DIES-Sommerschule: Empirische Schreibforschung

ZUR ANALYSE VON LERNERTEXTEN MITTELS RATINGVERFAHREN

Dr. Ulrike Behrens, Michael Krelle

Köln, 25.6.09

Programm für heute

- Einführung: Probleme der Bewertung von Schülertexten
- SCHREIBEN im Projekt „Evaluation der Bildungsstandards“
- Was sind „Ratingverfahren“?
- Zur Anlage von Kodieranweisungen (Aufgabe „Leserbrief“)
- Gruppenarbeit: Entwicklung eigener Kodieranweisungen anhand von Schülertexten (Aufgabe: „Eine Fabel“)

Kaffee- (und Abschreib-)pause

- Sammlung: Weitere mögliche Auswertungskriterien
- Einzelarbeit: Kodierung der Texte
- Auswertung der Übereinstimmungen und Ergebnisdiskussion

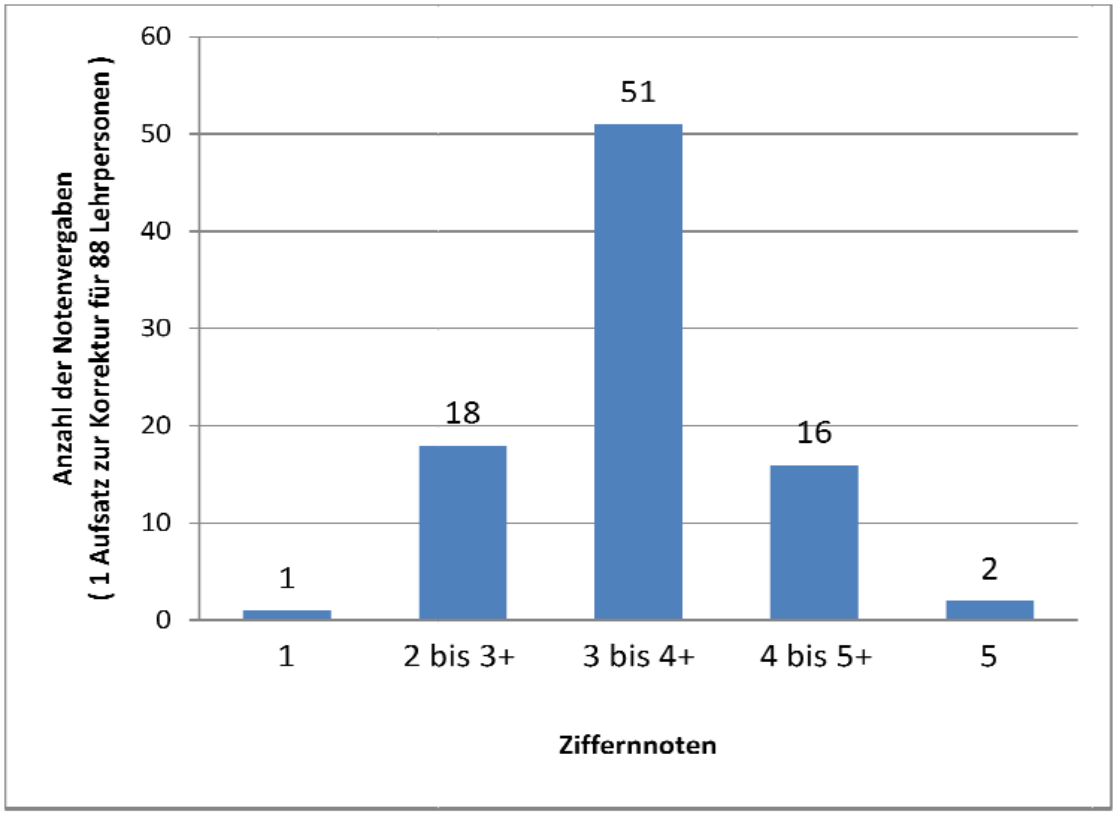
Problem der Bewertung von Texten:
„Die ungerechte Aufsatzzensur“ (Schröter 1970)

