



<https://bulletin.vals-asla.ch> · e-ISSN: 2813-7973 · p-ISSN: 1023-2044

Behrens, U. & Krelle, M. (2011). Schülertexte beurteilen im Licht von Bildungsstandards, Kompetenzrastern und Unterrichtsalltag. *Bulletin suisse de linguistique appliquée*, 94, 167-183.
<https://doi.org/10.5169/seals-978582>

Dieser Artikel ist unter der *Creative Commons Attribution 4.0 International* Lizenz veröffentlicht (CC BY):
<https://creativecommons.org/licenses/by/4.0>



© Ulrike Behrens, Michael Krelle, 2011

Schülertexte beurteilen im Licht von Bildungsstandards, Kompetenzrastern und Unterrichtsalltag

Ulrike BEHRENS & Michael KRELLE

Universität Duisburg-Essen, Fakultät für Geisteswissenschaften – Germanistik
Universitätsstraße 2, D-45141 Essen
ulrike.behrens@uni-due.de & michael.krelle@uni-due.de

In the course of the standard-based diagnostic of writing skills, instruments meeting fairly high psychometric standards have been developed. The evaluation charts used for this purpose, however, are not suitable for teachers' needs, as they are too voluminous or too time-consuming. For classroom purposes, scoring criteria must be reduced, concentrated and modified. The article exemplifies the development of such evaluation charts for teachers using one essay writing task developed in the project "Evaluation and implementation of educational standards – German". Relevant linguistic categories as well as task-specific criteria are considered.

Keywords:

Bildungsstandards Deutsch, Kompetenzraster, Leistungsbeurteilung, Texte bewerten

1. Einleitung: "Die ungerechte Aufsatzzensur"

Im Jahr 1970 macht eine Untersuchung von Schröter von sich reden, die eindrucksvoll die Schwierigkeit der angemessenen und "gerechten" Bewertung von Schüleraufsätzen auf dem Punkt bringt, indem sie in einem gross angelegten Versuch Beurteilerübereinstimmungen ermittelt: Insgesamt 617 Aufsätze wurden von jeweils mindestens zehn Lehrpersonen¹ bewertet. Die Ergebnisse sind ernüchternd: Über 10 % der Texte erhalten alle Zensuren zwischen "sehr gut" und "mangelhaft" (1-5) bzw. zwischen "gut" und "ungenügend" (2-6), in 95 % der Fälle umfassen die abgegebenen Bewertungen mindestens drei Schulnoten. Der Befund ist inzwischen vielfach repliziert worden, u.a. 2002 von Birkel & Birkel, die 88 Lehrpersonen ohne Vorgabe eines Auswertungsschemas jeweils vier Texte von ViertklässlerInnen bewerten liessen.

Als verantwortlich für die mangelnde Objektivität der Aufsatzbeurteilung erweisen sich verschiedene Faktoren, u.a. der Umfang der zu bewertenden Texte und Halo-Effekte grammatikalischer und orthografischer Fehler (Birkel & Birkel, 2002) sowie der Handschrift (Briggs, 1970). In weiteren Studien fand man Varianten des Rosenthal-Effekts sowie Einflüsse der Beliebtheit der SchülerInnen (Hadley, 1954 zit. n. Ingenkamp, 1971); Baumann (1995) weist ausserdem auf

¹ Teilweise liegen Urteile von 35 LehrerInnen vor, im Schnitt gab es 18 Noten pro Aufsatz (Schröter, 1970: 18).

Reihenfolgeeffekte hin. Auch für den von Studierenden und Lehrpersonen als "sichere Bank" wahrgenommenen Bereich der Bewertung und Beurteilung orthografischer Leistungen berichtet Birkel (2009) erst kürzlich ähnliche Befunde.

Es geht hier also nicht in erster Linie um die Ermessensspielräume, die Lehrpersonen ausnutzen, indem sie – je nach Unterrichtskontext – Gewichtungen vornehmen oder wenn sie in motivierender oder aufrüttelnder Absicht "pädagogisch" benoten. Die berichteten Ergebnisse beziehen sich durchgängig auf Effekte, die sich "hinter dem Rücken" der Lehrpersonen auswirken und im Unterrichtsalldag gute Chancen haben, unbemerkt zu bleiben.

In den letzten Jahren werden nun auch im deutschsprachigen Raum zunehmend grössere Lernstandserhebungen durchgeführt, von zentralen Abschlussarbeiten bis hin zu landesweiten Vergleichsarbeiten. In allen Fällen spielt die Objektivität und Vergleichbarkeit der Leistungsbewertungen eine herausgehobene Rolle, sei es, weil es um die Zuweisung von künftigen Karrierechancen geht wie im Falle des Zentralabiturs, sei es, weil Bildungsforschung und -politik auf empirisch robuste Daten angewiesen sind, wie im Falle vergleichender Gruppentestungen (z.B. PISA, IGLU).

Bei der Erhebung von Schülerleistungen im Kompetenzbereich "Schreiben" (vgl. DESI, Vergleichsarbeiten) stellt sich also die Frage, wie eine grösstmögliche Objektivität der Leistungsmessung gewährleistet werden kann. Neben der Berücksichtigung allgemeingültiger Testgütekriterien spielen hier die Auswertungsinstrumente eine bedeutende Rolle. Am Beispiel der bundesdeutschen Normierung der Bildungsstandards im Kompetenzbereich Schreiben soll im Folgenden anhand einer argumentativen Schreibaufgabe aus der Primarstufe ("Klassenfahrt") ein Auswertungsschema vorgestellt werden, mit dem versucht wird, diesen Ansprüchen gerecht zu werden.

In einem weiteren Schritt sollen Möglichkeiten der Implementation diskutiert werden: Im Idealfall sollen Schulen und Lehrpersonen ja nicht nur von den zurückgemeldeten Testergebnissen sowie ggf. von einer sich allmählich in Richtung Kompetenzorientierung verändernden Aufgabenkultur profitieren. Wünschenswert wäre vielmehr zusätzlich eine Rückwirkung der in diesem Zusammenhang entwickelten und erprobten Messinstrumente auf den Schulalltag. Die Frage ist also nicht nur, mithilfe welcher Instrumente in large-scale-Untersuchungen Schreibkompetenz "gerecht" bewertet werden kann, sondern auch, wie man von solchen Instrumenten zu für die schulische Praxis nützlichen Werkzeugen kommt.²

² Gelungene Beispiele eines solchen Transfers finden sich noch zu selten. Einschlägig sind beispielsweise die Arbeiten von Becker-Mrotzek & Böttcher (2006), die das so genannte "Zürcher Textanalyseraster" (Nussbaumer, 1996) für die Belange des Deutschunterrichts anpassen und auf einem mittleren Abstraktionsniveau verschiedene Erwartungshaltungen formulieren.

