



Klassenarbeiten bewerten und beurteilen

Schwerpunkt: Möglichkeiten, aber auch Sinnlosigkeiten von Klassenarbeiten und Leistungsbeurteilungen

Aus: BEHINDERTENPÄDAGOGIK, 39. Jg., Heft 3/2000, Seite 312-322

Manfred Bönsch

[Das Problem](#)

[Die häufigsten Fehler](#)

[Sinnlosigkeiten](#)

[Die positive Wende: Vertretbare Klassenarbeiten und die damit verbundenen Anforderungen](#)

[Die Rahmenbestimmungen der Leistungsbewertung mit Noten: Gewichtung und Verortung](#)

[Informelle Tests - die bessere Variante?](#)

[Das schwierigste Feld: Aufsätze, freie Texte](#)

[Schluss](#)

[Literatur](#)

[Anschrift des Verfassers:](#)

© Manfred Bönsch 2000



Das Problem

Zum scheinbar oder tatsächlich unaufgebbaren Bestandteil des Schulalltags gehören Klassenarbeiten. Ihre Zahl ist in Erlassen vorgeschrieben und so tauchen sie mit schöner Regelmäßigkeit auf. Von Schüler/innenbefragungen wissen wir, dass die Modi der Leistungsbewertung bedrängend sind, dass sie den größten Leidensdruck verursachen. Und Schulpädagogen wissen, dass die Tücken beträchtlich sind auf dem Wege zu Gerechtigkeit, Objektivität und Nachvollziehbarkeit. Das Thema ist insgesamt ein unerledigtes Kapitel schulpädagogischen Denkens und Handelns.



Die häufigsten Fehler

Nur kurz soll daran erinnert werden, was wir recht sicher über Fehler und Tücken wissen. Die Unklarheiten fangen schon bei der Identifizierung des Bezugsrahmens an, in dem Klassenarbeitsbeurteilungen stehen: Handelt es sich um einen objektivierten, weitrahmigen Bezugsrahmen (alle Schüler/-innen der 6. Schuljahre z.B. Hessens schreiben etwa zur gleichen Zeit die gleichen Mathematikarbeiten) oder um einen klasseninternen Bezugsrahmen (die Mathematikarbeit in Klasse 6a unterscheidet sich schon von der in 6b in Anspruch, Umfang und Ansetzung und ist mit der für eine 6. Klasse in einer anderen Schule gar nicht mehr vergleichbar) oder um einen intraindividuellen Bezugsrahmen (es werden nur die Lernfortschritte eines Sechstklässlers von einer Leistungskontrolle zur anderen bewertet). Über die Fächer hinweg gibt es fachspezifische Hierarchien: Die Leistungsfächer (Deutsch, Mathematik, Fremdsprachen, Naturwissenschaften) erfahren die strengste Zensurierung, die sog. Beteiligungsfächer (z.B. Geographie, Geschichte) eine weniger strenge, die sog. Gestaltungsfächer (Kunst, Sport, Musik) die mildeste. Geschlechtsspezifische und schichtenspezifische Zensurierungsunterschiede bestehen nach wie vor. Und das Beurteilungssystem "Lehrer/-in" ist nicht valide, objektiv und reliabel! Bekannt sind Befunde, nach denen verschiedene Lehrer/-innen den gleichen Arbeiten unterschiedliche Noten geben (in der Streuung von sehr gut bis mangelhaft), dass sogar die gleichen Lehrer/-innen zu verschiedenen Zeiten der gleichen Arbeit unterschiedliche Noten geben. Bekannt sind bestimmte Fehlerarbeiten. Da gibt es den Strenge-/Mildeeffekt (ein(e) Lehrer/-in gibt Zensuren vorwiegend im oberen oder unteren Notenbereich), da gibt es die Tendenz zur Mitte als Scheu, Extremwerte zu benutzen. Die Orientierung an der Gaußschen Normalverteilung ist ausgeprägt vorhanden (je Lerngruppe/Klasse hat ein breites Mittelfeld und jeweils wenige im oberen und unteren Bereich). Der sog. Reihungseffekt führt dazu, dass vorausgegangene Urteile/Eindrücke die nachfolgende Leistungsbewertung bestimmen (nach drei guten Arbeiten kann es nicht eine vierte auch noch geben). Der Halo-Effekt lässt allgemeine Eindrücke von einer Person (Kleidung, Haartracht, allgemeines Verhalten) in die Leistungsfeststellung und -beurteilung einfließen. Der Ermüdungseffekt im Lauf der Durchsicht einer Klassenarbeit lässt flüchtiger,

milder oder strenger werden.

