 (2008) entwickelten Aufgaben zu ermitteln, wurde, wie oben (S. 55) ausgeführt, in sechs Ländern in verschiedenen Schulformen

im Jahre 2007 eine Pilotierung durchgeführt, an der insgesamt 4045 Schülerinnen und Schüler der Jahrgangsstufe 6 teilnahmen.

Alle zu überprüfenden Aufgaben zu *Sprache und Sprachgebrauch*, *Textschreiben* und *Leseverständnis* wurden in insgesamt 19 unterschiedlichen Testheften im Multimatrixdesign vorgegeben. Das Multimatrixdesign hat den Vorteil, dass in den Testheften insgesamt eine hohe Anzahl von Aufgaben enthalten ist, da innerhalb der 19 Testhefte die gleichen Aufgaben nur zweimal an unterschiedlichen Positionen vorkommen. In zwei Heften waren auch die Schreibaufgaben enthalten. Zur Kontrolle von Positionseffekten (Rost, 2004) waren die Schreibaufgaben einmal im ersten Teil des Testheftes und in einer Variante im zweiten Teil platziert.

Die für jede der beiden Schreibaufgaben vorliegenden ca. 400 Texte wurden mit Hilfe der oben beschriebenen theoretisch entwickelten 31 bzw. 34 dichotomisierten Items ausgewertet.

Die holistische Einschätzung im Hinblick auf die Einhaltung der Schreibkonventionen wurde in der Pilotierung dadurch ergänzt, dass Rechtschreibung und Grammatik in einer Stichprobe mit einem Richtig-Koeffizienten ausgewertet wurden. Um diesen Koeffizienten zu ermitteln, wurden alle Wörter des Textes gezählt. Ziffern wurden nicht mitgezählt. Weiterhin wurden alle notwendigen Satzzeichen (Kommata, Punkte, etc.) gezählt, auch wenn sie fehlten, da Satzzeichen als Verstehenshilfe für den Leser fungieren. Anschließend wurden alle rechtschriftlich und grammatisch richtigen Wörter gezählt. Aus diesen Daten wurde der Prozentkorrekt-Wert der Wörter incl. Satzzeichen errechnet (richtige Wörter incl. Satzzeichen/alle Wörter incl. Satzzeichen x 100). Als verständlich wurde ein Text eingestuft, wenn 80 Prozent der Wörter richtig sind.

Um im Anschluss an die Pilotierungsergebnisse die Aufgaben für die Lernstandserhebung und die dazugehörigen Beurteilungskriterien auszuwählen, wurden folgende Gütekriterien zugrunde gelegt:

- angemessene Schwierigkeit,
- Auswertungsobjektivität,
- Passung zum Rasch-Modell,
- Fairness,
- inhaltliche Gesichtspunkte.

Um zu leichte und zu schwierige Items zu identifizieren, wurde die Lösungshäufigkeit berechnet und Items mit einer Lösungshäufigkeit über 80 % bzw. unter 20 % wurden markiert.

Hinsichtlich guter Auswertungsobjektivität sollten Bewertungen gleicher Lösungen durch unterschiedliche Personen möglichst übereinstimmen. Um dies zu prüfen, wurden die Testheftergebnisse in mindestens 100 Fällen dreifach kodiert. Aus diesem Vorgehen resultieren drei Übereinstimmungswerte zur Bestimmung der Interraterreliabilität:

1. Übereinstimmungswert= Vergleich Ergebnis Bewerter 1 mit Ergebnis Bewerter 2
2. Übereinstimmungswert= Vergleich Ergebnis Bewerter 1 mit Ergebnis Bewerter 3
3. Übereinstimmungswert= Vergleich Ergebnis Bewerter 2 mit Ergebnis Bewerter 3

Als Maß der Interraterreliabilität wurde Cohens Kappa mit SPSS berechnet (Wirtz & Caspar, 2002, S. 17-19). Der Kappa-Wert (κ) kann Werte zwischen -1 (völlig unterschiedliche Einschätzung von zwei Bewertern) und +1 (perfekte Übereinstimmung) annehmen. In der Literatur wird ein Wert von $\kappa > 0,75$ als sehr gute Übereinstimmung und Werte $\kappa > 0,5$ als befriedigend angesehen. Bei der Bewertung ist allerdings zu berücksichtigen, dass der Kappa-Wert auch von anderen Merkmalen des Bewerterverhaltens, wie z. B. Strenge, Sorgfalt und auch von der Anzahl der Kategorien und von der Grundhäufigkeit der Kategorienwahl abhängt und daher keine allgemeingültigen Richtwerte gegeben werden können.

Von den in der Pilotierung überprüften *Sprachegebrauchsaufgaben*, die als geschlossene Aufgabenformate (z. B. Multiple-Choice) vorlagen, wurden nur Items für den Haupttest ausgewählt, die eine Kodierübereinstimmung $\kappa \geq 0,7$ aufwiesen. Im Hinblick auf die Bewertung freier Antwortformate (freie Formulierung der Antwort) – wie sie auch für die *Schreibaufgaben* vorliegen – wurden zum Teil deutlich geringere Übereinstimmungen erreicht. Diese Problematik wurde bereits für andere Tests, die Schreibaufgaben umfassen, belegt (vgl. Neumann 2007; Arras & Grotjahn, 2002). Für die offenen Aufgaben musste daher aus Praktikabilitätsgründen großzügiger verfahren werden. Als Gütekriterium wurde der mittlere Kappa-Wert herangezogen, da – bei mehr als zwei Kodierern – der Median der Gütemaße aller Rater die Gesamtstichprobe am besten charakterisiert (Wirtz & Caspar, 2002, S. 120). Um diesen Wert zu ermitteln, bildet man eine Rangreihe aus den Übereinstimmungswerten der drei Kodierer und wählt den mittleren aus.

Die Aufgaben aus den Bereichen *Sprachegebrauch*, *Textschreiben* und *Lesen* wurden einer getrennten Skalierung im Rasch-Modell mit ConQuest (Wu, Adams & Wilson, 1998) unterzogen.