2. Schreibkompetenz in den KMK-Bildungsstandards

Das Schreiben zählt zu den wesentlichen Kompetenzen, die Schülerinnen und Schüler in der Schule erwerben sollen. Im Deutschunterricht werden Schreibkompetenzen entwickelt, um zu kommunizieren, um Informationen aufzuzeichnen, um sich gedanklich mit Inhalten auseinanderzusetzen sowie zum kreativen und gestalterischen Umgang mit Sprache. Die Schreibprozesse und -produkte dienen auch dazu, die eigene Meinung zu äussern und argumentativ überzeugend darzulegen.

Zumeist wird Schreibkompetenz dabei als ein hochkomplexes Fähigkeitsbündel angesehen, das sich aus verschiedenen Teilkompetenzen zusammensetzt. In diesem Sinne heisst es in den Bildungsstandards der KMK für die Grundschule:

Die Schülerinnen und Schüler [...] gestalten den Schreibprozess selbstständig und verfassen ihre Texte bewusst im Zusammenhang von Schreibabsicht, Inhaltsbezug und Verwendungszusammenhang. Dabei greifen die Teilprozesse des Schreibens ineinander: Texte planen, aufschreiben und überarbeiten. (KMK, 2005: 8)

Mit dem expliziten Ziel, die Bildungsstandards zu präzisieren, weiterzuentwickeln und mithilfe von Aufgaben bzw. Tests zu operationalisieren, wurde gleichzeitig mit der Verabschiedung der Bildungsstandards das Berliner Institut zur Qualitätsentwicklung im Bildungswesen (IQB) gegründet. Hier werden im Auftrag der Kultusministerkonferenz der Bundesländer seit 2005 Testinstrumente entwickelt, die sich dazu eignen, erreichte Leistungsstände der Schülerinnen und Schüler festzustellen und mit den Anforderungen der Bildungsstandards abzugleichen. Die entsprechenden Projekte im Fach Deutsch bilden den Hintergrund des vorliegenden Beitrages.

Der Kompetenzbereich "Schreiben" gliedert sich in den KMK-Bildungsstandards in die Aspekte "über Schreibfertigkeiten verfügen", "richtig schreiben" und "Texte verfassen", die weiter unterteilt werden (vgl. Abb. 1).

Es fällt ins Auge, dass sich einige dieser (Sub-)Standards der normierten Gruppentestung entziehen – hier ist besonders die Entwicklung von geeigneten Verfahren für den Unterricht gefordert. Andere Bereiche, wie z.B. die orthografischen Leistungen, wurden in spezifischen Tests erfasst. Im Folgenden geht es um diejenigen Kompetenzen, die sich valide nur auf der Basis frei formulierter, längerer Schülertexte ermitteln lassen, im Kern also um die Substandards, die unter dem Namen "Texte schreiben" firmieren.

Substandard	Charakterisierung
über Schreibfertigkeiten verfügen	eine gut lesbare Handschrift flüssig schreiben
	Texte zweckmässig und übersichtlich gestalten
	den PC – wenn vorhanden – zum Schreiben verwenden und für Textgestaltung nutzen
richtig schreiben	geübte, rechtschreibwichtige Wörter normgerecht schreiben
	Rechtschreibstrategien verwenden: Mitsprechen, Ableiten, Einprägen
	Zeichensetzung beachten: Punkt, Fragezeichen, Ausrufezeichen, Zeichen bei wörtlicher Rede
	über Fehlersensibilität und Rechtschreibgespür verfügen
	Rechtschreibhilfen verwenden – Wörterbuch nutzen, – Rechtschreibhilfen des Computers kritisch nutzen
	Arbeitstechniken nutzen – methodisch sinnvoll abschreiben – Übungsformen selbstständig nutzen – Texte auf orthographische Richtigkeit überprüfen und korrigieren
Texte verfassen	
Texte planen	Schreibabsicht, Schreibsituation, Adressaten und Verwendungszusammenhang klären
	sprachliche und gestalterische Mittel und Ideen sammeln: Wörter und Wortfelder, Formulierungen und Textmodelle
Texte schreiben	verständlich, strukturiert, adressaten- und funktionsgerecht schreiben: Erlebtes und Erfundenes; Gedanken und Gefühle; Bitten, Wünsche, Anforderungen und Vereinbarungen; Erfahrungen und Sachverhalte
	Lernergebnisse geordnet festhalten und auch für eine Veröffentlichung verwenden
	nach Anregungen (Texte, Bilder, Musik) eigene Texte schreiben
Texte überarbeiten	Texte an der Schreibaufgabe überprüfen
	Texte auf Verständlichkeit und Wirkung überprüfen
	Texte in Bezug auf die äussere und sprachliche Gestaltung und auf die sprachliche Richtigkeit hin optimieren
	Texte für die Veröffentlichung aufbereiten und dabei auch die Schrift gestalten

Abbildung 1: Substandards im Bereich "Schreiben" und ihre Charakterisierung (KMK, 2005: 11)

3. Testung von Schreibkompetenz

Kompetenz zeigt sich nach Hartig & Klieme (2007: 13)

im je situativen Bewältigen von Anforderungen (in der "Performanz" des Handelns), wird aber als Disposition interpretiert. Dementsprechend ist Kompetenz kontextualisiert und spezifisch, aber auf Transfer und Verallgemeinerung angelegt. Kompetenz bezieht sich sowohl auf Handlungsvollzüge als auch auf die ihnen zugrunde liegenden mentalen Prozesse und Kapazitäten, zu denen Kognition, Motivation und Volition bzw. Wissen und Können gehören. Im wissenschaftlichen Kontext verbindet sich mit dem Kompetenzkonstrukt die Erwartung, dieses Wissen und Können in ökologisch valider Weise beschreiben und womöglich messen zu können.