Die bisher entwickelte Problemliste erfährt eine Verstärkung durch eine häufig bestehende Unklarheit über die Funktion von Klassenarbeiten. Haben sie "nur" eine Orientierungs-/Berichts-/Rückmeldefunktion im Sinne von Lernerfolgskontrollen (können wir das jetzt oder doch noch nicht?) oder haben sie auch eine sog. pädagogische Funktion im Sinne der Belohnung/Bestrafung/Disziplinierung von Schülern/innen (Ansporn, Motivation, Warnung, "Racheabsichten" können schnell den subjektiven Faktor aufladen) oder steht die Berechtigungsfunktion im Vordergrund (Arbeiten sind die wichtigste Grundlage für die Vergabe von Zensuren und Abschlüssen)? Der prognostische Wert von Zensuren (das Abitur sage ein gutes Medizinstudium voraus, eine sehr gute Lehramtsprüfung bestätigt einem Kandidaten, dass er ein sehr guter Lehrer sein wird) ist offensichtlich sehr begrenzt. Und wenn es zu einer Mischung der Funktionen kommt, was meistens der Fall ist, wird es vollends problematisch. Dies heißt in der Summe, dass die Diffusität der Leistungsbeurteilung außerordentlich hoch ist und die Konturen des Verantwortbaren verschwimmen.

Sind dies die spätestens seit den INGENKAMPschen Publikationen bekannten Fehler, die auftreten können, so sind unter didaktischen Aspekten, die Problematik zuspitzend, folgende Sinnlosigkeiten zu markieren.



Sinnlosigkeiten

Eine Klassenarbeit ist relativ sinnlos, wenn sie nicht primär als Lernerfolgskontrolle gilt und nicht direkt Folge wirklich durchgeführten Unterrichts ist. Sollte es den Fall geben, dass eine Klassenarbeit konzipiert und geschrieben wird, weil halt wieder eine "dran" ist und sie in keinem Zusammenhang mit realisierten Lernprozessen steht, ist sie als sinnlos zu bezeichnen.

Eine Klassenarbeit ist relativ sinnlos, wenn ihr Erwartungshorizont nicht vorher expliziert worden ist, wenn also die Vorstellung von dem, was im besten Fall erreicht werden kann, nicht klar ist und damit eine "Beurteilungsfolie" nicht gegeben ist. Es kann sich dann nur um eine willkürlich addierte Zahl von Aufgaben handeln.

Eine Klassenarbeit ist relativ sinnlos, wenn der Bewertungs- und Beurteilungsmodus nicht vorher festgelegt worden ist. Es treten dann regelmäßig die Fehler auf, die oben kurz genannt worden sind.

Eine Klassenarbeit ist relativ sinnlos, wenn Schüler/innen nicht mitgeteilt bekommen, was sie nicht vollständig oder fehlerhaft bearbeitet haben. Das Anstreichen von Fehlern und die Note unter der Arbeit sind an Information zu wenig, um eine Klassenarbeit als Lernerfolgskontrolle und Lernmöglichkeit zu verstehen.

Eine Klassenarbeit muss auch in ihrer Relation zu anderen Leistungen näher bestimmt sein. Welchen Anteil hat sie im Rahmen der schriftlichen Arbeiten und wie sind Klassenarbeiten in ihrem Verhältnis zu mündlichen Leistungen und weiteren fachspezifischen Leistungen gewichtet? Haben sie unter Umständen wechselnde Bedeutungen (werden also einzelne Arbeiten plötzlich stärker gewichtet, als vorher angegeben worden war), wachsen Irritationen

für die Schüler/-innen, wird Leistungsbeurteilung zu einem Lotteriespiel?