Bei der Schreibaufgabe wurde das holistisch eingeschätzte Item *Rechtschreibung und Grammatik* im Bereich IV, *Schreibkonventionen* nicht mit skaliert, da keine ausreichende Kodierübereinstimmung erzielt werden konnte.

Aufgrund zu geringer Lösungshäufigkeit (2,4 %) wurde das Item *Ort/Datum* aus den weiteren Berechnungen herausgenommen.

Mit Hilfe der Raschskalierung wurden die einzelnen Items auf ihre Schwierigkeit hin untersucht. Zu diesem Zweck werden die Antworten aller Probanden analysiert und in mehreren Analyseläufen (Iterationen) bestimmten Aufgabenschwierigkeiten zugewiesen. Aufgrund dieser mathematischen Berechnungen erhalten die Items einen Passungswert, den sog. WMNSQ-Wert: Der WMNSQ hat einen Erwartungswert von 1. Abweichungen nach oben und nach unten sind im Bereich zwischen 0.8 und 1.2. tolerierbar. Stärkere Differenzen kennzeichnen eine Verletzung der modellimpliziten Annahme (vgl. Adams, 2002, S. 105).

Zur Berechnung der Itemfairness wurden die Lösungsunterschiede nach Gruppen durch das *Differential Itemfunctioning* mit Conquest für Geschlecht, Migrationsstatus und Land (Nord-Süd; Ost-West) berechnet. Die Itemfairness gibt Auskunft darüber, ob eine Aufgabe allen Testpersonen dieselben Chancen bietet oder ob eine Benachteiligung für bestimmte Gruppen besteht.

Neben diesen statistisch-formalen Auswertungskriterien wurden Kriterien inhaltlicher Art berücksichtigt. Alle Kriterien wurden im Hinblick auf ihre inhaltliche Bedeutsamkeit

für die Konstruktvalidität durch Experten überprüft. Items, die aufgrund ihrer formal-statistischen Werte als nicht geeignet erscheinen, wurden dennoch beibehalten, wenn es die Konstruktvalidität – als übergeordnetes Kriterium - erforderte.

5. Ergebnisse und Interpretation

5.1. Entwicklung der Kodieranleitung und Auswahl der Schreibaufgabe

Aus den Kodieranleitungen beider Schreibaufgaben wurde unter Berücksichtigung der Gütekriterien etwa die Hälfte der Items herausgenommen. Bei der Auswahl der Items wurden die fünf oben genannten Gütekriterien betrachtet und gegeneinander abgewogen. So wurde z. B. das Item Ort/Datum wegen der zu geringen Lösungshäufigkeit eliminiert. Dass nur in 2,4 % der Schülertexte (Brieffreundschaft) diese für die Briefform wichtige Formalie vorhanden war, deutet darauf hin, dass sie nicht im Unterricht behandelt wurde. Daher sollte es auch nicht zur Kompetenzüberprüfung herangezogen werden. Andere Items, wie z. B. der Einleitungssatz, wurden wegen zu hoher Lösungshäufigkeit (87 %) entfernt.

Im Hinblick auf das Gütekriterium *Auswertungsobjektivität* wurden Items mit einem mittleren Kappa-Wert unter .3 eliminiert.

Wegen fehlender Passung zum Rasch-Modell fielen bei der Schreibaufgabe *Brieffreundschaft* fünf Items heraus (WMNSQ über 1.2; vgl. Anhang 2). Die Passung als Ergebnis dieser mathematischen Operation liefert jedoch keine inhaltlichen Aussagen darüber, warum ein Item nicht zu den übrigen passt. Inhaltliche Gründe können sein, dass die Aufgabenstellung missverständlich ist, dass es für bestimmte Items sehr abweichende Antwortverhalten gibt oder dass nicht dieselbe Kompetenz wie durch das Gesamtkonstrukt gemessen wird.

Betrachtet man die aufgrund der fehlenden Passung eliminierten Items unter inhaltlichen Gesichtspunkten des Schreibentwicklungsmodells von Bereiter, so fallen nur Nuancierungen weg, während die für die Aufgabenstellung zentralen Aspekte erhalten bleiben.

Die Wahl der Schreibaufgabe für den Haupttest wurde auf Grundlage der vergleichenden Auswertung vorgenommen.

Die durchschnittliche Lösungshäufigkeit für die beiden Schreibaufgaben lag in einem angemessenen Bereich (55 % vs. 76 %). Die Reliabilität der Gesamtaufgabe und die Trennschärfen der Items lagen ebenfalls für beide Schreibaufgaben in akzeptabler Höhe (Cronbachs Alpha > 0,85, Trennschärfen zwischen 0,49 bis 0,78).

Hinsichtlich der Kodierübereinstimmung erzielten die Texte zur *Brieffreundschaft* eine höhere Bewerterübereinstimmung als die *Briefe an den kranken Mitschüler*.

Die Untersuchung der beiden Schreibaufgaben im Hinblick auf die Fairness, d. h., ob die Aufgaben von Jungen und Mädchen sowie von Schülern mit und ohne Migrationshintergrund gleich gut zu lösen sind, ergab ebenfalls bessere Werte für die Schreibaufgabe *Brieffreundschaft*. Wie die folgende Übersicht (Tabelle 2) zeigt, besteht im *Brief an den kranken Mitschüler* ein hoher Bias sowohl für das Geschlecht als auch für den Migrationsstatus. Dabei wurde als Kriterium ein Unterschied der Schwierigkeit der Items zwischen den Gruppen (Differential

Itemfunctioning) > 0,5 logit festgelegt.

	<i>Brieffreundschaft</i>	<i>Brief an denkranken Mitschüler</i>
Geschlecht	7 von 16 Items sind für Jungen schwerer als für Mädchen.	Alle Items sind für Jungen schwerer als für Mädchen.
Migrationsstatus	6 von 16 Items sind für Kinder mit Migrationsstatus schwerer als für Kinder o. M., 4 sind leichter und 6 gleich.	11 von 16 Items sind für Kinder mit Migrationsstatus schwerer als für Kinder o. M., 2 sind leichter und sechs gleich.

Tabelle 2: Ergebnisse der Überprüfung der Aufgabenfairness

Die Befunde zur Kodierübereinstimmung und Fairness waren ausschlaggebend dafür, die Aufgabe *Brieffreundschaft* für den Haupttest auszuwählen.