Für die Ermittlung von Schreibkompetenz folgt aus dieser verbreiteten Lesart, dass Anlässe für das Verfassen frei formulierter Texte geschaffen werden müs-

sen, damit Schülerinnen und Schüler ihre Fähigkeit zur Bewältigung realistischer Schreibanforderungen zeigen können.³ Aus der erfolgreichen Bewältigung dieser situativen Handlungsanforderungen kann mit einiger Sicherheit auf die zugrunde liegende Kompetenz des Schreibers bzw. der Schreiberin geschlossen werden, denn "gute Texte kommen nur selten zufällig zustande; sie erfordern in der Regel eine ausgebaute Schreibfähigkeit." (Becker-Mrotzek & Böttcher, 2006: 90). Ein misslungener Text hingegen kann viele Ursachen haben, weswegen hier bei Schlussfolgerungen einige Vorsicht angebracht ist. In grossen Leistungsstudien bezieht sich diese Vorsicht vor allem auf die Reichweite der Aussagen, die man auf Basis der Daten machen kann. Zwar werden pro Schüler jeweils zwei Texte zu unterschiedlichen Textsorten erhoben, dennoch erfolgt die Darstellung der Ergebnisse ausschliesslich gruppenbezogen, damit situative und individuelle Faktoren möglichst nicht ins Gewicht fallen.

Zugleich findet aber – anders als beispielsweise in Schröters Untersuchung – die Auswertung der Texte nicht nur global statt, sondern hinsichtlich verschiedener Kriterien, die sich aus den Teilaspekten der KMK-Bildungsstandards ableiten. Auf diese Weise können an einen einzelnen Text eine grosse Zahl von Kriterien angelegt werden, die Rückschlüsse auf Teilkompetenzen des Schreibens auf Schülerseite erlauben.

3.1 Komplexe Schreibaufgaben

Um im Rahmen der Bildungsstandards-Normierung die Schreibkompetenz von bundesdeutschen Grundschülerinnen und Grundschülern zu ermitteln, wurden insgesamt 15 Schreibaufgaben eingesetzt, in denen die SchülerInnen längere eigene Texte zu verfassen hatten. Einige dieser Aufgaben wurden nur in der dritten, andere nur in der vierten und eine Schnittmenge in beiden Jahrgangsstufen eingesetzt.

In der sehr umfangreichen Studie zur Pilotierung der Aufgaben bearbeiteten 5032 Dritt- und Viertklässler jeweils zwei Schreibaufgaben,⁴ im Testheft kombiniert wurden immer unterschiedliche Textsorten und -muster. Pro Aufgabe und Klassenstufe liegen zwischen 260 und 540 Schülertexte vor. Die Konzeption der Normierungsstudie entspricht weitgehend der der Pilotierung; sie wurde im Frühjahr 2007 durchgeführt. Die Normierungsstudie wird die Grundlage für die Definition von nationalen Skalen für die dritte und vierte Jahrgangsstufe in der Grundschule liefern, über die spätestens 2011 berichtet werden soll.

³ Nur in Ausnahmefällen, wie etwa beim Ausfüllen von Formularen, bietet sich eine Leistungsmessung in Form von klassischen Testitems an.

⁴ Die Pilotierungsstudie fand im Rahmen der Datenerhebung der Internationalen Grundschul-Lese-Untersuchung (IGLU, 2006) statt.

Die Schülerinnen und Schüler sollten sowohl im Rahmen der Pilotierungs- als auch der Normierungsstudie entsprechend gängiger schulischer Textsorten und Textmuster *narrative, informative und argumentative Texte* schreiben (vgl. Fix, 2006). Dazu wurden Bilder, Texte oder Text-Bild-Kombinationen mit entsprechenden Aufgabenstellungen vorgegeben.

3.2 *Argumentatives Schreiben*

Dass auch bei den Aufgaben für die Grundschule auf Argumentation gesetzt wurde, ist bemerkenswert: Zwar gibt es eine Reihe empirischer Hinweise darauf, dass Kinder bereits früh, jedenfalls weit vor Schulbeginn, bei Bedarf mündlich argumentieren können. In der Grundschule spielt das schriftliche Argumentieren bisher aber eine untergeordnete Rolle; man konzentriert sich häufig auf das schriftliche Erzählen.

Wie andere Typen komplexer sprachlicher Handlungen (z.B. eine Geschichte erzählen, eine Wegauskunft geben, ein Spiel erklären) dienen Argumentationen dazu, ein Problem mit sprachlichen Mitteln zu lösen. Um das Schreiben argumentativer Texte bei Schülerinnen und Schülern zu evozieren, sind eine Reihe von Bedingungen notwendig: In Bezug auf das *Schreibziel* ist es beispielsweise nötig, etwas Strittiges vorzugeben, zu dem die Schülerinnen und Schüler begründet Stellung nehmen können, oder ein Problem, das sie im Rahmen der Aufgabe lösen können. Das Strittige sollte zudem auf irgendeine Weise in der Lebenswelt der Schülerinnen und Schüler verankert sein. Bei der Aufgabenstellung sind u.a. die Modalität der Themenentfaltung, Textsorten-spezifisches und Inhaltliches zu beachten: Lässt sich auf Grundlage der Aufgabenstellung beispielsweise überhaupt problem- bzw. objektbezogen argumentieren? Sollen Aspekte der Adressatenorientierung eine Rolle spielen? Wie lässt sich im Rahmen der Vorgaben und der Aufgabenstellung plausibel argumentieren? Spielen Gegenargumente eine Rolle, werden z.B. schon Argumente vorgegeben, denen man widersprechen soll?

Solche Fragen waren auch bei der folgenden Aufgabe "Klassenfahrt" leitend, die im Rahmen der Pilotierungsstudie eingesetzt wurde, aus Gründen des Testdesigns allerdings nicht Teil der Normierung war. Grundlage des Folgenden sind also Daten, die im Rahmen der Pilotierung erhoben wurden.