Wird nicht deutlich zwischen Leistungsfeststellung und -beurteilung unterschieden, kann schnell das Phänomen auftreten, dass ein 'Befriedigend' z.B. nicht nachvollziehbar ist. Dann kommt der Eindruck fehlender Plausibilität auf, Sinn ist nicht zu finden. Nach wie vor ist eine umstrittene Frage, ob im Prinzip nicht die inhaltliche Lernerfolgskontrolle (Rückmeldung 'erreicht - nicht erreicht') reichen würde und warum immer eine Bewertung durch Umsetzung in eine Note notwendig ist. Aber hier spielen die Erwartungen von Eltern eine große Rolle. Ein 'Befriedigend' sagt offenbar mehr als eine Fehlerrückmeldung bzw. eine Punktangabe (von 20 erreichbaren Punkten sind 14 erreicht worden).



Die positive Wende: Vertretbare Klassenarbeiten und die damit verbundenen Anforderungen

Die Dramaturgie der Darstellung ging absichtlich zunächst von der Problem- und Fehlerhaftigkeit von Klassenarbeiten und generell der Leistungsfeststellung und -beurteilung aus. Es ist jetzt aber die Stelle, an der die positive Wende genommen werden soll. Sinn und Soll sind in folgender Weise zu bestimmen:

1. Schriftliche Klassenarbeiten machen in einer ganzen Reihe von Fächern (Deutsch, Mathematik, Fremdsprache, sie sog. Sachunterrichtsfächer vor allem) Sinn, wenn sie als zwischenzeitliche Lernerfolgskontrollen verstanden werden.
2. Dazu ist es notwendig, dass sie sehr eng an den gehaltenen Unterricht anschließen. Klassenarbeiten sollten auch unter dem Aspekt der Erfolgsprüfung des Unterrichts(!) gesehen werden. Sie sind nicht nur Überprüfungen von Schülern/-innen.
3. Da Klassenarbeiten den Testgütekriterien Validität, Objektivität und Reliabilität in den allermeisten Fällen nicht entsprechen können, ist Transparenz das vorherrschende Kriterium. Transparenz bei Klassenarbeiten heißt:
 - Die Ankündigung einer Klassenarbeit weist auf die Unterrichtseinheiten, ihre Inhalte und Anforderungen hin, die bearbeitet worden sind und die in Schulbüchern, Unterrichtsergebnissen (Hefte, Mappen u.a.m.) ihren Niederschlag gefunden haben. Schüler/innen müssen sich in verlässlichem Rahmen vorbereiten können. (Um was wird es gehen?)
 - Die Anforderungen werden im Inhalt und Leistungsniveau klar beschrieben. (Was wird verlangt werden?)
 - Der Auswertungsmodus wird im vornhinein festgelegt und bekanntgemacht (Fehlerfeststellung und Gewichtung von Leistungen und ihre Beurteilung).



Die Rahmenbestimmungen der Leistungsbewertung mit Noten: Gewichtung und Verortung

Wenn z.B. in der Grundschule Diktate mit Noten bewertet werden sollen, ist zunächst die Verortung im Fach wichtig. Diese wird hier einmal exemplarisch vorgenommen. Natürlich ist die Frage bedeutsam, ob Noten überhaupt zweckmäßig sind. Darauf wird später noch eingegangen.

Im Fach Deutsch gibt es verschiedene Lernbereiche. Im 4. Schuljahr könnte man sie folgendermaßen benennen und gewichten:

Sprechen, Hören	0,6 Std.	(15% der 5 Wochenstunden)
Lesen	0,5 Std.	(10% der 5 Wochenstunden)
Rechtschreiben	1,3 Std.	(25% der 5 Wochenstunden)
Texte erfassen	1,3 Std.	(25% der 5 Wochenstunden)
Sprache betrachten und untersuchen	1,3 Std.	(25% der 5 Wochenstunden)
Gesamt	5 Std.	100%

Damit ist im Grunde auch die Gewichtung der Lernbereiche festgestellt. Die Note für das Fach Deutsch wird dann folgendermaßen ermittelt:

- Festlegen der Teilnoten in jedem Teilbereich.
- Multiplizieren dieser Teilnoten mit der zugehörigen Prozentzahl; diese wäre für die Rechtschreibung $25\% = 0,25$.
- Die Teilnoten werden mit der Prozentzahl multipliziert und ergeben Teilwerte, die addiert die Gesamtnote ergeben. Ein Beispiel:

<i>Lernbereich</i>	<i>Teilnote</i>	<i>Prozentzahl</i>	<i>Teilwert</i>
Sprechen, Hören	2 x	0,15 =	0,30
Lesen	3 x	0,10 =	0,30
Rechtschreiben	3 x	0,25 =	0,75
Texte verfassen	2 x	0,25 =	0,50
Sprache	4 x	0,25 =	1,00
		gesamt	2,85
		Zensur also	3

Mit dieser Gewichtung generell und der Verortung des Rechtschreibens ist erreicht, dass (1.)