5.2. Untersuchungsergebnisse zur Schreibaufgabe *Brieffreundschaft*

In Tabelle 3 sind die Ergebnisse der Untersuchung für die Schreibaufgabe *Brieffreundschaft* zusammengestellt: Die Beurteilungskriterien, die der empirischen Überprüfung standgehalten haben, die durchschnittliche Lösungshäufigkeit für jedes Item und die Bewerterübereinstimmung.

Die Kodierübereinstimmung, erfasst über das mittlere Kappa, liegt zwischen 0,19 bei der Bewertung der Rechtschreibung und Grammatik bis 0,9 für die Kodierung des Vorliegens der Unterschrift. Die Kodieranweisung für Rechtschreibung und Grammatik wurde daher präzisiert (s. Anhang 3). Die angestrebten Übereinstimmungen von mindestens 0,7 können, wie oben bereits ausgeführt wurde, für die Kriterien der Schreibaufgabe nicht erreicht werden. Die Trennschärfen der Einzelkriterien liegen zwischen 0,43 und 0,90, wobei die Trennschärfe als Produkt-Moment-Korrelation der Werte eines Raters mit dem Mittelwert aller Rater definiert ist (Wirtz/Caspar 2002, S. 210). Die interne Konsistenz berechnet über Cronbachs Alpha beträgt 0,85 und liegt damit in einem sehr zufriedenstellenden Bereich.

Zur Berechnung der Lösungshäufigkeit wurden alle Items in der Schreibaufgabe *Brieffreundschaft* mit einem Punkt bewertet. Die Lösungshäufigkeiten liegen in der Pilotierung zwischen 25 und 95%, wobei die durchschnittliche Lösungshäufigkeit über alle Items mit 9,5 Punkten 55% beträgt (Standardabweichung=4, N=419). Damit sind die Aufgaben im mittleren Schwierigkeitsbereich zwischen 20 und 80% (Lienert & Raatz, 1998). Zwei Drittel der Schülerinnen und Schüler lösen zwischen 32% und 79% der Aufgabe. Bei der Auswahl der verbliebenen Kriterien wurde auch darauf geachtet, dass die theoretisch begründeten Teilbereiche durch eine Anzahl von Kriterien repräsentiert sind. Die Beurteilung der Gesamtleistung im Schreiben erfolgte über die Zusammenfassung der Kriterien.

Die Korrelation des Richtig-Koeffizienten für die *Items Rechtschreibung/Grammatik* mit der Gesamtlösung der Schreibaufgabe *Brieffreundschaft* beträgt $r=0,45$ ($p<0,000$, $N=121$). Da der Wert $r=0$ bedeutet, dass kein Zusammenhang und der Wert $r=1$, dass ein vollkommener Zusammenhang besteht, zeigt der Wert von $r=0,45$ an, dass zwischen der Einhaltung der Schreibkonventionen (Rechtschreibung/Grammatik) und den Teilkompetenzen Inhalt/

Kriterien der Kodieranleitung	Lösungshäufigkeit in %	Kappa (mittleres)
I. Inhalt		
1) Bezugnahme zu Angaben der Brieffartnerin		
a) Wohnort	25,1	0,30
b) Hobbys	36,0	0,60
2) Ausführungen zu den eigenen Angaben		
a) Erläuterung/Kommentierung/Beispiele zur eigenen Person, Wohnort, Hobbys, Eigenschaften	51,2	0,48
b) Etwas Außergewöhnliches aus dem eigenen Leben wird berichtet, um das Interesse an der Brieffreundschaft zu wecken	34,9	0,60
3) Ausführungen zur Bezugnahme auf die Brieffartnerin		
a) Besonderes Interesse an der Person: Fragen zu den Angaben	33,7	0,75
II. Aufbau (Textsorte Brief)		
a) Anrede/Begrüßungsformel	95,0	0,31
b) Hauptteil (Angaben und Ausführungen zum Thema)	85,4	0,45
c) Abschied (Verabschiedung bzw. Grußformel)	59,3	0,41
d) Unterschrift	62,4	0,90
III. Sprache/Stil		
1) Zur Erläuterung/Kommentierung/Begründung		
a) Verben, die das eigene Denken und Fühlen explizit ausdrücken	62,0	0,48
b) Wertung durch Adjektive	43,5	0,44
c) Nachgestellte Erklärung und Erläuterung	49,8	0,36
2) Sprachliche Komplexität		
a) Satzgefüge (Haupt-Nebensatz-Verbindungen)	56,2	0,36
b) Konjunktionen (komplexe Beziehung)	49,8	0,43
c) Anschluss mit Relativ- oder Fragepronomen	35,9	0,25
IV. Rechtschreibung, Grammatik, Schrift		
a) Beurteilung von Rechtschreibung und Grammatik	87,1	0,19
Richtig-Koeffizient (> 80%)	82,6	-
b) Die Schrift ist leserlich, so dass der Text zu verstehen ist.	91,6	0,36

Tabelle 3: Kodierschema auf der Grundlage der Pilotierungsergebnisse

Aufbau/Sprache und Stil ein mittlerer Zusammenhang besteht.⁹ Von der gesamten auftretenden Varianz (r^2) wird rund ein Fünftel (20%) durch die Beziehung zwischen diesen beiden Variablen aufgeklärt.

Das *Schriftbild* als ein Maß für eine nicht-kognitive Fähigkeit ist bei knapp 10% als nicht lesbar gewertet. Mit der Schreibleistung korreliert das *Schriftbild* zu $r=0,64$ ($p<0,000$, $N=121$). Von der gesamten Varianz werden durch diesen Zusammenhang 41% erklärt, was einem hohen Effekt entspricht. Bei den 10 % Schülerinnen und Schüler mit nicht lesbarer Schrift wird auch die Schreibkompetenz eher schwächer beurteilt.

Die Schwierigkeit der einzelnen Items, die auf der Grundlage der Lösungshäufigkeiten mit probabilistischen Verfahren geschätzt wurde, ist Abbildung 3 zu entnehmen.