3.3 Aufgabenbeispiel: Klassenfahrt

Die Klassenfahrt

Die Klasse 4a möchte eine 3-tägige Klassenfahrt machen.
Zwei Ziele stehen zur Auswahl: Weinhausen am Breitsee und
Landstein an der Murre. Bald soll in der Klasse entschieden werden.
Lies dir die Informationen zu den beiden Zielen durch:

Weinhausen am Breitsee



<u>In Weinhausen gibt es:</u> - einen Abenteuerspielplatz - einen Grillplatz - einen Naturlehrpfad	<u>Ausflugsmöglichkeiten:</u> - eine Badestelle (Freibad) - eine Tropfsteinhöhle - eine Staumauer
--	---

Anfahrt: 1 Stunde mit dem Bus; Kosten für jeden Schüler: 8 Euro
Übernachtung: kostet in der Jugendherberge pro Schüler 30 Euro

Landstein an der Murre



<u>In Landstein gibt es:</u> - einen Minigolfplatz - einen Bolzplatz - eine alte Wassermühle	<u>Ausflugsmöglichkeiten:</u> - ein Picknickplatz im Wald - eine Sommerrodelbahn - ein Freilichtmuseum
--	--

Anfahrt: 2 Stunden mit dem Zug; Kosten für jeden Schüler: 12 Euro
Übernachtung: kostet im Schullandheim pro Schüler 35 Euro

1. Stell dir vor, du würdest mitfahren!
Wohin würdest du lieber fahren?

- 0 Weinhausen
0 Landstein

Abbildung 2: Schreibaufgabe "Klassenfahrt"

Bei der Aufgabe "Klassenfahrt" haben die Schreibenden die Wahl zwischen zwei verschiedenen Reisezielen, die insgesamt gleichwertig sind, jedoch je nach individuellen Vorlieben, Überzeugungen und Erfahrungen erkennbare Vor- bzw. Nachteile haben. So spielen in Hinblick auf das Ausflugsziel verschiedene Überlegungen eine Rolle, z.B. die Freizeitmöglichkeiten "vor Ort" oder auch Überlegungen zu den "Ausflugsmöglichkeiten". Schliesslich können auch die Kosten, die Fahrtzeit und die Unterbringung argumentativ verarbeitet werden.

Um die Schreibenden in Hinblick auf das Aufgabenziel zu entlasten, wurde eine Entscheidungsfrage (Stell dir vor, du würdest mitfahren! Wohin würdest du lieber fahren? "Weinhausen" oder "Landstein") im geschlossenen Format vorgegeben und dem eigentlichen Schreibauftrag vorangestellt. Dieser lautet dann wie folgt:

2. Versuche, die Kinder der Klasse 4a von **deiner Wahl** zu überzeugen!

Schreibe dafür einen Text! Du kannst die Kinder dann überzeugen, wenn du deine Wahl gut begründest.

Nutze dazu die Informationen auf den vorherigen Seiten!



Abbildung 3: Aufgabenstellung Schreibaufgabe "Klassenfahrt"

Insgesamt wurden zwei linierte Seiten vorgegeben. Die Notwendigkeit des Argumentierens ergibt sich aus dem Anspruch, die "Kinder der Klasse 4a" zu überzeugen. In dieser Situierung dient das Schreiben sicherlich auch der Selbstvergewisserung, primär ist jedoch ein an die Kinder gerichteter Text in der Aufgabe angelegt. Dieser muss keinen formalen Vorgaben (z.B. der Briefform) entsprechen, sondern kann ganz auf die Qualität der Argumentation hin untersucht werden. Dabei spielt die Adressatenorientierung eine besondere Rolle: Es ist zu berücksichtigen, was für MitschülerInnen überzeugende Argumente sind.

Thematisch dürfte es Viertklässlern leicht fallen, die Frage nach dem Ort einer Klassenfahrt mit ihren eigenen Erfahrungen zu verknüpfen und bedeutsame Begründungen für die eigene Wahl zu finden. Die Aufgabe stellt insofern einen kommunikativ und lebensweltlich relevanten Schreibanlass dar.

Für die Auswertung ist besonders bedeutsam, dass einerseits die vorgegebene Rahmung eine Vergleichbarkeit der Lösungen ermöglicht, andererseits individuelle Lösungen gefragt sind, die das freie Schreiben überhaupt erst rechtfertigen und ausserdem für die nötige inhaltliche Differenz der resultierenden Schülertexte sorgen. In diesem Spannungsfeld muss ein Auswertungsschema bestehen, dessen Kriterien an alle Texte angelegt werden können, und das gleichzeitig die nötige Differenzierung zwischen gelungenen und weniger gelungenen Texten erlaubt.

4. Darstellung des Auswertungsschemas⁵

Aus den beschriebenen Charakteristika und Qualitätsanforderungen der Aufgabe ergab sich ein erstes, überaus differenziertes Auswertungsschema, bestehend aus 45 Einzelvariablen, die teils dreistufig (zumeist Sprachvariablen), teils dichotom (zumeist Inhaltsvariablen) kodiert wurden.

Eine solche Kodierung kann als "analytisch" bezeichnet werden, d. h., bei der Kodierung wird ein Konstrukt sehr fein ausdifferenziert und in seine einzelnen Bestandteile zerlegt.

Diese Einzelbestandteile, wie z.B. bestimmte inhaltliche oder sprachliche Aspekte, werden dann zumeist dichotom bzw. auf Skalen mit nur wenigen, etwa drei, Stufen hinsichtlich ihres Auftretens (ja/nein) bewertet und anschließend aggregiert. Beispielsweise können für die Beurteilung des Inhalts eines Aufsatzes vorgegebene Schlagwörter oder Schlüsselphrasen im Text identifiziert und diese einzeln und unabhängig voneinander hinsichtlich ihres Vorhandenseins dichotom (0/1) kodiert werden. So verstanden sind einzelne Wörter bzw. Wortgruppen die Basis einer analytischen Kodierung. (Böhme, Bremerich-Vos & Robitzsch, 2009: 319)

⁵ Die im Folgenden verwendeten Schülertexte sind authentische (unveränderte) Aufgabenlösungen.

