

Prof. Dr. Renate Valtin

Probleme der Leistungsbeurteilung –  
Was ist ein gutes Zeugnis?

## Hildegard Baumgarten 1917:

- *Die Schulnote ist der "Giftpilz des Haus- und Schullebens" und eine unerschöpfliche Quelle des Lügens.*

# Wovor hast du manchmal Angst? (Schmude)



## Mündliche Interviews:

- Zu Beginn von Klasse 2:
  - Vorwiegend ängstigende Ereignisse außerhalb der Schule (z. B. Trennung der Eltern, Besuch beim Zahnarzt)
  - Schulische Ängste etwa ein Drittel der Nennungen
- Mitte 2. Klasse, als die meisten Kinder Noten erhielten:
  - Nur noch 11 % außerschulische Ängste
  - 75 % schulische Ängste, wie Angst vor Klassenarbeiten und Noten

## Leistungsangst (%) (Klasse 6-9)

	Klasse	1 ja
Wenn du abends im Bett liegst, machst du dir dann manchmal Sorgen, wie du am nächsten Tag im Unterricht abschneiden wirst?	6.	32,2
	7	44,0
	8	36,8
	9	36,3
Machst du dir auf dem Schulweg manchmal Sorgen, die Lehrerin/der Lehrer könnte heute nachprüfen, wie gut du in der Schule bist?	6	30,5
	7	36,4
	8	36,0
	9	36,5
Machst du dir Sorgen, ob du in die siebte/achte Klasse versetzt wirst?	6	24,0
	7	32,4

# Gliederung

- Warum braucht man ein Zeugnis?
- Wie beurteilen Kinder, Eltern und Lehrkräfte die Formen der Zeugnisse?
- Werden Noten und Berichtszeugnisse den in sie gesetzten Erwartungen gerecht?
- Haben Noten und verbale Beurteilungen unterschiedliche Auswirkungen auf die Leistungs- und Persönlichkeitsentwicklung der Schülerinnen und Schüler?
- Welche Form der Leistungsrückmeldung ist zu empfehlen?

# Funktionen von Zeugnissen

- Gesellschaftliche Funktion:
  - Dokumentation eines erreichten Bildungsabschlusses (Zertifizierung)
  - Legitimation für Ausleseentscheidungen
- Pädagogische Funktion:
  - (für Eltern und Kind):
  - Information über Lernstand
  - Hinweis auf Fördermöglichkeiten

# Pädagogische Funktion für Lehrer

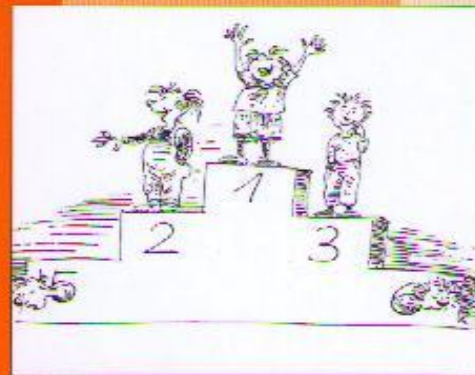


Ergebnisse  
aus  
dem Projekt  
NOVARA

Renate Valtin

# Was ist ein gutes Zeugnis?

Noten und verbale  
Beurteilungen auf  
dem Prüfstand



JUVENTA



# NOVARA (Noten- oder Verbalbeurteilung – Akzeptanz, Realisierung und Auswirkungen)

- NOVARA ist das bislang umfassendste empirische Projekt zur Leistungsbeurteilung in der Grundschule.
- Die von der DFG geförderte Längsschnittstudie umfasst Leistungstests, Schüler-, Eltern und Lehrerbefragungen.
- Kinder aus 40 Klassen wurden zu 7 Messzeitpunkten von Beginn der 2. Klasse an bis zur Mitte der 6. Klasse zunächst in mündlichen Einzelinterviews, dann schriftlich zu verschiedenen Aspekten von Zeugnissen befragt, wie Funktion von Zeugnissen, Präferenz von Beurteilungsformen, Zufriedenheit mit der Beurteilung sowie ihre Erinnerung an ihre jeweiligen Noten- oder Berichtszeugnisse.
- Die Eltern wurden zweimal befragt nach ihren Erfahrungen, Wünschen und Einstellungen zu Zeugnissen und zu schulischen Reformen.
- Erforscht wurden ferner die Realisierung der Leistungsbeurteilung durch die Lehrkräfte (die Praxis der Notengebung und die Abfassung der verbalen Beurteilung) sowie
- die Auswirkungen dieser beiden Formen der Leistungsbeurteilung durch einen längsschnittlichen Vergleich der Leistungs- und Persönlichkeitsentwicklung von Schülerinnen und Schülern aus Klassen mit Noten bzw. verbaler Beurteilung.