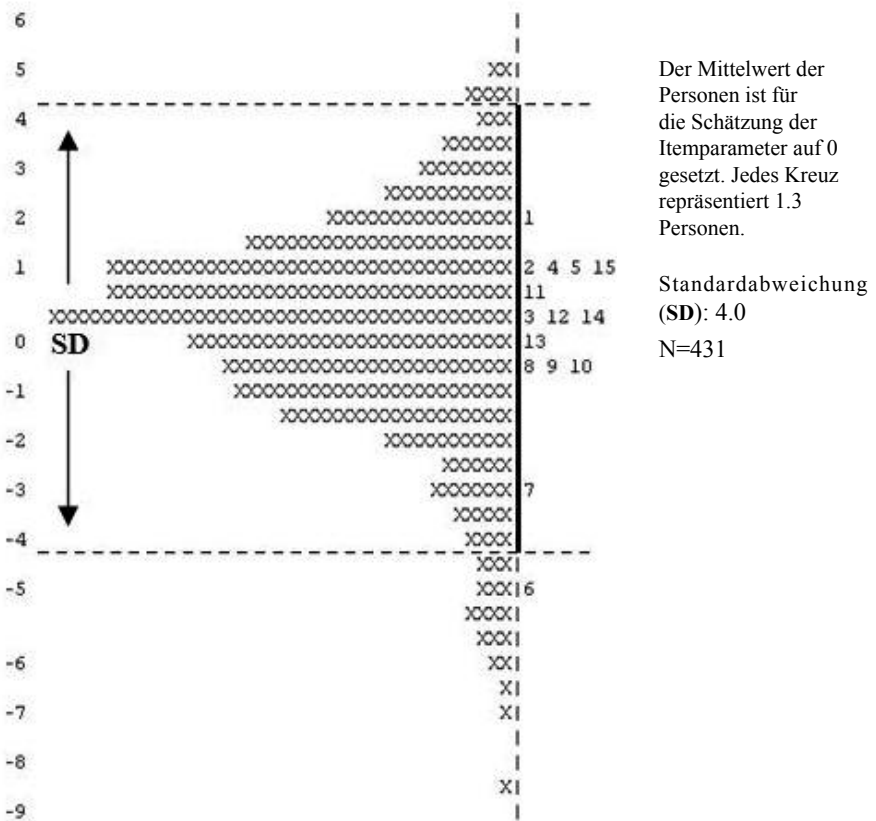


Abbildung 3: Schwierigkeit der einzelnen Items auf der Grundlage der Lösungshäufigkeiten

9 Nach Bortz, 2005 werden die Korrelationswerte wie folgt interpretiert: $r=0,1$ schwacher Effekt, $r=0,3$ mittlerer Effekt, $r=0,5$ großer Effekt.

Liste der Items

I. Inhalt

BEZUGNAHME ZU ANGABEN DER BRIEFPARTNERIN

1. *Wohnort*

2. *Hobbys*

AUSFÜHRUNGEN ZU DEN EIGENEN ANGABEN

3. Erläuterung/Kommentierung/Beispiele zur eigenen Person, Wohnort, Hobbys, Eigenschaften

4. Etwas Außergewöhnliches aus dem eigenen Leben wird berichtet, um das Interesse an der Brieffreundschaft zu wecken

AUSFÜHRUNGEN ZUR BEZUGNAHME AUF DIE BRIEFPARTNERIN

5. Besonderes Interesse an der Person: Fragen zu den Angaben

II. Aufbau (Textsorte Brief)

6. *Anrede/Begrüßungsformel*

7. *Hauptteil (Angaben und Ausführungen zum Thema)*

8. *Abschied (Verabschiedung bzw. Grußformel)*

9. *Unterschrift*

III. Sprache/Stil

ZUR ERLÄUTERUNG/KOMMENTIERUNG /BEGRÜNDUNG

10. Verben, die das eigene Denken und Fühlen explizit ausdrücken

11. Wertung durch Adjektive

12. Nachgestellte Erklärung und Erläuterung

SPRACHLICHE KOMPLEXITÄT

13. Satzgefüge (Haupt-Nebensatz-Verbindungen)

14. Konjunktionen (komplexe Beziehung)

15. Anschluss mit Relativ- oder Fragepronomen

Items, die aus theoretischer Sicht zur elementaren Textkompetenz gehören (s. o. Abb. 2).

Items, die aus theoretischer Sicht zur entfalteten Textkompetenz gehören (s. o. Abb. 2)

Von Interesse ist hierbei, inwieweit die empirische Überprüfung mit den theoretischen Annahmen einer elementaren bzw. entfalteten Schreibkompetenz übereinstimmt. Was die Items des Bereichs II Aufbau betrifft, so lässt sich die theoretische Zuordnung zur elementaren Schreibkompetenz bestätigen. Alle Items in der Kategorie *Aufbau* liegen unterhalb des Mittelwertes und sind also als leicht bzw. sehr leicht einzustufen.

Die übrigen Items liegen alle im Bereich einer Standardabweichung über dem Mittelwert (Items 1, 2, 4, 5, 15, 11, 3, 12, 14), im Mittelwert (Item 13) bzw. knapp darunter (Item 10) und entsprechen somit im Schwierigkeitsniveau den Fähigkeiten der getesteten Schüler.

Von der Vorannahme abweichend erweisen sich die beiden inhaltlichen Items 1 und 2. Es handelt sich dabei um eine Bezugnahme zu den Angaben der Briefpartnerin (Wohnort, Hobbys). Sie wurden aus theoretischer Sicht zur elementaren Schreibkompetenz gerechnet, weil zur Erfüllung dieser Kriterien ein einfacher Bezug zu den Angaben der Briefpartnerin ohne nähere Ausführung als Bewertungsgrundlage ausreichte. Die gemessene Aufgabenschwierigkeit

dieser Items liegt jedoch ebenso hoch bzw. – was die Bezugnahme zum Wohnort betrifft – sogar noch höher als Items, die einer entfalteten Schreibkompetenz zugeordnet wurden. Die der entfalteten Schreibkompetenz zugewiesenen Items verlangen nähere inhaltliche Ausführungen und liegen erwartungsgemäß im oberen Bereich. Die erwartungswidrig hohe Aufgabenschwierigkeit der Items 1 und 2 lässt sich so erklären, dass auch bei einer einfachen Bezugnahme die Leserperspektive eingenommen werden muss. Nach den Skalierungsergebnissen zu schließen, ist die Einnahme einer Leserperspektive auch unabhängig von der inhaltlichen Ausführung den Anforderungen einer entfalteten Schreibkompetenz zuzuordnen. Ein erwartungsgemäß bestätigter Aspekt der entfalteten Schreibkompetenz, der bei den Items 3-5 erhoben wird, ist der inhaltliche Ideenreichtum.