- Holistische Urteile können stark streuen.
- Resultat am Beispiel eines Aufsatzes in Form von Ziffernnoten:

1: 1 mal,
2 bis 3+: 18 mal
3 bis 4+: 51 mal
4 bis 5+: 16 mal
5: 2 mal

(= Normalverteilung)

vgl. Birkel/Birkel 2002



Problem der Bewertung von Texten: Auswertungsobjektivität

Verminderte Auswertungsobjektivität durch …

- Vorinformation: Rosenthal-Effekt (z. B. Weiss 1995)
- Beliebtheit der Schüler/innen (Hadley 1954, 1995)
- Länge der Texte (z. B. Birkel/Birkel 2002)
- grammatikalische und orthografische Fehler: Halo-Effekt (z. B. Birkel/Birkel 2002)
- Handschrift (z. B. Briggs 1970)
- Reihenfolgeeffekte (z. B. Baurmann 1995)

Problem der Bewertung von Texten: Teilkompetenzen

- Unklarheit über das Verhältnis verschiedener Teilkompetenzen untereinander

→ Wie sind hohe Korrelationen zwischen verschiedenen Teilkompetenzen zu interpretieren?

→ Welche Teilqualitäten sollen wie gewichtet werden?

→ Keine „empirischen Fragen“

IQB-Schreibprojekt

- Hintergrund: Normierung der Bildungsstandards im Fach Deutsch, hier: Schreibkompetenz

„Die Schülerinnen und Schüler kennen die vielfältigen Möglichkeiten des Schreibens als Mittel der Kommunikation, der Darstellung und der Reflexion und <u>verfassen selbst adressatengerecht Texte</u>." (KMK Bildungsstandards)

Umsetzung und Beherrschung verschiedener
- Textsorten/ -muster,
- kommunikativer Zielsetzungen,
- sprachlicher Mittel.

Einige Hinweise zur Aufgabenkonstruktion

- Aufgabe = Schreibanlass für einen längeren, frei formulierten Text

Entwicklungskriterien:

- Möglichst motivierende, ökologisch valide Aufgabenstellung,
- verbale und grafische Hinweise zur erwarteten Textlänge,
- Klärung der AdressatInnen des Textes,
- ggf. kurze Lesetexte integriert (z. B. Leserbrief),
- gelegentlich Teilaufgaben (mit Blick auf Auswertung),
- Berücksichtigung verschiedener „Textsorten“.

Aufgaben und Stichproben (Sek 1)

- Pilotierung: insgesamt 24 Schreibaufgaben
  Stichprobe: 2150 Schülerinnen und Schüler
- Normierung: 11 Schreibaufgaben
  - 3x argumentieren
  - 4x informieren
  - 2x erzählen
  - 1x überarbeiten
  - 1x zusammenfassen

  zwischen 200 und 500 SchülerInnen pro Aufgabe

→ sehr großer Korpus unterschiedlichster Schülertexte

Verschärfte Qualitätsanforderungen im Large Scale-Assessment

- für geschulte KodiererInnen verständliche Auswertungsanweisungen
- hinreichende Übereinstimmung der Kodiererurteile
- Maß für „hinreichend“ ist strittig: Faustformeln, aber keine strikte Grenzziehung
- Abhängigkeit von Charakteristika der Variablen:
  - hoch- vs. niedrig inferent
  - zwei- vs. mehrstufig
  - ein- vs. mehrdeutig

Aber: Das alles gilt grundsätzlich auch für kleine Korpora.

Ratingverfahren

- „Rating“ = Urteile und Wertungen von bestimmten Textaspekten
- Ziel im besten Fall:
  Eine zusammenfassende Bewertung (für einen Schülertext)

  hier: Schreibprodukte!