Allgemeines	
1. Schriftbild/Lesbarkeit (dreistufig)	3. Textlänge (Anzahl der Wörter)
2. Gesamteindruck vom Text (vierstufig)	4. Beendigung des Textes (dreistufig)
Inhalt	
5. These/Vorschlag (dichotom)	
6. Expliziter Adressatenbezug (dichotom)	
Argumente	
7. Abenteuerspielplatz/Minigolfplatz (dichotom)	14. An-/Abreise (dichotom)
8. Grillplatz/Bolzplatz (dichotom)	15. Eigenschaften des nicht gewählten Ortes (dichotom)
9. Naturlehrpfad/Wassermühle (dichotom)	16. Weitere Argumente (dichotom)
10. Badestelle, Freibad/Picknickplatz im Wald (dichotom)	17. Vorwegnahme denkbarer Gegenargumente (dichotom)
11. Tropfsteinhöhle/Sommerrodelbahn (dichotom)	18. Gewichtung von Argumenten (dichotom)
12. Staumauer/Freilichtmuseum (dichotom)	19. Ausgestaltung von Argumenten (dichotom)
13. Kosten (dichotom)	
Bewertungen	
20. Wertungen (dichotom)	22. Adressatenbezogenheit der Begründungen
21. Begründung von Wertungen (dichotom)	
Wertbezüge	
23. Lernmöglichkeiten (dichotom)	27. Exklusivität/Einmaligkeit (dichotom)
24. Vielfalt (dichotom)	28. Kosten (dichotom)
25. Erholung/Gesundheit (dichotom)	29. Weiteres (dichotom)
26. Natur (dichotom)	
Stimmigkeit	
30. Inhaltliche Schlüssigkeit der Argumentation (dichotom)	
Struktur	
31. Überschrift (dichotom)	34. Durchgängiger argumentativer Charakter des Textes (dichotom)
32. Einleitung (dichotom)	
33. Argumentatives Gesamtkonzept (dichotom)	35. Abschluss (dichotom)
	36. Absätze (dichotom)
Sprache – argumentativ	
37. Gebrauch von Verknüpfungen (dreistufig)	
Sprache – allgemein	
38. Wortschatz/Wortwahl (dreistufig)	42. Grammatik (dreistufig)
39. Orthografie (dreistufig)	43. Wiederaufnahmen (dreistufig)
40. Sätze/Satzbau allgemein (dreistufig)	44. Tempusgebrauch (dreistufig)
41. Satzgrenzen (dreistufig)	

Abbildung 4: Variablen zur Schreibaufgabe "Klassenfahrt"⁶

⁶ Es wurden drei weitere globale Variablen zu Inhalt, Sprache und Aufbau entwickelt, die – zusammen mit den analytischen Kategorien – bei der Definition von Kompetenzabstufungen eine Rolle spielen (vgl. Böhme, Bremerich-Vos & Robitzsch, 2009). Diese sind hier nicht abgedruckt.

Bei einer "holistischen" Kodierung hingegen wird den BewerterInnen eher ein Globalurteil abverlangt, welches die Leistung des Schülers insgesamt auf einer oft mehrstufigen Ratingskala einordnet. Eine solche Entscheidung muss etwa beim "Gesamturteil" (vgl. Abb. 4, Variable Nr. 2) getroffen werden. Die Grenzen zwischen "analytischem" und "holistischem" Kodieren können allerdings auch fließend sein. Beispielsweise ist nach der obigen Definition partiell holistisches Kodieren im Spiel, wenn der Wortschatz (Abb. 4, Variable 38) in Hinblick auf den gesamten Text anhand einzelner Wörter bzw. Wortgruppen eingeschätzt wird.

4.1 Zur Reliabilität der Instrumente

Entscheidendes Kriterium für die Qualität des Auswertungsschemas in einer large-scale-Untersuchung ist das Mass der Übereinstimmung verschiedener KodiererInnen desselben Textes. Dieses Mass hängt dabei zunächst von zwei Bedingungen ab: 1. Die Übereinstimmung ist umso höher, je weniger Stufen die Skala für eine Variable umfasst; 2. sie ist umso höher, je klarer das Kriterium auf der Textoberfläche zutage tritt, je weniger inferent die Variable also ist. So gibt es beispielsweise zum Kriterium "These/Vorschlag" (Variable 5) nur sehr wenige Zweifelsfälle. Kodiert werden soll, ob der gewählte Vorschlag (Weinhäuser oder Landstein) ausdrücklich genannt wird – die Übereinstimmungswerte sind auch bei ungeschulten KodiererInnen nahe Kappa (κ) = 1⁷. Eine hohe Übereinstimmung ist bei dieser Variable insofern erwartbar, als es vergleichsweise eindeutige Hinweise gibt, die einen Richtig-Code rechtfertigen, z.B. wenn Schülerinnen und Schüler schreiben "Wir sollten/wollen nach ... fahren." Oder "... ist das bessere Klassenfahrtsziel." Der auf drei Stufen einzuschätzende Wortschatz erweist sich hingegen als problematisches Kriterium, sofern ja von dem Wortmaterial des Textes auf einen zugrunde liegenden Wortschatz des Schülers bzw. der Schülerin geschlossen werden muss. Hier lag die Übereinstimmung bei ungeschulten Raterinnen und Ratern zunächst bei κ = 0.7, nach zwei Schulungen mit je 15-20 Beispieltexen bei immerhin bei κ = 0.84.

Die Entwicklung eines Auswertungsschemas stellt also immer eine Gratwanderung zwischen objektiver Beurteilbarkeit einerseits und angemessener Bewertung andererseits dar. Das hier vorgestellte Auswertungsschema erwies sich nach intensiven Schulungen der KodiererInnen insgesamt als geeignet für die Datengewinnung in einer wissenschaftlichen Studie. Es konnten in einer Gruppe von KodiererInnen variablenbezogene Übereinstimmungen von mindestens κ =

⁷ Cohens Kappa (κ) ist ein zufallsbereinigtes Mass der Raterübereinstimmung. Der Koeffizient kann Werte zwischen -1 und +1 annehmen und ist für nominale Daten geeignet. In der Literatur finden sich verschiedene Empfehlungen, welcher Wert als Mindestanforderung interpretiert werden kann. Wirtz & Caspar (2002) verweisen auf Studien, die – je nach Kriterium und Anzahl der Codes – bereits bei einem Wert von κ > .40 von einer zufriedenstellenden Reliabilität sprechen, andere nennen Werte von κ > .70 oder > .75.

0.6 (bei dreistufigen Variablen) und $\kappa = 0.8$ und höher (bei dichotomer Kodierung) erreicht werden. Grundlage waren drei Schulungstermine, verteilt über ca. drei Wochen, bei denen parallel kodierte Texte analysiert und diskutiert wurden, um zu einer gleichsinnigen Lesart der Kodieranweisungen zu kommen. Im Folgenden werden einige der oben ausgeführten Auswertungskategorien näher erläutert.