Im Bereich *Sprache/Stil* liegen alle Items über dem Mittelwert, was eine Zuordnung zur entfalteten Schreibfähigkeit prinzipiell rechtfertigt. Den höchsten Wert erzielte der Relativanschluss von Sätzen. Dies ist nach den Ergebnissen der Testaufgaben zu *Sprache und Sprachgebrauch* in Lernstand 6 ebenfalls erwartungskonform. Bei diesen Aufgaben zeigte sich, dass die richtige Verwendung von Relativanschlüssen selbst für Gymnasiasten bzw. Gymnasiastinnen eine zu hohe Herausforderung darstellt.

5.3. Ergebnis der Modellprüfung und Korrelationsanalysen

Um zu untersuchen, ob es sich bei der Textkompetenz um eine globale oder differentielle Kompetenz handelt (Forschungsfrage 3), wurde eine Modellprüfung mit Hilfe der Raschskalierung vorgenommen. Dazu wurden die Auswertungsdaten der Pilotierung herangezogen, die auf einer Kodieranweisung mit 32 Kriterien beruht (s. Anhang 1). Neben einem eindimensionalen Modell, das die Schreibkompetenz (Aufbau, Inhalt, Sprache) global erfasst, wurde auch die Hypothese überprüft, ob die Schreibkompetenz differentiell strukturiert ist. Zum einen wurde ein zweidimensionales Modell geprüft, das zwischen einer elementaren und entfalteten Schreibkompetenz unterscheidet (s. o. Tabelle 2). Zum anderen wurde im Skalierungsmodell nach *Inhalt* und *Aufbau* auf der einen Seite und *Sprache/Stil* auf der anderen Seite unterschieden.

Die Skalierung ergab für das eindimensionale Rasch-Modell die beste Passung, und zwar sowohl für den umfangreichen als auch für den reduzierten Itemsatz. Entgegen der Erwartung, dass sich Items theoretisch begründet gruppieren ließen, zeigte die Datenstruktur der Lösungen höhere Zusammenhänge *aller* Items miteinander. Damit hat sich die Hypothese einer differentiellen Kompetenz im Falle der vorliegenden Schreibaufgabe nicht bestätigt. Damit ist aber noch keine Aussage über andere Schreibaufgaben verbunden. Für eine valide Modellüberprüfung im Hinblick auf differentielle Teilkompetenzen wären umfangreichere Schülertexte sowie eine höhere Zahl von Items in jedem Teilbereich erforderlich. Beispielweise werden in der PISA- und IGLU-Studie mehr als 100 Items eingesetzt, um die Konstruktdimensionalität zu überprüfen. Bei diesen Ergebnissen ist auch zu berücksichtigen, dass die analysierten Daten aus dem Korpus einer Lernstandserhebung stammen, die wissenschaftlich begleitet wurde, und nicht um Grundlagenforschung zur Struktur der Textkompetenz.

Aus didaktischer Sicht ist es auch von Interesse, den Einfluss weiterer schriftsprachlicher

Fähigkeiten auf die Gesamtleistung beim Textschreiben zu ermitteln. Zu diesem Zweck wurden Korrelationen zu den Ergebnissen der Sprach- und Lesetests und eines Orthographietests errechnet.

Die zur Überprüfung der *Sprachkompetenz* eingesetzten Aufgaben bezogen sich auf die Bereiche: Kasus, Konjunktionen, Verben und Zeiten, Wortarten, Wortfamilie und Wortfeld.

Die im Folgenden aufgeführten Ergebnisse aus dem Haupttest stammen aus dem Land Schleswig-Holstein, da wegen des Multimatrixdesigns in der Pilotierung zu wenige Daten für einen Vergleich von Schreib- und Sprachkompetenz zur Verfügung stehen. Im Haupttest wurden in Schleswig-Holstein zwei Testheftvarianten eingesetzt. Die leichtere Testheftvariante wurde von rund 2600 Schülerinnen und Schülern bearbeitet, die anspruchsvollere von 7200.

Die Zusammenhänge zwischen den Kompetenzbereichen stellen sich in den beiden Testheftvarianten in gleicher Rangfolge dar (Tabelle 4). Die Werte in Klammer beziehen sich auf die leichtere Testheftvariante.

Tabelle 4: Zusammenhang zwischen den Leistungen im *Textschreiben*, *Leseverständnis* und *Sprache/Sprachgebrauch*

	Leseverständnis	Sprache/ Sprachgebrauch
Textschreiben	r= 0,25 (0,18)	r= 0,32 (0,27)
Sprache/ Sprachgebrauch	r= 0,59 (0,54)	

Der engste Zusammenhang besteht zwischen der Leistung in *Sprache/Sprachgebrauch* und *Leseverständnis* und der geringste zwischen *Textschreiben* und *Leseverständnis*. Alle Zusammenhänge sind hoch signifikant ($p < 0,00$). Das bedeutet, dass der Zusammenhang der Leistungen in den Kompetenzbereichen *Sprache/Sprachgebrauch*, *Leseverständnis* und *Textschreiben* substantiell ist. Da die Zusammenhänge aber nicht sehr eng ($r > 0,8$) ausfallen, spricht dies dafür, dass es sich um unabhängige Kompetenzen handelt, die auch unabhängig voneinander gemessen werden müssen.