# Wozu braucht man ein Zeugnis?

Kinder der 2. Klasse:

- Zwei Drittel betonen die *pädagogische* Funktion der Rückmeldung über ihren Leistungsstand:  
*„damit man weiß, ob man gut oder schlecht ist“.*

# Wozu braucht man ein Zeugnis?

Kinder der 2. Klasse:

- Zwei Drittel betonen die *pädagogische* Funktion der Rückmeldung über ihren Leistungsstand:  
*„damit man weiß, ob man gut oder schlecht ist“.*

*“Weil man das nicht wissen kann, wie gut man ist und wie gut man nicht ist”.*

# Wozu braucht man ein Zeugnis?

Kinder der 2. Klasse:

- Ein Drittel betont die *pädagogische* Funktion der Rückmeldung über ihren Leistungsstand an die Eltern:

*„damit die Eltern wissen, ob man gut oder schlecht ist“.*

*“ damit sie wissen, ob man was ausgefressen hat.“*

# Wozu braucht man ein Zeugnis?

Kind der 2. Klasse:

*„Ein Zeugnis unterscheidet die Schwächeren von den Guten.“*

## Cartoon von Amelie Gliencke



# Die Mehrheit der Kinder wollen keine Notenzeugnisse

- Ein reines Notenzeugnis wünschten sich im Projekt NOVARA nur 14% für die 1. Klasse und 21% für die 6. Klasse.
- Die Mehrheit wünscht Ziffernzeugnisse, versehen mit schriftlichen Kommentaren und/oder Elterngesprächen.
- Sie möchten keine ‚Kopfnote‘, sondern einen Beurteilungstext.

# Eltern schätzen Noten, verbunden mit erläuternden Kommentaren

- „*Noten sind objektiver als die verbale Beurteilung*“.
- Selbst 80 % der Anhänger der Verbalbeurteilung unterstützen die Aussage: *"Bei einem Notenzeugnis weiß man genau, wo das Kind steht"*
- Etwa 40% der Eltern sehen in Noten ein wichtiges Mittel zur Disziplinierung von Schülern.
- Gleichzeitig sehen Eltern als Vorteile der verbalen Beurteilungen an, dass diese konkrete Hinweise zum Lernstand und zur Förderung geben können.
- Insofern ist es verständlich, dass sich Eltern am häufigsten ein Zeugnis wünschen, das Noten und erläuternde Kommentare verbindet.