4.2 Zum Auswertungsschema: Beispiele

Während unter "Allgemeinem" (vgl. Abb. 4) übergreifende Kategorien zu finden sind, verbindet sich mit den "*inhaltlichen*" Kriterien der Anspruch, die Varianz der Schülerlösungen in Hinblick auf die thematische, also argumentative Umsetzung der Aufgabenlösung abzubilden. Anders als beispielsweise bei Kriterien zur sprachlichen Angemessenheit und Richtigkeit (Variable 37-44) handelt es sich hierbei um "aufgabenspezifische Kriterien". So wird beispielsweise (dichotom) kodiert, ob "ein gewählter Vorschlag genannt wird" (Variable 5) und ob ein direkter Adressatenbezug im Text sprachlich ausgedrückt wird (Variable 6). Dabei haben viele der Kriterien rein deskriptiven Charakter – sie können detaillierten Auswertungen zur Textkonstruktion dienen, eignen sich aber nur bedingt für einen bewertenden Zugriff. So ist z.B. die Frage, auf welche Überlegungen sich die Argumentation bezieht (Variablen 7-16), für *Arten von Aufgabenlösungen* von Interesse, für die Qualität der Argumentation ist jede Variable (für sich genommen) jedoch kein hinreichender Prädiktor; ein Verrechnen bzw. "Summieren" solcher Codes verhilft vermutlich nicht dazu, die Leistungen eines Schülers adäquat einzuschätzen. Andererseits können sich einzelne Kriterien als besonders bedeutsam für die argumentative Qualität herausstellen. Ein Beispiel, die "Vorwegnahme von Gegenargumenten" (Variable 17), ist im Folgenden dokumentiert. Zur Veranschaulichung sind die Original-Kodieranweisungen abgedruckt.

Vorwegnahme von denkbaren Gegenargumenten	Ein oder mehrere denkbare Gegenargumente werden entkräftet. Dabei kann es u.a. um Bedenken bezüglich der Gültigkeit von angeführten Argumenten gehen (als Antwort auf den Einwand: Dein Argument trifft nicht oder nur zum Teil zu.) sowie um die Entkräftung möglicher Verweise auf Nachteile des favorisierten Vorschlags.	Erfüllt: "Jeder der keine lust auf den Bolzplatz zu gehen kann ja Minigolf spielen." "und die Länge der Fahrt ist für mich total in ordnung." "Es ist zwar teurer aber bestimmt viel schöner." Nicht erfüllt: "Ich habe Landstein an der Murre/ gewelt weil ich gerne mal eine alte Wassermühle/ sehen möchte und gerne Minigolf/ spiele. Mit der Sommerrodelbahn zu fahren/ macht bestimmt viel Spaß und/ ein Picknickplatz im Wald ist/ bestimmt toll. In einem/ Freilichtmuseum war ich/ noch nie."
---	--	---

Abbildung 5: Beispiele für Inhaltsvariablen (dichotom)⁸

⁸ Unabhängig davon, welche Variable zu kodieren ist, besteht die Notwendigkeit, die zu beurteilenden Kriterien zu beschreiben und mithilfe von Schülerbeispielen zu illustrieren, die entweder die

Neben der "Gewichtung" (Variable 18) und der "Ausgestaltung von Argumenten" (Variable 19) firmieren weitere inhaltliche Kriterien unter dem Label "Wertung", d.h., ob "eine oder mehrere Eigenschaften des gewählten Reiseziels positiv bewertet" werden (Variable 20) und ob dafür Gründe angeführt werden, was über eine einfache Attribution hinausgeht (Variable 21). Zusätzlich wurde auch kodiert, ob im Text jeweils ein expliziter Hinweis auf den normativen Hintergrund zu finden ist (Variablen 23-29) und ob die Argumentation inhaltlich schlüssig ist (Variable 30).

In struktureller Hinsicht wurden hauptsächlich solche Kriterien kodiert, die sich auf die Qualität des Aufbaus beziehen, z.B.

- ob am Beginn des Textes deutlich wird, worum es im Text gehen soll, indem z.B. die Wahl des Klassenfahrtziels genannt wird (Variable 32),
- ob ein satzübergreifendes Konzept hinsichtlich des Textaufbaus erkennbar ist (Variable 33),
- ob innerhalb des Textes alle Teile in einem erkennbaren Zusammenhang zur Argumentation stehen (Variable 34),
- ob der Schluss des Textes explizit signalisiert wird, z.B. durch eine Konklusion oder einen Appell (Variable 35).

Sprachliche Kriterien wurden zudem in Hinblick auf Angemessenheit und Richtigkeit entwickelt, und zwar auf Wort-, Satz- und Textebene (Variablen 37-44).

Das vorgestellte Beurteilungsraster kann insofern als "summativ" bezeichnet werden, als es der abschliessenden Beurteilung eines Textes dient und ggf. in einer "Note" bzw. einem Punktwert mündet. Einer solchen Verrechnung gehen allerdings Entscheidungen über die Gewichtung der einzelnen Kriterien voraus. Man kann etwa einzelne Aspekte besonders herausheben oder andere (wie etwa alle deskriptiv gemeinten Kriterien; s.o.) für die Textbewertung ganz unberücksichtigt lassen.

Trotz der Ähnlichkeit mit bereits etablierten "Erwartungshorizonten" ist offenkundig, dass das Schema für den Einsatz im Unterricht bei Weitem zu aufwändig ist.

5. Zur Implementation im Unterricht

Um die Variablen zu vereinfachen, haben wir auf die empirischen Ergebnisse und Erfahrungen aus der Auswertung der Bildungsstandards-Schreibaufgaben gesetzt. So wird z.B. die vormals dreistufige Kategorie "Orthografie" einerseits auf Basis der Verteilungsdaten mit einem "erfüllt"-Kriterium von 5 % dichotom

Grenzen zwischen "erfüllt" und "nicht erfüllt" markieren oder auch für einen der Codes besonders typisch sind. So kann den Ratern die Entwicklung einer an Vorgaben orientierten Norm erleichtert werden.

gestaltet, andererseits wird aber altersgerecht differenziert zwischen Fehlern der Gross- und Kleinschreibung und solchen der Laut-Buchstaben-Beziehungen, um Lehrpersonen Hinweise auf Förderbedarfe zu geben. Die bereits erwähnte hohe Zahl deskriptiver Variablen zur Argumentation geht in einer einzelnen Variable mit einem Mengenkriterium auf. Das Kriterium "Vorwegnahme von Gegenargumenten" erwies sich auch empirisch als besonderes Qualitätsmerkmal und wurde daher unverändert übernommen. Von den vormaligen 44 Kategorien verbleiben nunmehr 16.