Leistungen in unterschiedlichen Lernbereichen entsprechend gewürdigt werden und (2.) der Stellenwert des Rechtschreibens festgelegt ist und nicht übermächtig oder auch unterbewertet werden kann. Die Gewichtungen mögen als willkürlich aufgefasst werden. Sie geben aber Verlässlichkeit über verschiedene Klassen hinweg (in einer Schule oder gar in mehreren/vielen Schulen). Wenn die Note im Fach Deutsch auch im Zeugnis so aufgeschlüsselt werden würde, wäre dem Postulat der Transparenz auf eine erste Weise Genüge getan.

Im Bereich des Rechtschreibens geht es dann um die Feststellung der Zahl der jährlich mit Benotung zu schreibenden Diktate, um die Festlegung des Umfangs (z.B. 6 Diktate mit dreimal 80 und dreimal 90 Wörtern). Immer willkürlich und schwer begründbar ist die Zuordnung der Notenskala zu Fehlerzahlen, also z.B.

0	Fehler = sehr gut
1-2	Fehler = gut
3-5	Fehler = befriedigend
6-8	Fehler = ausreichend
9-12	Fehler = mangelhaft
mehr	Fehler = ungenügend

Je nach vorausgegangenem Rechtschreibunterricht (Diktat als Lernerfolgskontrolle) könnte dies als zu hart oder zu weich bezeichnet werden. Das Individuum und seine vielleicht immensen Lernfortschritte seit dem letzten Diktat (z.B. von 18 auf 13 Fehler verbessert) können keine Würdigung mehr erfahren. Die Noten sind für beide Diktate ungenügend. Eine lernprozessorientierte Würdigung der Verbesserungen und damit eine pädagogisch gehaltene Rückmeldung an ein einzelnes Kind bleiben von der Note her ausgeschlossen.

Bleiben Verabredungen für eine möglichst gleichmäßige und verfahrensgerechte Leistungsfeststellung und -beurteilung. Auch hier einmal ein Beispiel:

1. Nur lebensvolle, aus dem Unterricht erwachsene Texte werden verwendet. Es wird keine Häufung von Schwierigkeiten in den Text eingebaut.

2. Diktierregeln

- Der Text des ganzen Diktats wird einmal vom Lehrer/von der Lehrerin vorgelesen.
- Das Diktat erfolgt satzweise, längere Sätze werden in einzelnen, sinngemäß zusammenhängenden Abschnitten diktiert.
- Kein Schüler darf schreiben, wenn der Lehrer/die Lehrerin noch diktiert.
- Zum Abschluss wird das ganze Diktat noch einmal vorgelesen.

3. Es sind einheitliche Korrekturzeichen anzustreben. Es genügen drei:

Rechtschreibfehler im Wort/Verdrehungen im Satzbau:	/
Zeichenfehler:	-
Auslassungsfehler:	V

4. Alle Fehler, auch die Zeichensetzungsfehler (soweit die Zeichensetzung gewusst werden muss), zählen als ganze Fehler. Mehrere Fehler in einem Wort gelten als ein Fehler.

Wiederholt auftretende gleiche Fehlerwörter zählen nur einmal. Es gibt keine halben Fehler.

5. Unleserliche Wörter gelten als Fehler.

6. Nach Einführung der Regeln für die Trennung gilt das Über-den-Rand-Schreiben als Fehler.
7. Die Schrift wird unabhängig von der Rechtschreibleistung bewertet.

Es wurde für die vorstehenden Punkte von Verabredungen gesprochen. Sie könnten auch anders gesetzt werden. Dass sie aufgestellt werden, ist angesichts der bekannten Phänomene subjektiver Fehlerdefinition und -buchung, die alles vorneweg Festgelegte konterkarieren könnten, sicher sinnvoll.