# Einschätzungen der Lehrkräfte

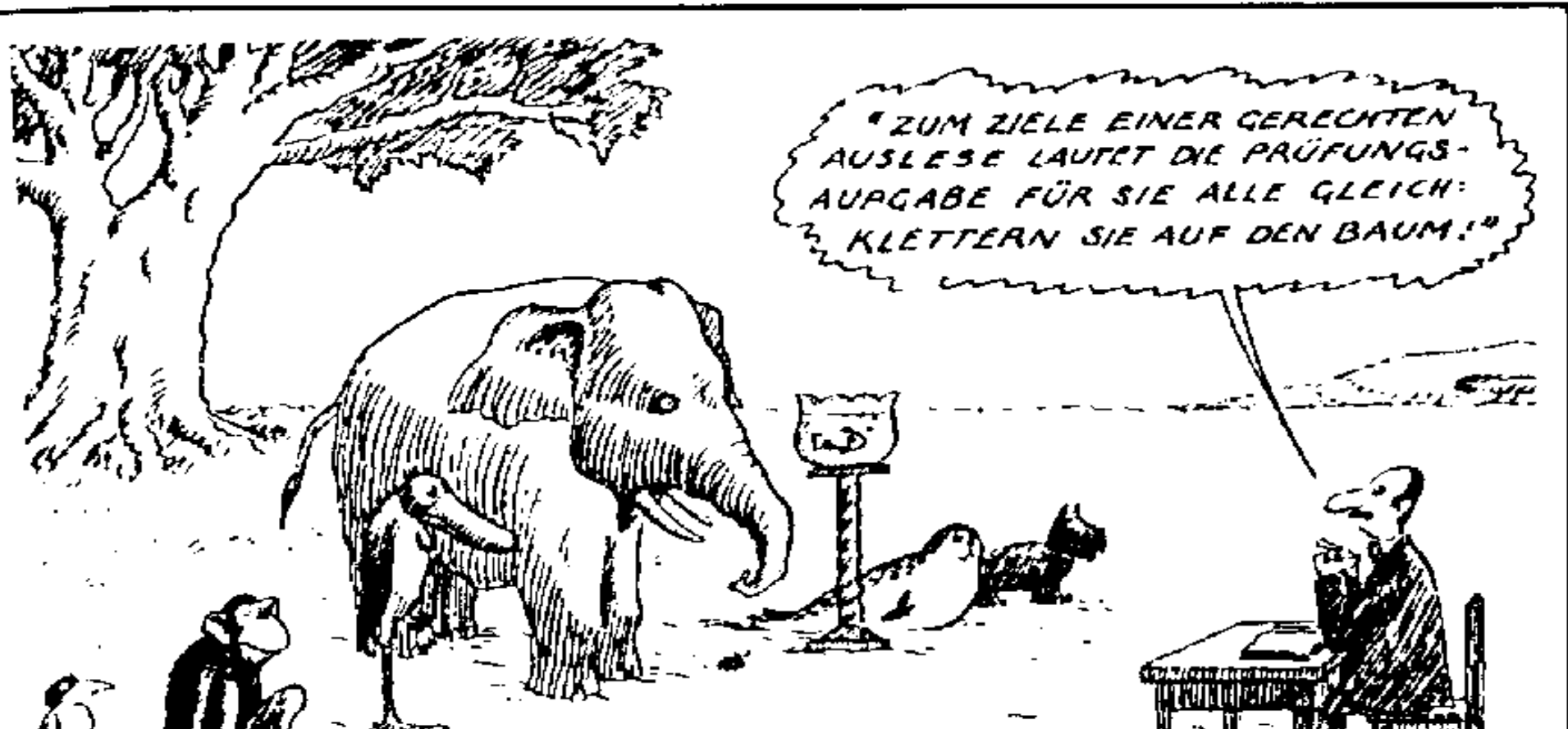
- *„Noten gehören zur Schule dazu“ (75%)*
- *„Noten sind notwendig, um Schüler zum Lernen anzuspornen“ (50%)*
- Allerdings stimmten auch 44% der Äußerung zu: *„Noten sind überflüssig, ausformulierte Beurteilungen sind wesentlich aussagekräftiger“*
- Mehr Lehrkräfte aus Grund- und Sonderschulen (60%) befürworteten ausformulierte Beurteilungen als Lehrkräfte von Realschulen (21%) oder Gymnasien (28%).

# Problematik der Leistungsbeurteilung

Unterscheidung von drei Ebenen:

- (1) die Grundlage der Bewertung (Frage: Auf welchem Wege gelange ich zur Feststellung der Leistung?)
- (2) die Wahl des Maßstabs (Frage: An welchen Kriterien orientiere ich mich bei der Bewertung?)
- (3) die Darstellungsform der Leistungsrückmeldung (Frage: Wähle ich Ziffern, wie beispielsweise Noten bzw. die Anzahl der richtigen/der falschen Lösungen oder verwende ich sprachliche Formulierungen, d.h. kürzere oder längere verbale Beurteilungen?).

# Die Grundlage der Bewertung



# Verfahren zur Feststellung der Leistung

- Gelegentliche oder systematische Beobachtungen im Unterricht, mündliche, schriftliche und andere fachspezifische Lernkontrollen, informelle oder standardisierte Aufgaben (Tests oder Lernstandserhebungen).
- Diese Verfahren unterscheiden sich in ihrer Güte und sind von unterschiedlicher Objektivität, Zuverlässigkeit und Gültigkeit sind.