Der Weg dahin:

- **Benennung** einer konkreten Zahl von **Teilqualitäten**
- **Bewertung** jeder **einzelnen Qualität *für sich***
- **Gewichtung** der Teilqualitäten
- **Verrechnung** zu einer Gesamtbewertung

Zur Anlage von Kodieranweisungen

Ratings basieren auf **Kategorien(systemen)**

Konkret: „Kodieranweisungen“ (auch: „Codebook“, „Manual“ etc.)

- *Grundlage von Ratingprozessen*
- *folgen einer bestimmten „Logik“*
- *lassen sich anhand diverser Kriterien systematisieren.*

Im Folgenden : *„Zur Anlage von Kodieranweisungen“*

1.) „Verwendungssituation, Aufgabenspezifik, funktionale Einbettung“:

Aufgabenstellung, Stimulus, Textmuster und -funktion

2.) Zu Form- und Formataspekten und „Logik“

- *„Kodierstrategie“, „typische“ Kategorien/systeme,*
- *Anzahl der Codes etc.*

1.) „Verwendungssituation, Aufgabenspezifik, funktionale Einbettung“. Ein Beispiel: „Leserbrief“

primäre Textfunktion nach Schreibziel und Modalität der Themenentfaltung (vgl. u.a. Fix 2006)		Nr.
argumentieren (inkl. analysieren, interpretieren)	*Schreibziel:* Ein Problem, ein Sachverhalt wird beleuchtet, eine eigene Stellungnahme begründet ggf. der Leser von einer These überzeugt. *Modalität der Themenentfaltung:* Problem- bzw. objektbezogen, adressatenorientiert, ggf. persuasiv, argumentativ mit deskriptiven Anteilen.	1
		2
		3
		4 (V)
		5 (V)
informieren	**Berichten** *Schreibziel:* Vom Resultat her wird über ein abgeschlossenes Ereignis in der Vergangenheit informiert.	6
	Modalität der Themenentfaltung: sachlich, informierend, referierend, ereignisbezogen, deskriptiv mit narrativen Anteilen.	7
	Beschreiben *Schreibziel:* Ein Sachverhalt, ein Objekt oder ein Prozess wird zur Widererkennung oder zum Nachvollzug für einen Adressaten möglichst genau dargestellt.	8
	Modalität der Themenentfaltung: sachlich, informierend, anschaulich, objekt- oder prozessbezogen, deskriptiv mit ggf. explikativen Anteilen.	9
erzählen **(inkl. anderer kreativer Schreibformen)**	*Schreibziel:* Ein erlebtes, rezipiertes oder fiktionales Geschehen mit überraschender Wende soll unterhaltend, betroffen machend, Erstaunen erzeugend, spannend etc. nahe gebracht werden.	10
	Modalität der Themenentfaltung: subjektbezogen, emotional, szenisch, perspektivisch, stimmungsvoll, narrativ	11
Sonderformen in den KMK-Bildungsstandards	**Überarbeiten**	12
	Zusammenfassen	13