Allgemein		Noch nicht erreicht	erreicht	übertroffen
Schriftbild/Lesbarkeit	Der Lesefluss wird durch das Schriftbild nicht oder kaum beeinträchtigt.			
Textumfang	Zahl der Wörter (als Wort zählt, was zwischen zwei Leerstellen steht).			
Auswertbarkeit	Eine Auswertung ist möglich, da - der Text mehr als 20 Wörter umfasst, - ein durchgehender Text vorliegt (nicht nur Stichwörter), - der Text erkennbar fertiggestellt wurde, - keine blosse Abschrift des Impulstextes vorliegt.			
Sprache				
Orthografie: Gross- und Kleinschreibung (Grenze bei ca. 5 % der Wörter)	Es gibt keine oder kaum Fehler im Bereich der Gross- und Kleinschreibung (ohne Nominalisierungen).			
Orthografie: Laut-Buchstaben-Beziehungen (Grenze ca. bei 10 % der Wörter)	Es gibt kaum Fehler im Bereich der Laut-Buchstaben-Beziehungen (Ersetzungen, Auslassungen, Hinzufügung von Buchstaben).			
Kontrolle des Satzbaus	Die Sätze sind abwechslungsreich gestaltet (z.B. durch Satzgefüge, unterschiedliche Satzanfänge, Verwendung von verschiedenen Konnektoren...).			
Zeichensetzung (nur Satzgrenzenmarkierung)	Die Satzgrenzen sind (ggf. abgesehen von den Grenzen von Anrede und Grussformel/Unterschrift) durchgängig korrekt markiert. Fehlende oder falsche Kommas werden ignoriert.			
Grammatik (v.a. Verbformen, Satzgliedfolge, Kasusgebrauch, Konjunktionen) (Grenze ca. bei 5 % der Wörter)	Der Text enthält nur wenige Grammatikfehler, die sich jedoch nicht störend auf Lesefluss oder Verständnis des Textes auswirken.			
Wortschatz	Der Text weist auf einen umfangreichen Wortschatz hin (als Hinweise gelten v.a. verständlich gebrauchte Formulierungen,			

	die deutlich über das Wortmaterial im Aufgabentext und sog. Passe-Partout-Wörter hinausgehen).			
Struktur				
Aufbau	Zu Beginn des Textes (innerhalb des ersten oder zweiten Satzes) wird die Wahl des Klassenfahrtziels deutlich. Der Schluss des Textes wird explizit signalisiert, z.B. durch ein Fazit, eine Zusammenfassung etc. Hinsichtlich des Textaufbaus ist insgesamt ein satzübergreifendes Konzept erkennbar.			
Kohärenz	Der argumentative Gesamtzusammenhang des Textes ist erkennbar.			
Inhalt				
Aufgabenverständnis	Die Aufgabenstellung ist durchgängig verstanden worden. Insbesondere wird deutlich, dass die Merkmale der Reiseziele richtig verstanden wurden. Zudem wird durch eine direkte Anrede („Ihr/Wir“) und/oder Formulierungen mit explizitem Aufforderungscharakter ein direkter Adressatenbezug im Text sprachlich ausgedrückt.			
Zentraler Inhalt: eigene Position	Es wird ausdrücklich eine eigene Wahl getroffen (für Weinhausen, Landstein oder ggf. für eine eigene plausible Variante).			
Basiselemente laut Aufgabenstellung: Argumente	Es wird mindestens ein Argument für die eigene Auswahl angeführt (keine reine Wiedergabe/Beschreibung).			
Stützung/Betonung/ Hervorhebung	Die Argumentation basiert auf mindestens zwei verschiedenen der folgenden Bezüge: - Interessen der Adressaten (Schüler der 4a) - Lernmöglichkeiten - Vielfalt - Erholung/Gesundheit - Natur - Exklusivität/Einmaligkeit - Kosten - Weiteres oder Es wird der Stellenwert eines oder mehrerer Argumente hervorgehoben. Den Argumenten wird innerhalb der Argumentation damit eine unterschiedliche Bedeutung zugewiesen. oder Ein oder mehrere Argumente werden so ergänzt, dass die Vorstellungskraft des Rezipienten angeregt wird. Sie kann insbesondere durch Vergleiche oder die Nennung von Beispielen erfolgen.			
spezielles Qualitätsmerkmal	In mindestens einem Fall wird ein mögliches/denkbares Gegenargument zur eigenen Position genannt (und ansatzweise entkräftet).			

Abbildung 6: Summatives Auswertungsschema

So gefasst, kann das Beurteilungsraster prinzipiell dazu dienen, Schülerprodukte auch im Unterricht zu beurteilen.

Dort sollte es aber auch darum gehen, den Lernprozess zu begleiten und Textprodukte weiter zu überarbeiten. Wenn es also um ein "förderndes Beurteilen" geht, dann ist im Gegensatz zur Beurteilung von Testleistungen die Rückmeldung zunächst vorläufig. Man gibt dem Schreiber die Möglichkeit zum Weiterlernen und Weiterschreiben, der Überarbeitungsprozess kommt zusätzlich in den Blick.

Keineswegs trivial ist dabei die Frage, welche Textmerkmale generell als Qualitätskriterium gefasst werden können. Die allgemein formulierten Kriterien (etwa: "sinnvoller Textaufbau") sind zu übersetzen in konkrete Merkmalsbeschreibungen (etwa: "Zu Beginn des Textes (innerhalb des ersten oder zweiten Satzes) wird die Wahl des Klassenfahrtziels deutlich."). Offensichtlich hilft diese Konkretisierung dabei, kriterienorientiert zu beurteilen. Sie schliesst aber gleichzeitig Varianten aus, die rhetorisch ebenfalls sinnvoll sein können, wie etwa die Nennung des gewählten Ziels als Konklusion am Textende. Eine mögliche Lösung ist eine Formulierung wie: "*Typischerweise* zu Beginn des Textes wird die Wahl des Klassenfahrtziels deutlich." Sie reflektiert nicht nur Anforderungen des konkreten Schreibanlasses, sondern ggf. auch die Kompetenzentwicklung der jeweiligen Lerngruppe und die vorausgehenden Unterrichtsinhalte. Von allzu offenen Formulierungen ("Der Text ist sinnvoll aufgebaut.") ist abzuraten. Zwar bieten sie die Möglichkeit, viele denkbare Lösungen zu subsumieren, sie umgehen aber gleichzeitig die notwendige Explikation der Erwartungen (mithin Unterrichtsziele). Besonders originell gelungene Lösungen einzelner SchülerInnen, die durch ein Beurteilungsraster nicht erfasst werden, können vielmehr wertvoller Gegenstand weiterer Reflexion für Lehrpersonen und SchülerInnen werden. Sie schärfen den Blick für die Funktionalität sprachlicher Formen, wie wir in den Diskussionen bei Kodiererschulungen immer wieder erfahren haben.