# Subjektive Fehlerquellen

- Die ungleichmäßige Ausschöpfung des Beurteilungsspektrums:
  - Milde, Strenge oder Tendenz zur Mitte
- Voreingenommenheiten im Urteil:
  - *Halo-Effekt* (Hof- oder Überstrahlungseffekt): ein hervorstechendes Merkmal "überstrahlt" in positiver oder negativer Weise die Beurteilung anderer Bereiche
  - *Reihungs-* und *Kontrasteffekte*: Die ersten Schülerarbeiten werden häufig strenger zensiert als die zuletzt gelesenen. Nach einer Reihe mäßiger Arbeiten wird eine etwas bessere besonders gut, eine etwas schlechtere als besonders schlecht bewertet.

Die subjektiven Fehlerquellen tauchen sowohl bei der Notengebung als auch bei der verbalen Beurteilung auf.

Diese beeinträchtigen die Gütekriterien, die eine Leistungsmessung erfüllen sollte:

- **Objektivität** (Unabhängigkeit des Urteils von subjektiven Einstellungen des Beurteilenden),
- **Reliabilität** oder Zuverlässigkeit (Genauigkeit und Sicherheit der Leistungsfeststellung, so dass bei wiederholter Bewertung die Lehrkräfte dieselbe Einschätzung abgeben sollten) sowie
- **Validität** (Gültigkeit und Inhaltsangemessenheit der Aufgabe).

# Fehlerquellen

- Gut belegt ist, dass verschiedene Lehrkräfte dieselbe Schülerarbeit unterschiedlich beurteilen (nicht nur schwer einschätzbare Arbeiten wie Aufsätze, sondern auch Arbeiten mit eindeutigen Lösungen wie bei der Rechtschreibung oder Mathematik).
- Jedoch benoten auch dieselben Lehrer dieselbe Arbeit zu unterschiedlichen Zeiten unterschiedlich.
- Einige Merkmale von Schülern begünstigen positive Voreinstellungen: Beliebte Schüler und Schülerinnen werden besser benotet. Schriftliche Arbeiten, die in sorgfältiger Handschrift verfasst sind, erhalten bessere Noten als unsauber geschriebene.

# Die Studie von Rudolf Weiss, 1965

## Ein Hund allein auf der Straße

Meine Mutter und ich gingen einmal spazieren. Plötzlich stolperte ich. Die Mutter lachte und sagte: „Du bist ja über einen Hund gestolpert!“ Es war ein lieber, kleiner Budel. Da sahen wir, das der arme Hund ganz allein auf der Straße ist. Er schaute uns ganz treuherzig an. Wir gingen dann zur Polizei und lieferten ihn ab. Ich dachte, er muß sich verirrt haben.

Rechtschreiben:      Stil:      Inhalt:      Gesamtnote:

## Nikolaus und Krampus

Am 5. Dezember 1963 lud mein Freund den Harald und mich ein. Wir hatten vor, Nikolaus und Krampus zu spielen und zum Hausmeisterssohn zu gehen. Wir richteten uns her. Mein Freund ging als Nikolaus, Harald und ich gingen als Krampusser. Wir nahmen einen Sack, steckten Äpfel, Nüsse und zwei Tafeln Schokolade hinein. Schnell schrieben wir auf einen Zettel: „Weil du immer so brav warst, haben wir dir das gebracht!“

Der Nikolaus legte den Sack auf die Fußmatte. Wir Krampusser schrien recht. Der Hausmeisterssohn weinte recht, und wir lachten. Dann gingen wir nach hause.

Rechtschreiben:      Stil:      Inhalt:      Gesamtnote:



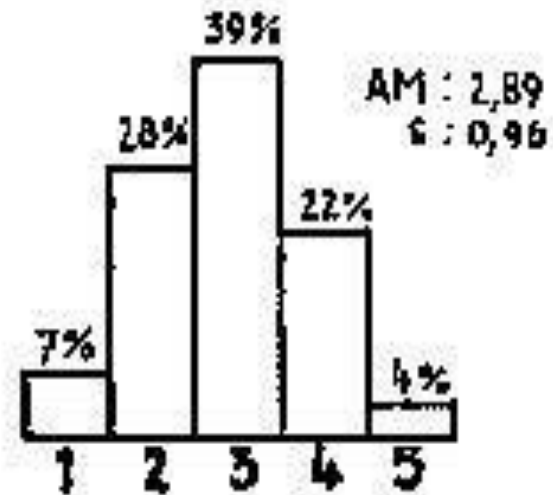
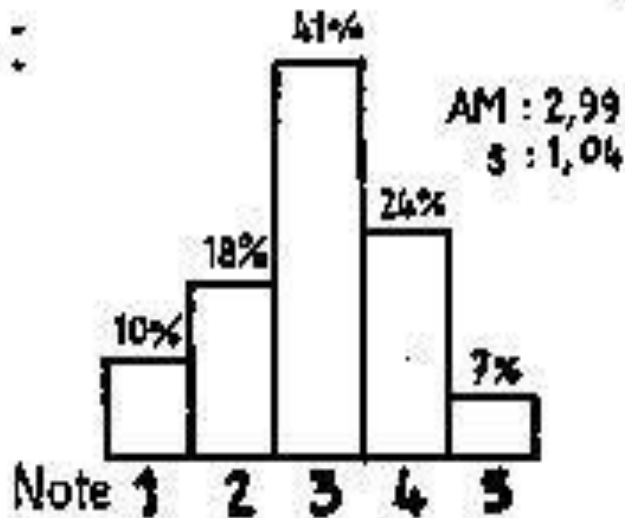
# Studie von Weiss

## Beurteilung der Aufsätze der 4. Schst.

„Nikolaus und Krampus“ „Ein Hund allein auf der Straße“

N = 92

Rsch. :



# Positive vs. negative Beeinflussung

Den Aufsätzen war bei einem Teil der Versuchspersonen folgende Einleitung vorangestellt:

Zwei Aufsätze aus der 4. Klasse einer Volksschule. Der erste stammt von einem durchschnittlichen Schüler (beide Elternteile berufstätig, liest gerne Schundhefte), der zweite von einem sprachlich begabten Buben (Vater Redakteur bei einer großen Linzer Tageszeitung).

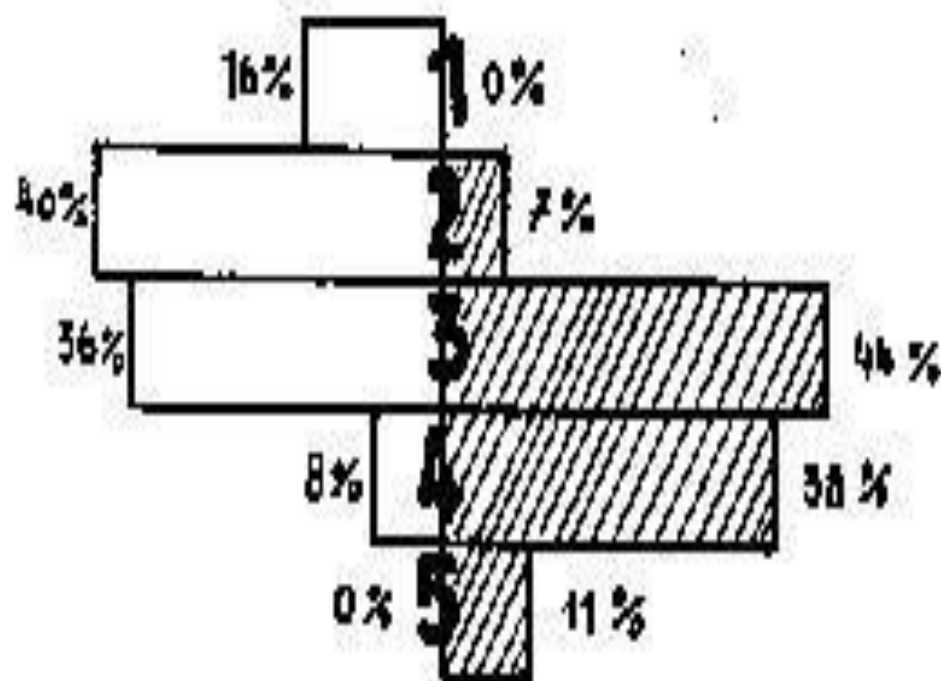
# Beurteilung der Aufsätze unter positiver bzw. negativer Beeinflussung

positives negatives  
Vorurteil

N = 92

Rschr.