(V): Validierungsaufgabe aus Fremdprojekten

Beispielaufgabe: Leserbrief

Nussbaumer/Sieber 1994: ZTaR zur Textqualität (+30 Variablen)	Becker-Mrotzek/Böttcher 2006 Basiskatalog:12 Variablen	Evaluation der Bidlungsstandards 11 (bzw. 13) Aufgaben, (ca.20-30 Variablen, Ziel: -10)	
0. Grundgrößen (u.a. Textlänge/Wortschatz/ Syntax/ thematische Komplexität/ Kohäsions-leistung	**1. Sprachrichtigkeit** (u.a. Orthografie, Grammatik)	**1. Allg. Variablen** - „Grundgrößen“ - Allg. Einschätzung zur Angemessenheit und Richtigkeit	**Textsorten-übergreifend**
A. Sprachrichtigkeit (u.a. in Bezug auf Semantik, Interpunktion, Orthografie, Morphologie	**2. Sprachliche Angemessenheit** (u.a. Wortschatz)	**2. Sprachliche Variablen,** u.a. zur Grammatik, Orthografie, Zeichensetzung etc.	
B1. Angemessenheit: Verständlichkeit/Kohärenz (u.a. sprachliche Mittel, Rezipienten-führung, Textmusternormen, thematische Entfaltung, Gesamtidee etc.	**3. Inhalt** (u.a. Gesamtidee, Umfang, Relevanz)	**3. Strukturbezogene Variablen („Aufbau“),** u.a. zur Textsorte bzw. -muster, u Textteilen , Absätzen etc.	**Textsorten- und Aufgabenspezifisch**
B2. Ästhetische Angemessenheit: formale Qualitäten (u.a. Attraktivität der sprachlichen Mittel, Wortwahl/ Registerwahl	**4. Aufbau** (u.a. Textmuster, Textaufbau, thematische Entfaltung)	**4. Inhaltsbezogene Variablen** (u.a. zur Angemessenheit und Richtigkeit in Bezug auf den Stimulus und die Aufgabenstellung	
B3. inhaltliche Relevanz: inhaltliche Qualitäten (u.a. inhaltliche Wagnis)	**4. Schreibprozess** (u.a. Planung/Überarbeitung, Wagnis, Kreativität)		

Nummer und Kurzlabel

Codes und -Beschreibung

Benchmarks, Beispiele, Schülertexte

G057018 Einleitung	Code 1: Der Text beginnt ohne explizite Bezugnahme auf den Leserbrief der Lina K. Code 2: In einem Einleitungssatz wird der Bezug des Textes (auf den Leserbrief der Lina K.) ausdrücklich thematisiert. Code 0: Die Variable kann nicht kodiert werden, weil der Text annähernd eine vollständige Abschrift des Stimulus ist (Code 0 in Variable 002).	Code 1: Heute um 13 Uhr stand ich wie üblich an der Bushaltestelle mit meinem Einkauf an der Bushaltestelle. [...] [5601010502] Nach meiner meinung liegt es [???] Teils an der Jugend doch die Busse sind auch stark überfuhlt und den Busfahrer interessiert das nicht [...] [24400709211] Code 2: Ich möchte mit diesem Leserbrief meiner Vorrednerin Lina K. völlig recht geben, in dem Punkt das die heutige Jugend respektlos ist und keine Manieren mehr besitzt. [...] [2106010124] Ich würde mich gerne zum Leserbrief von Lina K. äußern. [...] [420260405] Ich schreibe Ihnen, da ich vor kurzem einen Leserbrief las, indem sich jemand über eine Busfahrt bei Ihnen beschwerte. [...] [5601010619] Ich schreibe diesen Leserbrief weil ich doch gerne Stellung nehmen möchte zu Lina K`s Vorwurf. [...] [20602210108]

a.) „Kodierhandbuch“ zu jeder Aufgabe und jeder Variable.
b.) „Präpilotierung“, „Pilotierung“, Studie
c.) Mehrere Schulungen mit Beispielmaterial

Allgemeine Variablen: Beispiel

Gesamt-eindruck

G057002 – (Cohens Kappa = .68)
Zu bewerten ist der Gesamteindruck bezogen auf die gestellte Aufgabe auf folgender Skala:

Codes und Codebeschreibung:
Code 1: Sehr schlechter Text.
Code 2: Eher schlechter Text.
Code 3: Mittelmäßiger Text.
Code 4: Eher guter Text.
Code 5: Sehr guter Text.
Code 0: Der Text ist annähernd eine vollständige Abschrift des Stimulus.
In diesem Fall werden alle folgenden Variablen mit 0 kodiert.