Beurteilungen von Texten und Textentwürfen werden im Unterricht auf unterschiedliche Weise organisiert und teils von der Lehrperson ("Schreibberatungen", Randkommentare), teils von den SchülerInnen selbst vorgenommen. Ein zentrales Mittel hierfür sind "formative" Beurteilungsraster. Sie werden nicht zur abschliessenden Beurteilung an fertige (Schreib-)Produkte angelegt, sondern ermöglichen eine differenzierte, auf Förderung zielende Rückmeldung zu einem Text im Entwurfsstadium. Im Sinne eines kompetenzorientierten Unterrichts ist dabei wichtig, dass die Rückmeldung bezüglich spezifischer Textmerkmale erfolgt, die auf Schülerseite als Teilaspekte der Schreibkompetenz interpretiert werden. Solche Teilkompetenzen lassen sich auf Grundlage empirischer Studien vergleichsweise genau formulieren – die oben ausgeführten Kategorien können dafür leitend sein, müssen aber ggf. dem konkreten Schreibanlass und der Lerngruppe angepasst werden. Dies kann partiell auch durch die LernerInnen selbst erfolgen.

6. Fazit

Auswertungsschemata und sog. Kompetenzraster zur Selbst- und Fremdbeurteilung haben mit Recht in den letzten Jahren an Popularität gewonnen. Sie ermöglichen den Lehrpersonen eine transparentere und tendenziell "gerechtere" Einschätzung der Schülerleistungen, unter anderem deshalb, weil sie sie zwingen, sich über die eigenen Unterrichtsziele, Qualitätskriterien und Gewichtungen explizit Rechenschaft abzulegen. Ausserdem können sie detailliertere Leistungsprofile ihrer SchülerInnen gewinnen, die v.a. im Hinblick auf gezielte Förderung von Bedeutung sind. Auf Schülerseite gilt das Gleiche: Sie können sehen, was von ihnen erwartet wird und ihre Anstrengungen gezielt auf Bereiche richten, in denen sie (noch) Schwächen haben.

Bei aller Wertschätzung solcher Instrumente wird jedoch oft übersehen, wie komplex der Prozess ist, diese zu gewinnen. Bei der Vorstellung ihres Basiskataloges zur Bewertung von Schülertexten weisen Becker-Mrotzek & Böttcher darauf hin, dieser müsse

in zweifacher Weise an die jeweils konkreten Schreibaufgaben und Bewertungsziele angepasst werden: Zum einen müssen die spezifischen Merkmale der jeweiligen Textart eingearbeitet werden, das heißt, die eher allgemeinen Formulierungen müssen konkretisiert werden. Und zum anderen müssen die didaktischen Ziele der jeweiligen Unterrichtseinheit berücksichtigt werden. (Becker-Mrotzek & Böttcher, 2006: 94).

Ohne eine solche Anpassung stellt sich bei Anwendung auf konkrete Schülertexte sehr schnell der Eindruck ein, am "Eigentlichen" des Textes vorbei zu urteilen.⁹

Nun ist andererseits von Lehrpersonen nicht zu erwarten, dass sie ihre Beurteilungsraster in aufwändigen Voruntersuchungen erst gewinnen. Hier ergibt sich eine Transfergelegenheit wissenschaftlicher fachdidaktischer Forschung in die schulische Praxis: Gerade dort, wo sich Sets von Auswertungskriterien in grösseren Studien bewährt haben, können aus ihnen für Lehrpersonen handhabbare Raster entstehen. Werden diese *im Verbund mit der zugehörigen Schreibaufgabe* publiziert, können Lehrerinnen und Lehrer sowohl Aspekte der Aufgaben als auch die entsprechenden Auswertungskriterien für die Bedürfnisse ihrer Lerngruppen anpassen. Diese Anpassungsarbeit jedoch kann ihnen nicht zentral abgenommen werden – sie ist Teil der professionellen Anforderungen an die Entwicklung eines kompetenzorientierten Unterrichts.

⁹ Dies gilt u.E. auch für Kompetenzraster, Selbstevaluationen, Checklisten etc., die sich in aktuellen Lehrwerken und -materialien zum Sprachunterricht immer häufiger und in teils hoher Qualität finden; vgl. dazu in Deutschland etwa "Deutschprofi" (Oldenbourg), "Deutsch.punkt" (Klett) oder das Schweizer Lehrwerk "Die Sprachstarken" (Klett und Balmer).

Literatur

- Baurmann, J. (1995): Schreiben in der Schule. Orientierung an Schreibprozessen. In: J. Baurmann & R. Weingarten (Hg.), Schreiben – Prozesse, Prozeduren und Produkte. Opladen (Westdeutscher Verlag), 51-69.
- Becker-Mrotzek, M. & Böttcher, I. (2006): Schreibkompetenz entwickeln. Berlin (Cornelsen).
- Birkel, P. & Birkel, C. (2002): Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? In: Psychologie in Erziehung und Unterricht, 49, 219-24.
- Birkel, P. (2009): Rechtschreibung im Diktat: Eine objektiv beurteilbare Leistung? In: Didaktik Deutsch, 27, S. 5-32.
- Böhme, K., Bremerich-Vos, A. & Robitzsch, A. (2009): Aspekte der Kodierung von Schreibaufgaben. In: D. Granzer, O. Köller & A. Bremerich-Vos (Hg.), Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule. Weinheim (Beltz), 290-329.
- Briggs, D. (1970): The influence of handwriting on assessment. In: Educational research, 13, 1, 50-55.
- Fix, M. (2006): Texte schreiben. Schreibprozesse im Deutschunterricht. Paderborn (Schöningh).
- Hartig, J. & Klieme, E. (2007): Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In: Zeitschrift für Erziehungswissenschaft, Sonderheft 8, 11-29.
- Ingenkamp, K.-H. (1971): Die Fragwürdigkeit der Zensurengebung. Weinheim (Beltz).
- KMK (Kultusministerkonferenz, Beschlüsse der) (2005): Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4) – Beschluss vom 15.10.2004. München/Neuwied (Luchterhand).
- Schröter, G. (1971): Die ungerechte Aufsatzzensur. Bochum (Kamp).
- Wirtz, M. & Caspar, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen (Hogrefe).