AM: 2,35



AM: 3,35

# Nachfolgestudie 2010 (Kirsten Becker/Astrid Kaiser): Bedeutung des Vornamens

- Über 200 Grundschullehrer bewerteten schriftliche Aufgabenlösungen von Kindern – dabei waren dieselben Lösungen mal mit negativ („Kevin“), mal mit positiv etikettierten Vornamen („Maximilian“) belegt.
- Die Lehrer mussten die Antwortvorgaben aus dem Bereich Sachunterricht auf "Rechtschreibung", "Stil", "Inhalt" und "Gesamtnote" auf einer Punkteskala zwischen 1 und 10 bewerten.
- Obwohl die Antworten gleich waren, variierten die Bewertungen durch die Lehrer oftmals um neun Bewertungspunkte.

<http://bildungsklick.de/a/74753/schlechtere-noten-fuer-kevin/>

- Bessere Noten für Maximilian als für Kevin.
- Bei Aufgaben mit Namensnennungen aus dem negativ bewerteten Namenspool erhielten Jungen eine durchweg schlechtere Bewertung.
- Welche Jungennamen gute Noten erhielten:  
Alexander Jakob Lukas Maximilian
- Welche Jungennamen schlechtere Noten erhielten:  
Cedric Justin Kevin Marvin

Fachfremde Einflussgrößen bei der Notengebung eine Rolle spielen:

- Beliebtheit
- Reihenfolge
- Handschrift
- Geschlecht
- soziale Schicht
- Migrationsstatus

# W. C. Eells: Die Zuverlässigkeit wiederholter Benotung (1930)

61 Lehrer benoteten dieselben 4  
Aufgaben nach 11 Wochen noch einmal:

- Kein Lehrer wiederholte seine Bewertungen bei allen vier Fragen
- 2 Lehrer wiederholten ihre Bewertungen nur bei drei Fragen
- 17 Lehrer wiederholten ihre Bewertungen nur bei zwei Fragen
- 23 Lehrer wiederholten ihre Bewertungen nur bei einer Frage
- 19 Lehrer wiederholten ihre Bewertungen bei keiner Frage

- Weitere Studien  
in:





# Die Wahl des Maßstabs zur Bewertung einer Leistung

## Definition der Noten laut Kultusministerkonferenz (1968)

- **sehr gut** - wenn die Leistung den **Anforderungen** in besonderem Maße entspricht
- **gut** - wenn die Leistung den Anforderungen voll entspricht;
- **befriedigend** - wenn die Leistung im Allgemeinen den Anforderungen entspricht;
- **ausreichend** - wenn die Leistung zwar Mängel aufweist, aber im Ganzen den Anforderungen noch entspricht;
- **mangelhaft** - wenn die Leistung den Anforderungen nicht entspricht, jedoch erkennen lässt, dass die notwendigen Grundkenntnisse vorhanden sind und die Mängel in absehbarer Zeit behoben;
- **ungenügend** - wenn die Leistung den Anforderungen nicht entspricht und selbst die Grundkenntnisse so lückenhaft sind, dass die Mängel in absehbarer Zeit nicht behoben werden könnten.

# Noten sind nicht vergleichbar

- *Innerhalb eines Schuljahrs* gibt es harte und weiche Fächer, in denen unterschiedlich streng zensiert wird.
- *Über verschiedene Schuljahre hinweg*: Mit der Dauer der Schulzeit wird strenger zensiert.
- *Noten aus unterschiedlichen Klassen* sind nicht vergleichbar aufgrund des klasseninternen Maßstabs des Lehrerurteils. „*Eine Zwei ist eine Drei ist eine Vier*“ (Thiel/Valtin 2002),

# IGLU 2006

Tabelle III.15: Einschätzung der Lesekompetenz durch die Lehrkräfte (auf einer Zensurenkala von 1 bis 6)