Die Korrelation zwischen einem in der Pilotierung eingesetzten Orthographietest (Wortdiktat) und der Schreibaufgabe *Brieffreundschaft* wurde gesondert berechnet. Der Orthographietest bestand aus 40 Wörtern, die wiederum auf der Grundlage eines sprachsystematischen Rechtschreibkompetenzmodells in 177 Struktureinheiten eingeteilt waren (vgl. dazu Voss/Blatt/Kowalski 2006).

Die Korrelation des Gesamtergebnisses der Schreibaufgabe mit der korrekten Schreibweise der 40 Wörter beträgt $r = 0,57$ ($p < 0,01$, $N = 217$).¹⁰ Es handelt sich dabei um einen mittleren Zusammenhang zwischen der geprüften Text- und Rechtschreibkompetenz. Von der gesamten auftretenden Varianz (r^2) bei der Schreibaufgabe *Brieffreundschaft* wird rund ein Drittel (32%) durch die Beziehung zwischen diesen beiden Variablen aufgeklärt. Hier liegt

¹⁰ Die Korrelation auf Ganzwortebene ist mit 0,57 am höchsten, gefolgt vom Zusammenhang mit den Struktureinheiten Teilkompetenzen phonologisch (0,56), Peripheriebereich (0,55) und Wortbildung (0,53).

zwar ein hoher Effekt vor, aber es kann daraus keine kausale Erklärung abgeleitet werden. Auf Einzelschülerebene kann dies sowohl heißen, dass die Rechtschreibkompetenz die Schreibleistung stark beeinflusst, als auch, dass sie wenig Einfluss darauf ausübt und andere Einflussfaktoren wie Ideengenerierung und –strukturierung, Textsortenwissen oder die gestalterische Sprachfähigkeit im Vordergrund stehen.

Von didaktischem Interesse ist, dass die gemeinsame Varianz zwischen der Leistung im *Textschreiben* und der Leistung in den separaten Tests zur *Orthographie* und zu *Sprache und Sprachgebrauch untersuchen* mit 32 % bzw. 35 % weitaus höher liegt als die gemeinsame Varianz innerhalb der Textkompetenz. Die Teilkompetenzen Inhalt/Aufbau/Sprache und Stil einerseits und die Teilkompetenz Rechtschreibung/Grammatik auf der Grundlage des Richtig-Falsch-Koeffizienten andererseits weisen nur eine gemeinsame Varianz von 20 % auf. Mit den vorliegenden Daten lässt sich jedoch keine Erklärung für diesen Befund ermitteln. Dazu sind längsschnittliche Untersuchungen erforderlich.

6. Zusammenfassung und Einordnung in den Forschungsstand und Fazit

Die Ergebnisse dieser Untersuchung werden abschließend auf der Grundlage der Forschungsfragen zusammengefasst und in den Forschungsstand eingeordnet.

Analog zu den Ergebnissen von Neumann (2007) und Blatt/Voss/Matthießen (2005) kann die Forschungsfrage 2 positiv beantwortet werden, dass theoriebasierte Kriterien auch empirisch abgesichert werden können. Diese Absicherung gelang jedoch nur für ca. die Hälfte der ursprünglichen Items.

Im Unterschied zu Neumann wurden bei Lernstand 6 – analog zu Blatt/Voss/Matthießen – nur dichotome Kriterien eingesetzt. Weiterhin wurden nur solche Kriterien überprüft, die die Textkompetenz im engeren Sinne und nicht die Einhaltung der Sprachkonventionen testen. Rechtschreibung, Grammatik und Schrift wurden daher nur holistisch eingeschätzt. Dass die Interkoderreliabilität bei den Schreibaufgaben zu Lernstand 6 im Bereich *Inhalt* geringer ausfiel als bei Neumann, kann mit den Vorgaben der Aufgabenstellung erklärt werden, die in der Schreibaufgabe *Brieffreundschaft* weitaus offener waren als in den von Neumann untersuchten Schreibaufgaben.

Zur Forschungsfrage 3 nach unterschiedlichen Fähigkeitsniveaus konnte in der Untersuchung zu Lernstand 6 die theoretische Annahme einer elementaren und entfalteten Schreibkompetenz weitgehend bestätigt werden. Während der sehr formale Aufbau der elementaren Schreibfähigkeit zuzurechnen ist, ist vor allem der Leserbezug ein Qualitätsmerkmal für eine entfaltete Schreibfähigkeit. Er kann sich auf inhaltlicher Ebene daran zeigen, ob Inhalte unter Berücksichtigung der Leserperspektive ausgewählt werden, indem sie das Informationsbedürfnis des Lesers befriedigen und dessen Interesse wecken, und ob sie näher ausgeführt, erläutert und kommentiert werden. Auf sprachlicher und stilistischer Ebene wirkt sich qualitätsbestimmend aus, ob sprachliche und stilistische Mittel eingesetzt werden, mit denen eine Textkohäsion hergestellt wird, bzw. mit deren Hilfe die Inhalte für den Leser nachvollziehbar erläutert, erklärt, gewertet, kommentiert und begründet werden.

Diese Ergebnisse sind mit den vorliegenden Befunden von Augst u. a. (2007) und Neumann nicht unmittelbar vergleichbar, da sich die theoretische Modellierung der Textkompetenz

unterscheidet. Augst u. a. weisen textsortenspezifische Fähigkeitsniveaus aus, und zwar in einer Rangfolge vom Erzählen und Instruieren über das Beschreiben zum Argumentieren und Berichten.

Nach Neumanns Befunden bestehen Niveauunterschiede im Hinblick auf den der jeweiligen Textfunktion angemessenen schriftsprachlichen Ausdruck und die Strukturierung des Inhalts in eine logische Abfolge.

Die aus diesen Befunden zu ziehenden didaktischen Folgerungen sind unterschiedlicher Art. Aus den Befunden von Augst u. a. ergibt sich ein textsortenbezogener Schreibunterricht in einer festen Reihenfolge. Aus den Befunden von Neumann ist abzuleiten, dass die Schüler hinsichtlich unterschiedlicher Textfunktionen im sprachlichen Ausdruck und in der logischen Strukturierung des Inhalts geschult werden.