Aus dem vorgestellten Beispiel, das leicht analog für Mathematikarbeiten und verschärfend für Sachunterrichtsarbeiten oder gar Aufsätze ergänzt werden könnte, werden die Notwendigkeiten und Fragwürdigkeiten deutlich, auch wenn es als in gewisser Weise fortschrittliches Beispiel gelten kann. Häufig ist die Beurteilungspraxis sehr viel wildwüchsiger. Das vom Arbeitskreis Grundschule in einer Publikation vorgestellte Zensierungsmodell für differenzierte Diktate hat dieselben Tücken (Günther 1992):

Allen Schülern wird ein Grundtext diktiert, der die Grundanforderungen beinhaltet, zusätzlich können die Schüler/innen einen Ergänzungstext mit weiteren Anforderungen schreiben. Für die Zensierung wird folgende Tabelle angegeben:

Zensur	Grundtext	Erweiterter Text	Fehler
1	+	+	zusammen 0
2	+	+	zusammen bis 2
3	+ Fehler: 0, 1, 2	+ fehlt	zusammen bis 5
4	+ Fehler: 3, 4, 5	+ fehlt	zusammen bis 8
5	+ Fehler: mehr als 5	fehlt	

Dabei wird nichts gesagt über Grund- und weitergehende Anforderungen, über Kompensationsmöglichkeiten (etwa 3 Fehler im Grundtext und 0 Fehler im erweiterten (schwierigeren) Text führen zu einem Gut). Es wird rein additiv gedacht.



Informelle Tests - die bessere Variante?

Die Erörterung von Klassenarbeiten und ihrer Beurteilung muss zwangsläufig auch die Suche nach Alternativen einschließen. Zwischen den herkömmlichen Klassenarbeiten und standardisierten Tests stehen die sog. Informellen Tests, die Gesichtspunkten objektiver Leistungsmessung folgen. Der bisher nicht übertroffene Standard ist von GAUDE und TESCHNER schon Anfang der 70er Jahre entwickelt worden (Gaude u. Teschner 1971²). Sie gingen von der Intention aus, die Leistungsmessverfahren technisch zu verbessern, Beurteilungsmaßstäbe intersubjektiv konsistent und - wenn man so will - gleich ungerecht zu

machen! Die von ihnen entwickelten Informellen Tests sind nur insofern repräsentativ und gültig, als sie - berechenbar - auf die jeweiligen antizipierten und operationalisierten Ziele bezogen sind. Unter Informellen Tests verstehen sie eine nach bestimmten Kriterien auf objektive Auswertbarkeit angelegte Probearbeit für ausgewählte Gebiete eines Unterrichtsfachs mit definierter Zielsetzung. Dies ist im Grunde das zentrale Anliegen von Klassenarbeiten. Informelle Tests wollen keine Aussagen über "die" Leistungen von Schülern machen, sondern lediglich ermitteln, welche Leistungen zu den von Lehrern/innen ausgewählten Inhalten bei intersubjektiv konsistentem Beurteilungsmaßstab erbracht werden.

Der "heilsame" Ausgangspunkt für Informelle Tests sind also die möglichst genau formulierten Lehr-/Lernziele (man denke an die frühere Diskussion über operationalisierte Lernziele) in Bezug auf bestimmte Unterrichtsinhalte. Konstruktionsvarianten sind Auswahlantworten (aus zwei oder mehreren vorformulierten Lösungen (Wörter, Sätze, Symbole) soll die einzig richtige oder beste Lösung markiert werden), Ordnungsantworten (zwei oder mehrere vorformulierte Lösungen sollen nach einem angegebenen Kriterium in die richtige Reihenfolge gebracht werden) und freien Antworten, bei denen zwischen definitiven und komplexen Antwortmöglichkeiten unterschieden wird. Gilt dies für 'Papier-Bleistift-Verfahren', so lassen sich auch mündlich-akustische und manuell-experimentelle Verfahren, die hier aber nicht weiter verfolgt werden, entwickeln.