Zensur	Legastheniker			andere Leseschwache			mit Lese-schwierigkeiten			Lesestärke		
	n	%	(SE)	%	(SE)	n	%	(SE)	n	%	(SE)	
1	2	0.8	(0.6)	3	0.7	(0.3)	111	4.2	(0.6)	960	23.3	(1.2)
2	17	6.0	(1.9)	21	4.3	(1.1)	680	25.7	(1.3)	1983	48.1	(1.4)
3	89	31.0	(3.5)	142	29.5	(3.3)	1135	42.9	(1.4)	1018	24.7	(1.3)
4	137	47.9	(3.7)	186	38.8	(2.9)	614	23.2	(1.4)	140	3.4	(0.5)
5 und 6	41	14.3	(2.7)	128	26.7	(3.5)	106	4.0	(0.8)	16	0.4	(0.2)

# Auswirkungen unterschiedlicher Bezugsnormorientierung

- Die Orientierung an *individuellen* Lernfortschritten der Kinder fördert die Leistungsmotivation. Vor allem schwache Schülerinnen und Schüler haben weniger Furcht vor Misserfolg, mehr Hoffnung auf Erfolg und günstigere Selbsteinschätzungen, was zu besseren Leistungen führen kann.
- An der *sozialen Bezugsnorm* orientierte Leistungsrückmeldungen wirken bei leistungsschwächeren Schülern deutlich negativ, bei leistungsstärkeren neutral oder leicht positiv.

# Die Darstellungsform der Leistungsrückmeldung: Ziffern oder sprachliche Kommentare.

## Was leisten sie?

- Die Informationsarmut der Zensuren beeinträchtigt die pädagogische Funktion der Leistungsbeurteilung. Kleinere Lernfortschritte können den Kindern nicht rückgemeldet werden. Noten können wenig über Stärken, Schwächen und die Lernentwicklungen der Kinder aussagen.
- Ziffern eignen sich auch nicht zur diagnostischen Leistungsbeurteilung, die im alltäglichen Unterricht notwendig ist, um die Lernvoraussetzungen der Kinder festzustellen und um didaktische Maßnahmen zu entwerfen, durchzuführen und auf ihren Erfolg hin zu überprüfen.
- Die Zensur erfüllt damit ihre wesentlichen pädagogischen Funktionen nicht.

- Die Zensur wird auch ihrem gesellschaftlichen Anspruch der Legitimation von Ausleseentscheidungen nicht gerecht, da fachfremde Gesichtspunkte, z. B. Leistungsstand der Klasse, Geschlecht, Bildungs- und Migrationshintergrund, bei der Zensurenvergabe eine Rolle spielen.

# Intentionen der verbalen Beurteilungen

- Sie sollen den Mängeln der Notengebung begegnen und großen Ansprüchen gerecht werden:
- umfassende Beschreibung nicht nur des Lernproduktes, sondern auch des Lernprozesses,
- Diagnose von Stärken und Schwächen der Kinder,
- Empfehlungen für Fördermaßnahmen sowie
- Ermutigung.

# Empirischer Befunde zu Berichtszeugnissen

- Fehlender Bezug auf die individuelle Lernentwicklung,
- Ungleichgewicht der Fächer und Leistungsdimensionen, d.h. starke Dominanz der Lese-, Rechtschreib- und Rechenleistungen,
- Fehlende Fördervorschläge
- Beschönigung der Rückmeldungen
- Standardisierung der Aussagen durch Nutzung von Textbausteinen (Brügelmann u.a. 2006, S. 33).



Haben Noten und verbale Beurteilungen unterschiedliche Auswirkungen auf die Leistungs- und Persönlichkeitsentwicklung der Schülerinnen und Schüler?

# Einschränkung der Ergebnisse

- Der Sachverhalt, dass Berichtszeugnisse pädagogisch nicht ihren Möglichkeiten entsprechend genutzt werden, schränkt natürlich mögliche positive Auswirkungen ein.
- Allerdings ist auch noch ein weiterer Aspekt zu beachten: die Unterrichtsform. Verbale Beurteilungen machen nur Sinn in einem binnendifferenzierten, schülerorientierten Unterricht, der auf individuelle Förderung ausgerichtet ist. Ein derartiger Unterricht wird bislang jedoch nur selten praktiziert (vgl. u.a. Hanke 2005),

# Wirkungen der verbalen Beurteilung

- In Bezug auf die Schulleistungen ergeben sich keine Unterschiede zwischen Kindern aus Klassen mit Berichts- bzw. Notenzeugnissen (NOVARA, LAU-Studie, Lehmann u.a. 1997).
- Die in der Öffentlichkeit verbreitete Sorge, Kinder würden ohne Noten nicht lernen, ist also unbegründet.