Die Ergebnisse der vorliegenden Untersuchung aus Lernstand 6 sprechen dafür, in erster Linie das Lernziel zu verfolgen, dass sich die Schüler in die Leserperspektive hineinversetzen. Dazu kann eine Verbindung von Literatur- und Schreibunterricht beitragen. Im Literaturunterricht können die Schüler auch stilistische Mittel für eine lesergerechte Ausdrucksweise kennen lernen. In einem verbundenen Sprach- und Schreibunterricht kann das Lernziel verfolgt werden, dass die Schüler kohäsive Mittel kennen und anwenden lernen.

Die Forschungsfrage 4, ob es sich bei der Textkompetenz um eine globale oder eine differenzielle Kompetenz handelt, wird in der vorliegenden Untersuchung zugunsten einer globalen Kompetenz beantwortet. Neumann 2007 dagegen konnte – analog zu den Ergebnissen der Hamburger Aufsatzstudie – die empirische Evidenz für ein zweidimensionales Modell erbringen.

Die Ergebnisse beider Untersuchungen sind jedoch nicht miteinander vergleichbar, da den jeweiligen konfirmatorischen Analysen unterschiedliche Theorieannahmen und empirische Datenmodellierungen zugrunde liegen. Neumann fasste die Bereiche *Inhalt und formaler Aufbau* zur Dimension *Semantik/Pragmatik* und die Bereiche *Sprache, Stil und orthographische und grammatische Korrektheit* zur Dimension *Sprachsystematik* zusammen (Neumann 2007, S. 197ff.).

In der vorliegenden Untersuchung wurden theoretische Modellvarianten überprüft. Zum einen wurde zwischen einer elementaren („assoziatives Schreiben“) und entfalteteten Schreibkompetenz („leserbezogenes und sprachgestaltendes Schreiben“) unterschieden. Zum anderen wurde zwischen Inhalt und Aufbau auf der einen Seite und Sprache/Stil auf der anderen Seite unterschieden, um zu untersuchen, ob die Generierung und Strukturierung von Inhalten und deren sprachliche Gestaltung jeweils eigene Dimensionen darstellen. Die Einhaltung der Schreibkonventionen als Merkmal des performativen Schreibens wurde nicht mit in die Modellprüfung einbezogen. Aus theoretischer Sicht wurde diese Kategorie zunächst nur unter dem Aspekt der Lesbarkeit berücksichtigt und holistisch eingeschätzt. Unter den Bedingungen der Lernstandserhebung konnte eine exakte Auswertung nur für eine kleine Stichprobe vorgenommen werden. Wie oben aufgezeigt, besteht zwischen den Kategorien Inhalt, Aufbau, Sprache/Stil und der Kategorie Rechtschreibung/Grammatik ein mittlerer Zusammenhang ($r=.49$). Zwischen den Ergebnissen der Textkompetenz (ohne Rechtschreibung/Grammatik) und den Ergebnissen der Sprachaufgaben in Lernstand 6 be-

steht dagegen ein geringer Zusammenhang ($r=.32$).

Der gegenwärtige Forschungsstand zur Textkompetenz weist viele offene Fragen auf theoretischer und empirischer Ebene auf. Damit begründet sich die Notwendigkeit für vertiefende empirische Grundlagenforschung zur kognitiven Struktur der Textkompetenz der Schülerinnen und Schüler.

Die Ergebnisse zu den Zusammenhängen der Textkompetenz mit den weiteren schriftsprachlichen Kompetenzen (Lese-, Grammatik-, Rechtschreibkompetenz), auf die sich die Forschungsfrage 5 bezieht, zeigen, dass die Zusammenhänge nachweisbar auf mittlerem Niveau liegen. Es sind also eigenständige Kompetenzen, die jedoch eine gemeinsame Grundlage haben.

Die übergeordnete Forschungsfrage 1 *Lassen sich theoretisch verankerte und unterrichtstaugliche Kriterien zur Erfassung von Textkompetenz entwickeln, die zeitökonomisch einsetzbar und als Lernbeobachtungsinstrument geeignet sind?* soll abschließend differenziert zu beantworten versucht werden.

Die theoretisch verankerten und zugleich empirisch abgesicherten Kriterien ließen sich bereits im Vorwege der Erhebung im Schwierigkeitsgrad weitgehend richtig einschätzen. Das ist bedeutsam für den Unterricht, da es Lehrenden und Schülerinnen und Schülern wichtige Hinweise zum Stand der Schreibkompetenz gibt. Was bislang noch nicht befriedigend gelöst werden konnte, ist die Rückmeldung der Leistungsergebnisse in Prozent-Korrekt-Werten, wobei jedes Item gleich bewertet wird. Nach Einschätzung der Aufgabenentwicklerinnen vermittelt dies einen falschen Eindruck vom Stand der Schreibleistung. In Lernstand 6 (2008/2009) soll daher gemeinsam mit den Aufgabenentwicklerinnen nach einer Alternative gesucht werden. Eine Möglichkeit besteht darin, Items unterschiedlicher Schwierigkeit unterschiedliche Werte zuzuweisen. Inwieweit dies statistisch abgesichert werden kann, muss überprüft werden.

Bei der Entwicklung der Kodieranweisungen für die Schreibaufgaben, die in Lernstand 6 (2008/2009) eingesetzt werden sollen, zeigte sich, dass die Kategorien *Inhalt*, *Aufbau*, *Sprache/Stil* und *Schreibkonventionen* praktikabel sind. Auf ihrer Grundlage konnten zunehmend schneller Kriterien für unterschiedliche Schreibaufgaben entwickelt werden (Anleitung, Bericht). Auch bei der statistischen Überprüfung für diese Items auf der Grundlage der Pilotierungsdaten wurden bessere Werte erzielt. Für eine der neuen Schreibaufgaben (Textsorte Anleitung) ergab sich für alle theoretisch entwickelten Items eine Passung zum Raschmodell.