Beispiele für die oben genannten Antworttypen sind:

1. Auswahlantworten

- Die Alternativantwort: Öl ist eines der wichtigsten Exportgüter Venezuelas:

Richtig

Falsch

- Mehrfach-Wahl-Antwort: Wie viel Kilopond sind 1150 p?

a) 111 kp "

b) 11,5 kp "

c) 1,15 kp "

d) 0,115 kp "

2. Ordnungsantworten

- Die Zuordnungsantwort: Ordne bitte jedem der Begriffe aus der linken Spalte jeweils die drei passenden Begriffe aus der rechten Spalte zu:

- | | |
|---------------|-------------------------|
| 1. Szenarium | a) streiten |
| 2. Monolog | b) fertige |
| 3. Rückblende | Regieanweisungen |
| 4. Dialog | c) aus früherer Zeit |
| 5. Textbuch | d) umrisshaft festlegen |
| | e) Wechselrede |
| | f) Erinnerung |
| | g) allein |
| | h) Entwurf |
| | i) Selbstgespräch |
| | j) ausgearbeitet |
| | k) vergangen |
| | l) Zwiegespräch |
| | m) Stichworte |
| | n) ohne Antwort |
| | o) vollständiger Text |

3. Freie Antworten

- Die Ergänzungs- bzw. Kurz-Antwort: Find die Gegensätze:

Father's car ist old. It is not.....

- Die Kurzaufsatz-Antwort: Zu dem folgenden Satz sollst du jeweils Fragen stellen. Der Anfang der Frage wird dir gegeben:

He goes to school every day. When

Nun kann man mit diesen Aufgabenarten einen Informellen Test konstruieren. Über die Bedingungen der Konstruktion von Aufgaben können hier keine näheren Ausführungen gemacht werden. Aber es ist bekannt, dass die Ratewahrscheinlichkeiten bei den Auswahlantworten im Prinzip recht groß sind, dass die Distraktoren (die falschen Antworten) Plausibilität haben müssen, wenn sie nicht von vornherein als Antwortmöglichkeit ausfallen sollen. Die genauere Ausmessung der Reproduktions- bzw. Denkleistungen bei den aufgeführten Aufgabenarten sind unterschiedlich. Es gibt weitere zu klärende Probleme. Ein Informeller Test muss eine sachgemäße und einheitliche Durchführung in vorgegebener Zeit sichern. Der Auswertungsmodus muss im vornhinein geklärt sein. Bei der Auswertung wird es zunächst um die Erfassung der Rohpunktzahl gehen. Zweiter Schritt ist die Transformation der Rohpunktwerte (das ist die von einem Schüler erreichte Gesamtpunktzahl in einem Test, z.B. 53 von 80 zu erreichenden Punkten) in Prozentwerte. Nun könnte man es bei der Feststellung dieser Prozentwerte belassen (Du hast 50% der zu erreichenden Punkte erreicht). - Wenn aber eine Zensierung als unerlässlich angesehen wird, muss eine Zuordnung der Punktwerte zu Noten erfolgen. Ein Beispiel:

Punkte	Note
100-91	1
90-81	2
80-71	3
70-61	4
60-51	5
50-41	6

Hier treten auch beim Informellen Test regelmäßig die alten Probleme wieder auf: Reicht für ein Ausreichend nicht die Hälfte der zu erreichenden Punkte oder müssen es mehr sein? Will man zu Gunsten des tatsächlich ausgefallenen Ergebnisses die Notenskala verschieben (nach oben oder nach unten), will man unter der Hand Normalerziehungsvorstellungen doch wieder stärker berücksichtigen? Falls man Informelle Tests in Leistungslernsystemen, z.B. in Englisch oder Mathematik, durchführt, ergibt sich zusätzlich die Frage, wie man die Noten staffelt. Nehmen wir an, in Englisch ist ein A-, B-, C-System (Fortgeschrittenen-, Grund- und Förderkurs) eingerichtet, so könnte eine Staffelung folgendermaßen aussehen:

Punkte	Noten		
	A-Kurs	B-Kurs	C-Kurs
100-91	1		
90-81	2	1	
80-71	3	2	
70-61	4	3	1
60-51	5	4	2
50-41	6	5	3
40-31		6	4
30-21			5
20-11			6
10-00			

Es ist klar, dass zu der Note dann der Kurs angemerkt werden muss, in dem die jeweilige Note erreicht worden ist. Aber die Staffelung bleibt eine Ermessensfrage und unterliegt nicht Objektivierungsargumenten. Das heißt, dass die Reichweite Informeller Tests in Bezug auf Leistungsfeststellung valide, objektiv und reliabel ist, in Bezug auf Leistungsbeurteilung doch wieder Plausibilitätsgründen unterliegt, die aber immerhin eine größere Breite von Vergleichbarkeit haben kann.