Anmerkung:
Es soll vor allem berücksichtigt werden:
- der Grad der Abgeschlossenheit bzw. Gegliedertheit der Texte;
- der Grad der Beachtung von Details, die in der Aufgabe enthalten sind bzw. kreativ hinzugefügt wurden;
- die Kontrolle des Satzbaus, der Umfang und die Angemessenheit des Wortschatzes, die Orthografie und das Schriftbild.
  Empfohlen wird, den „inhaltlichen“ Aspekten Vorrang zu geben

Erwartung: „Gleichbehandlung der Aspekte“
Aber: Ratereffekte (Strenge/Milde), Texteffekte (Merkmale stechen heraus) etc.

Sprachliche Variablen: Beispiel

Beispiele

G057010 Markierung von Satzgrenzen (Zeichensetzung I)	<u>Code 1:</u> Die Grenzen von Sätzen sind häufig nicht oder falsch (durch Punkte, Fragezeichen, Ausrufezeichen, Großschreibung am Satzanfang) markiert. <u>Code 2:</u> Die Grenzen von Sätzen sind manchmal nicht oder falsch (durch Punkte, Fragezeichen, Ausrufezeichen, Großschreibung am Satzanfang) markiert. <u>Code 3:</u> Die Grenzen von Sätzen sind durchgängig richtig (durch Punkte, Fragezeichen, Ausrufezeichen, Großschreibung am Satzanfang) markiert. <u>Code 0:</u> Die Variable kann nicht kodiert werden, weil der Text annähernd eine vollständige Abschrift des Stimulus ist (Code 0 in Variable 002). <u>Anmerkung:</u> Ein fehlendes Satzzeichen zwischen zwei Hauptsätzen oder Satzgefügen wird nur dann als fehlendes Satzschlusszeichen gewertet, wenn eine Satzgrenze durch Großschreibung markiert ist. Wenn nach einem Punkt also klein weitergeschrieben wird, ist dies als Rechtschreibfehler und nicht als Satzgrenzenfehler zu werten!

Inhaltsbezogene Variablen: Beispiel

Hauptthese

Code 1:

Auf die Hauptthese Lina K.s (sinngemäß: „Verrohung der Jugend“) wird nicht Bezug genommen.

Code 2:

Die Hauptthese der Lina K. wird zwar erwähnt, sie steht jedoch nicht im Zentrum der Argumentation.

Code 3:

Die Argumentation zielt (explizit oder implizit) **im Wesentlichen** abstrakt auf die Generalisierbarkeit der Hauptthese Lina K.s ab, z.B.

- bei ablehnender oder vermittelnder Position durch **Abstreiten der Generalisierbarkeit der Aussage**, ggf. mit Anführen von Gegenbeispielen oder Erklärungen für die Verhaltensweisen,
- bei zustimmender Position durch **Bestätigung und Anführen weiterer Beispiele.**

Code 0:

Die Variable kann nicht kodiert werden, weil der Text annähernd eine vollständige Abschrift des Stimulus ist (Code 0 in Variable 002).

Einige Schlussfolgerungen:

1. Kategorien und Kodiersysteme sind (neben der konkreten Forschungsfrage) wesentlich *an die Aufgabenspezifik (Aufgabenstellung, Stimulus, Erhebungssituation etc.) gebunden.*
2. Darüber hinaus sind Aspekte wie *„Kodierstrategie“, „typische“ Kategorien(-systeme), Anzahl der Codes etc. bedenkenswert.*
3. *Das Anlegen von Codes ist immer ein bewusster Akt zu Gunsten oder zu Ungunsten von Textaspekten.*
4. *Durch Ratings mit mehreren Menschen werden Kategorien bereits im Entwicklungsprozess „objektiviert“ und auf ihre empirische Güte geprüft.*
5. *Mehrfachrating eignet sich auch für kleine Stichproben.*

Einige Anmerkungen

- Dimensionen (erste Analysen GS!):
  Sprache = eine Dimension (Orthografie?)
- Holistische- vs. analytische Urteile: r= 0.8 bis 0.9 (GS)
- Ziel: Differenzierung von Standards
  Was wird ausgeblendet? „Überarbeitungen“, „Stil“, u.a.