# Ergebnisse aus dem NOVARA-Projekt

- Kinder mit schlechten Noten entwickelten eine **stärkere Leistungsangst**, vor allem im Bereich der Besorgtheit.
- In Bezug auf die Lernmotivation gab es Anzeichen für Notenangst und Notengeilheit (Sacher 1996, S. 74):
  - bei Kindern mit schlechten Noten stärkere **Misserfolgsorientierung** in der Leistungsmotivation
  - bei Kindern mit guten Noten eine stärkere **externale Motivation**.
- Im 3. Schuljahr schätzten die Notenkinder die schulischen Anforderungen als schwieriger ein

# Verbale Beurteilungen als Alternative für Noten?

- Keine Beurteilungsform erfüllt bislang alle Anforderungen:

*„Die Risiken von Fehlbeurteilungen und die negativen Nebenwirkungen lassen sich nur reduzieren, nicht gänzlich aufheben“*

(Brügelmann u.a., 2006, S. 53)

# Maßnahmen zur Verminderung von Fehlbeurteilungen

- „Bei der Erhebung von Leistungen sind *verschiedene Verfahren* wie standardisierte Tests, klassenbezogen gestellte oder individuell gewählte Aufgaben sowie informelle Beobachtungen zu kombinieren.
- Bei der Rückmeldung sind Leistungen mit Bezug auf *verschiedene Maßstäbe* auszuweisen
  - als **individuelle Fortschritte** gegenüber früheren Leistungen,
  - als **Grad der Lernzielannäherung** und
  - als Rangplatz in einer **repräsentativen** Bezugsgruppe.
- In den Beurteilungsprozess sind *verschiedene BewerterInnen* einzubeziehen, einschließlich der **Selbsteinschätzung durch die Betroffenen**, und eventuelle Differenzen ihrer Urteile explizit auszuweisen“ (Brügelmann u.a., 2006, S. 53).

# Zusammenfassung

- Leistungsbeurteilungen und auch Zeugnisse sollen vielfältigen und in sich **widersprüchlichen Funktionen** gerecht werden: Einerseits gesellschaftlich als Dokumentation eines erreichten Bildungsabschlusses und als Legitimation für Ausleseentscheidungen, andererseits pädagogisch als Feststellung des Lernstands und Hinweis auf Fördermöglichkeiten.

# Zusammenfassung

- Während in der empirischen Forschung die Nachteile von Zensuren seit langem erforscht sind (sie leiden an mangelnder Objektivität, Zuverlässigkeit und Gültigkeit), erfreuen sich Noten aus der Sicht der Betroffenen, der Kinder und Eltern, nach wie vor großer Beliebtheit.
- Das verbale Zeugnis hat nur wenige Anhänger bei Kindern, Eltern und auch unter Lehrkräften. Am beliebtesten sind Notenzeugnisse mit erläuternden Kommentaren.



# Zusammenfassung

- Die Fehlurteile, die bei Noten auftreten, gibt es auch bei der verbalen Beurteilung.
- Die Vorzüge der verbalen Beurteilung (Möglichkeiten der Orientierung am individuellen Lernfortschritt, der differenzierten Beschreibung des Lernstandes, der Hinweise auf Fördermöglichkeiten und der Ermutigung) werden bislang – so zeigen Analysen von Berichtszeugnissen – nicht ausreichend realisiert.

# Zusammenfassung

- Eine Zeugnisreform ist nur wirkungsvoll bei **verändertem Unterricht** und bei einer veränderten Funktion der Schule, die nicht auf Auslese, sondern auf die **gemeinsame und individuelle Förderung aller Kinder** gerichtet ist.
- Auch sollten Lehrkräfte eine bessere Ausbildung in Bezug auf Lern- und Entwicklungspsychologie, Fachdidaktik sowie Förderdiagnostik erhalten.

Erst nach Schaffung dieser Voraussetzungen ist damit zu rechnen, dass die verbale Beurteilung auch angemessen realisiert wird und nicht das ist, was einem Kind aus dem NOVARA-Projekt in einem Interview als Versprecher unterlief:  
eine “*verbale Verurteilung*”.