Das schwierigste Feld: Aufsätze, freie Texte

Zum Schluss muss auf das schwierigste Thema eingegangen werden: Wie sind

Leistungsfeststellungen und -beurteilungen bei Aufsätzen, freien Texten, Klausuren auf eine einigermaßen verlässliche Grundlage zu stellen? Hier liegen aus Untersuchungen die größten Mängelberichte vor. Es war schon erwähnt worden, dass die gleichen Texte von Lehrern/innen die Noten 1 bis 5 bekommen, der gleiche Lehrer zu verschiedenen Zeiten den gleichen Text unterschiedlich bewertet. Alle anderen genannten Fehlermöglichkeiten haben hier die größten "Chancen".

Wenn man ein Stück aus den Zufälligkeiten, Willkürlichkeiten und Sinnlosigkeiten heraustrennen will, muss man "weite Wege" gehen. Drei Möglichkeiten seien aufgezeigt:

1. Der Erwartungshorizont

Bei einem Geschichtsunterricht z.B. werden mehrere Aufgaben gestellt:

1. Stelle den Verlauf des 30-jährigen Krieges dar!
2. Nimm Stellung zu der Entscheidung von x!
3. In welchem Zusammenhang stand die Gewinnung von Soldaten mit der Lage der Bevölkerung im Jahre x?

Wenn die Antworten einigermaßen verlässlich in ihrer Richtigkeit und Komplexität identifiziert und bewertet werden sollen, ist es notwendig, die jeweiligen erwarteten Antwortelemente im Vorhinein genau zu identifizieren und mit Bewertungen zu versehen.

2. Das Optimum-Beispiel

Eine andere Hilfe wäre, wenn der Lehrer/die Lehrerin für eine Nacherzählung z.B. selbst ein optimales Beispiel schreibt und dabei die unerlässlichen Inhalts- und Darstellungselemente identifiziert und mit Punktbewertungen versieht. Es ergibt sich dann eine erreichbare Gesamtpunktzahl, von der aus wieder die Staffelung der Noten vorgenommen werden kann.

3. Kriterienauflistung

Das Mindestanforderungsmerkmal ist eine Kriterienauflistung, die für einen Text eine Rolle spielen soll, die Kriteriengewichtung mit Punkten und der Auswertungsmodus vorher festgelegt und den Schülern/innen bekanntgemacht werden. Nehmen wir als schwieriges Beispiel die Fantasieerzählung. Sie steht in der Spannung zwischen der Freude am Fabulieren und dem Satzbau, der Wortwahl und der Rechtschreibung. Saloppe Wendungen - also die Eigenart der Sprache -, Wiederholungen, dramaturgische Gesichtspunkte, die Aufgabenstellung und das Verhältnis der Schüler/innen zu ihr können schließlich zu sehr unterschiedlichen Texten führen, die sich einer Bewertung fast entziehen, so dass man provokativ sagen könnte, die noch plausibelste Bewertung läge in der Resonanz, die eine vorgetragene Fantasieerzählung in der Klasse erfährt (der Zuhörer als Beurteiler!).

Soll die Fantasieerzählung aber eine Klassenarbeit mit Bewertung und mit Berücksichtigung für die Deutschnote sein, müssen die Bewertungskriterien klar formuliert werden und über das Meinen des Lehrers/der Lehrerin hinausgehen. Könnte folgende Kriterienauflistung ein gutes Beispiel sein?

a)Dramaturgie	12 Punkte
- folgerichtig erzählt	3 Punkte
- Einleitung	1 Punkt
- Höhepunkt	2 Punkte
- Spannung	2 Punkte
- Schluss	1 Punkt
- Unwirkliches gut dargestellt	3 Punkte
b)Verwendete Sprache	12 Punkte
- richtig formulierte Sätze	2 Punkte
- differenzierte Syntax (zusammengesetzte Sätze)	4 Punkte
- reichhaltige Wortwahl (Verben, Adverbien, Attribute)	4 Punkte
- wörtliche Rede	2 Punkte
c)Rechtschreibung wird nicht gewertet	0 Punkte
Insgesamt ergeben sich so 24 Punkte.	