Workshop

Entwicklung eigener Kodieranweisungen zur Beispielaufgabe „Fabel“

Arbeitsphase in kleineren Gruppen

- Wie lassen sich Aspekte der Textqualität bezogen auf die vorliegende Aufgabe beschreiben?
- Bitte wählen Sie **einen** Aspekt aus, z. B.:

Argumentative Qualität	Zeichensetzung
Stil/Ausdruck	Wortschatz
Inhaltliche Richtigkeit	Satzbau
Grammatik	…

- Notieren Sie ggf. Teilaspekte zu diesem Bereich
- Beschreiben Sie verschiedene Qualitätsniveaus von (sehr) gelungen bis (vollkommen) misslungen

Arbeitsergebnisse

Kodieranweisungen für fünf Beispielvariablen (Ergebnisse der Grupenarbeit)

Qualitäts-aspekt	Teilaspekt (Variable)	Qualitätsniveaus (Codes)
Inhalt	positive Wendung	0 = keine 1 = Überleben der Stiere
	Anschluss	0 = schließt nicht an 1 = schließt an
Wortschatz	Differenziert-heit (bezogen auf den Umfang)	??? Beschreibung von Einzelniveaus macht Probleme versuchsweise: 0 = nicht differenziert 1 = differenziert

Qualitäts-aspekt	Teilaspekt	Qualitätsniveaus (Kodierung)
Textzusam-menhang	Kohäsion	1 = nicht angemessen 2 = 3 = 4 = angemessen
	Kohärenz	1 = nicht angemessen 2 = 3 = 4 = angemessen
	Grad der Abgeschlos-senheit	1 = nicht angemessen 2 = 3 = 4 = angemessen
Satzbau/ syntaktische Komplexität (stark verkürzt gegenüber Gruppenerg.)	Kriterien: Nebensätze, Hauptsätze, Ellipsen, Einbettungs-tiefe	1 = keine Komplexität 2 = geringe Komplexität: außer HS auch HS+NS 3 = gute Komplexität HS, HS+NS, …

Kodier-Übereinstimmungen

…bei vier Beispieltexten

Kodierentscheidungen TN (Auszug)

	positive Wendung		Wortschatz		Satzbau		
Code…	1	2	1	2	1	2	3
Text 1	1	26	4	16	3	14	1
Text 2	3	22	2	14	0	9	7
Text 3	3	13	3	14	1	10	6
Text 4	1	22	9	8	4	7	5

Fazit

- Die Entwicklung von Kodieranweisungen gestaltet sich gerade dann schwierig, wenn mehrere damit zuverlässig arbeiten sollen. Dieses ist aber im Sinne der empirischen Güte wünschenswert.
- Viele Fragestellungen sind nur vermeintlich klar zu beantworten (vgl. die inhaltliche Variable „positive Wendung“, aber auch Sprachbezogenes wie „Satzbau“).
- Bezüglich der Wahl der „Kodierstrategien“ (etwa: holistisch oder analytisch) ergeben sich weitere Schwierigkeiten, z.B.:
  - „Tendenz zur Mitte“ (evtl. vierstufige Variablen verwenden?)
  - Entscheidung zwischen „zählen“ (geht nur bei wenigen Aspekten, z. B. „Fehlern“) und/oder „schätzen“ (Übereinstimmung erfordert besonders klare Kriterien).
  - „dichtom“ vs. „mehrstufig“ (bei dichotomen Codes oft höhere Raterübereinstimmung).
- Mit Ratings durch mehrere KodiererInnen werden Kategorien bereits im Entwicklungsprozess auf ihre empirische Güte geprüft – gerade auch bei kleineren Stichproben!

Herzlichen Dank

…an alle TeilnehmerInnen für Ihre Mitarbeit und konstruktiven Beiträge.

Kontakt

ulrike.behrens@uni-due.de

michael.krelle@uni-due.de

Universität Duisburg/Essen