Die Rückmeldung von den Aufgabenentwicklerinnen zur Unterrichtstauglichkeit und zum zeitökonomischen Einsatz der Kriterien ist generell positiv, wie das folgende Beispiel zeigt: „Wie vereinbart habe ich unsere Kodieranweisung an meinen Aufsätzen erprobt und kann sagen, dass ich richtig gut zurecht gekommen bin. Ich halte die Sache also für praktikabel. Wenn ich die Ergebnisse mit meiner ‚normalen‘ Bewertung vergleiche, stelle ich fest, dass die Kodierung besonders im Bereich Sprache/ Stil genauer funktioniert.“

Ob sich die Kriterien zur Lernbeobachtung eignen, hängt von der Thematisierung im Unterricht ab. Die Lehrkräfte erhalten zusammen mit den Leistungsergebnissen didaktische Hinweise für den Unterricht. Diese zielen darauf, dass die Lernenden mit Hilfe der Kriterien

erfahren, was die Qualität von Texten bestimmt, um auf dieser Grundlage mit Unterstützung ihrer Lehrerinnen und Lehrer gezielt an der Weiterentwicklung ihrer Textkompetenz arbeiten zu können.

Literatur und Internetquellen

Adams, R. J. (2002). Scaling PISA cognitive data. In R. Adams & M. Wu (Hrsg.), PISA 2000 technical report (S. 99-108). Paris: OECD.

Arras, U., Grotjahn, R. (2002). TestDaF: Aktuelle Entwicklungen. Fremdsprachen und Hochschule, 66, 65-88. PDF-Datei (581 KB)

Augst, G., K. Disselhoff, A. Henrich, Th. Pohl, P. L. Völzing (2007): Text - Sorten - Kompetenz: Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter.

Baurmann, J., Dehn, M. (2004). Beurteilen im Deutschunterricht. Praxis Deutsch 31 (184): 6-13.

Bereiter, C. (1980). Development in writing. Cognitive processes in writing: An interdisciplinary approach. . L. Gregg, E.R. Steinberg Hillsdale, NJ, Erlbaum: 1-64.

Blatt, I., Voss, A., Matthießen, I. (2005). Kinder schreiben Briefe aus der Zukunft. Qualitative Analysen von Kindertexten und fachdidaktische Diskussion. IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingen und Zusatzstudien. . W. Bos, E-M. Lankes, et al. Münster et al., Waxmann: 109-157.

Bremerich-Vos, A. (2004). Bildungsstandards, Rechtschreibkompetenz und IGLU – einige Anmerkungen, in: A. Bremerich-Vos u. a. (Hg): Neue Beiträge zur Rechtschreibtheorie und -didaktik. Freiburg 2004, 85-104.

Bortz, J. (2005). Statistik für Human und Sozialwissenschaftler. Heidelberg: Springer.

Dehn, M. (1990). Die Zugriffsweisen 'fortgeschrittener' und 'langsamer' Lese- und Schreibanfänger: Kritik am Konzept der Entwicklungsstufen? Muttersprache Bd. 100, S. 305-316.

Deutsch Englisch Schülerleistungen International (DESI). <http://www.dipf.de/desi/> (letzter Zugriff: 30.8.2008)

Feilke, H., Augst, G. (1989). Zur Ontogenese der Schreibkompetenz. Textproduktion. Ein interdisziplinärer Forschungsüberblick. G. Antos, H.P. Krings. Tübingen, Niemeyer: 297-327. .

Hartmann, W., Blatt, I. (1991). Nützt empirische Forschung dem Schreibunterricht? Folgerungen aus der Hamburger Aufsatzstudie. Schreiben - Schreiben in der Schule. J. Baurmann and O. Ludwig. Hildesheim: 39-66.

Ivo, H. (1982). Lehrer korrigieren Aufsätze. Beschreibungen eines Zustands und Überlegungen zu Alternativen. Frankfurt a.M., Diesterweg.

Lernausgangslagenuntersuchung (LAU). <http://www.hamburger-bildungsserver.de/welcome.phtml?unten=/schulentwicklung/lau/welcome.htm> (letzter Zugriff: 30.8.2008)

Neumann, A. (2007). Briefe schreiben in Klasse 9 und 11. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen Münster et al. , Waxmann.

Nussbaumer, M. (1991). Was Texte sind und wie sie sein sollen. Tübingen, Max Niemeyer Verlag.

- Nussbaumer, M., P. Sieber (1994). Texte analysieren mit dem Zürcher Textanalyseraster. Sprachfähigkeit - besser als ihr Ruf und nötiger den je! Ergebnisse und Folgerungen aus einem Forschungsprojekt. P. Sieber. Aarau, Sauerländer: 141-186.
- Nutz, M. (2003). Beurteilung sprachlicher Leistungen. Didaktik der deutschen Sprache. Ein Handbuch, Teilbd. 2. U. Bredel, H. Günther, P. Klotz, J. Ossner and G. Sieber-Ott. Paderborn, Schöningh: 924-937.
- Ossner, J. (2006 b). Kompetenzen und Kompetenzmodelle im Deutschunterricht. Didaktik Deutsch 21: 5-19.
- Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in Berufs- und Fachschulen (ULME). <http://fhh.hamburg.de/stadt/Aktuell/pressemitteilungen/2006/mai/12/2006-05-12-bbs-ulme-II.html> (letzter Zugriff: 30.8.2008)
- Voss, A., Blatt, I., Kowalski, K. 2007: Zur Erfassung orthographischer Kompetenz in IGLU 2006: Dargestellt an einem sprachsystematischen Test auf Grundlage von Daten aus der IGLU-Voruntersuchung. In: Didaktik Deutsch 23, S. 15-33
- Wirtz, M., Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen: Hogrefe.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). ACER ConQuest. Generalised item response modelling software. Melbourne: Acer Press.

Anschrift der Verfasser:

Prof. Dr. Inge Blatt, Universität Hamburg, Fachbereich Erziehungswissenschaft, Sektion 4 „Didaktik der sprachlichen und ästhetischen Fächer“, Von-Melle-Park 8, 20146 Hamburg

Dr. Gesa Ramm, Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein (IQSH), Qualitätsagentur/IQSH 1, Schreiberweg 5, 24119 Kronshagen

Prof. Dr. Andreas Voss, Hochschule für Angewandte Wissenschaften (HAW) Hamburg, Fakultät Wirtschaft & Soziales, Studiengang Bildung & Erziehung in der Kindheit, Saarlandstrasse 30, 22303 Hamburg