Der Autor nimmt hier eine ausgesprochen zögerliche Haltung ein! Aber um einigermaßen das Postulat der Transparenz zu wahren, könnte die Kriterienauflistung mit Auswertungsmodus noch der beste Weg sein.



Schluss

Mit den vorstehenden Ausführungen ist der Versuch gemacht worden, Schneisen in ein schulpädagogisches Aufgabenfeld zu schlagen, die einigermaßen Verlässlichkeit und Transparenz für die Schüler/innen sichern helfen können. Der grundsätzliche Vorbehalt bleibt bestehen. Da bleiben Tücken und Vorläufigkeiten. Irritation und Zufälligkeiten können aber minimiert werden, wenn man den verschiedenen dargestellten Ansätzen folgt. Wir wissen von Schülerbefragungen, dass Leistungsfeststellungen und -beurteilungen das leidvollste Kapitel der Schule sind. Die Herausforderung ist daher groß, hier etwas zu verändern!



Literatur

BÖNSCH, M.: Didaktisches Minimum. Neuwied 1986. - ders.: Didaktisches Additum. -

Neuwied 1998. - BOHNSCH, O.: Neue Kriterien für die Aufsatzbewertung. München 1975². - BOLSCO, D., SCHWARZER, Chr.: Beurteilen in der Grundschule. München 1979. - DUMKE, D.: Schülerleistung und Zensur. Hannover 1973. - GAUDE, P., TESCHNER, W.-P.: Objektivierete Leistungsmessung in der Schule. Frankfurt/M. 1971². - GÜNTHER, Chr.: Zensierungsmodell. In: Bartwitzky, H., Portmann, R.: Leistung der Schule - Leistung der Kinder. Frankfurt/M. 1992. - HELLER, K. (Hrsg.): Leistungsbeurteilung in der Schule. Heidelberg 1975². - INGENKAMP, K. (Hrsg.): Die Fragwürdigkeit der Zensurengebung. Weinheim 1971 (viele Auflagen). - ders.: Lehrbuch der pädagogischen Diagnostik. Weinheim, Basel 1988. - KLAUER, K.J. (Hrsg.): Handbuch der pädagogischen Diagnostik, 4 Bde. Düsseldorf 1978. - KLEBER, E.W. u.a.: Beurteilung und Beurteilungsprobleme. Weinheim 1976. - MEYENBERG, R.: Zensurengebung in der Schule. Hannover 1986². - RHEINBERG; F.: Leistungsbeurteilung und Lernmotivation. Göttingen 1979. - SACHER, W.: Praxis der Notengebung. Bad Heilbrunn 1984. - SCHRÖBER, G.: Die ungerechte Aufsatzzensur. Bochum 1971. - STRITTMATTER, P. (Hrsg.): Lernzielorientierte Leistungsmessung. Weinheim 1973. - ULICH, D., MERTENS, W.: Urteile über Schüler. Weinheim 1973.



Anschrift des Verfassers:

Prof. Dr. Manfred Bönsch
(Universität Hannover, FB Erziehungswissenschaften, Inst. f. Erziehungswissenschaften)
Bismarckstr. 2, 30173 Hannover

* * *

Quelle:

Manfred Bönsch: Klassenarbeiten bewerten und beurteilen. Schwerpunkt: Möglichkeiten, aber auch Sinnlosigkeiten von Klassenarbeiten und Leistungsbeurteilungen

Erschienen in: BEHINDERTENPÄDAGOGIK, 39. Jg., Heft 3/2000, Seite 312-322

[bidok](#) - Volltextbibliothek: Wiederveröffentlichung im Internet

<ftp://ftp.uibk.ac.at/pub/uni-innsbruck/bidok/bhp/bhp3-00-klassenarbeiten.zip> (RTF-Version)

<http://bidok.uibk.ac.at/bhp/bhp3-00-klassenarbeiten.html> (HTML-Version)

Stand: 25. April 2001

