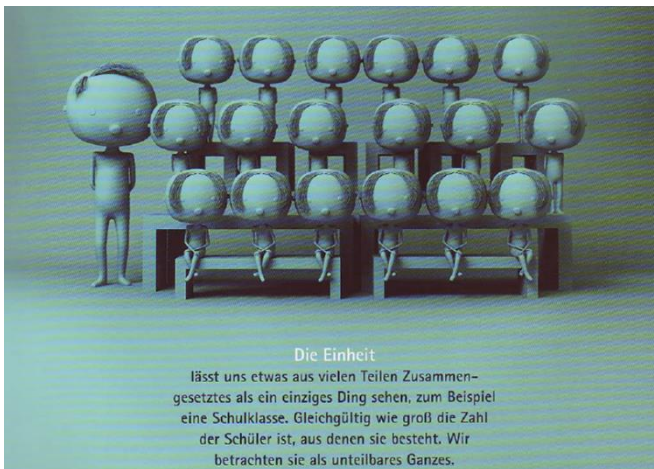


## Die Not mit den Noten

### Lernbeobachtung und Leistungsbeurteilung in der inklusiven Grundschule<sup>1</sup>

Wer eine Gruppe von 18 Kindern vor sich hat, kann ganz Unterschiedliches wahrnehmen. Die Frau vom Schulverwaltungsamt und der Schulleiter sehen eine Gruppe von 6-bis 7-Jährigen, für die ein Klassenzimmer bereitgestellt werden muss. Früher sahen auch die



Autor/inn/en von Lehrgängen eine „Klasse“, beispielsweise eine erste, die in den Wochen zuvor gemeinsam die Buchstaben F, U, R und T „gelernt“ hatten und in der nächsten Einheit alle das A lernen sollten.

Spätestens seit der Schulanfangsuntersuchung von Rathenow/ Vöge (1982) wissen wir aber, dass das eine Illusion ist. Nimmt man den altersbezogenen Durchschnitt als

Maßstab, liegen sie in ihren Voraussetzungen um etwa drei Jahre auseinander. Dies hat der große Schweizer Längsschnitt von Largo (2009) noch einmal eindrucksvoll bestätigt. Vor allem sind es nicht nur die kognitiven Voraussetzungen und die Fachleistungen, in denen sich die Kinder unterscheiden, sondern auch ihre Interessen, ihre Erfahrungen, ihre Persönlichkeit, ihre Wahrnehmungs- und Denkformen. Eine Lehrerin sieht deshalb etwas ganz Anderes:

<sup>1</sup> Dieser Beitrag ist ein teilweise gekürzter und zum Teil erweiterter Nachdruck aus b&w („Bildung und Wissenschaft“ der Gewerkschaft Erziehung und Wissenschaft Baden-Württemberg), H. 12/2013, 15-19, bzw. aus Peters, S./ Widmer-Rockstroh, U. (Hrsg.) (2014): *Gemeinsam unterwegs zur inklusiven Schule. Beiträge zur Reform der Grundschule* Bd. 138. Grundschulverband: Frankfurt.

Dieser Blickwechsel ist es, was „Inklusion“ von „Integration“ unterscheidet. Es geht nicht darum, die „Anderen“ in eine homogene Gruppe von „Normalen“ aufzunehmen, sondern darum, alle in ihrer Besonderheit wahrzunehmen und den Unterricht für diese Unterschiede zu öffnen.

Wie aber kann das gelingen, wenn der Lernerfolg an gleichen Anforderungen für alle zu demselben Zeitpunkt gemessen und in Form von Ziffernnoten nach der Glockenkurve der „Normalverteilung“ gemessen wird?



Die Vielfalt  
lässt uns eine Sache in ihre vielfältigen  
Elemente zergliedern. Wir haben dann nicht  
mehr die Schulklasse vor Augen, sondern ganz  
unterschiedliche Schüler  
und Schülerinnen.

Es ist inzwischen mehr als 40 Jahre her, dass Karlheinz Ingenkamp seinen Sammelband „Die Fragwürdigkeit der Notengebung“ (1971) veröffentlicht hat, der bis 1995 in neun Auflagen erschienen ist. In der Folge gab es zwar einige Versuche mit Berichtszeugnissen und zeitweise konnte in Grundschulen auf Ziffernnoten sogar bis Klasse 4 verzichtet werden. Aber andernorts wurden die Bestimmungen dann auch wieder verschärft. Aus diesem Grund hat der Grundschulverband unsere Arbeitsgruppe Primarstufe vor fast zehn Jahren erneut mit einer Expertise beauftragt, die wir dann 2006 unter dem Titel „Sind Noten nötig und nützlich?“ vorgelegt haben. Die Ergebnisse decken sich mit denen von Ingenkamp: Wie vielfältige Studien zeigen, sind Noten entgegen den oft unterstellten Ansprüchen

- nicht valide: in eine Beurteilung gehen immer fachfremde Merkmale mit ein;
- nicht objektiv: dieselbe Leistung wird von verschiedenen Lehrpersonen sehr unterschiedlich beurteilt;
- nicht verlässlich: dieselbe Leistung wird von denselben Lehrer/inne/n zu verschiedenen Zeitpunkten unterschiedlich bewertet;
- nicht vergleichbar: für dieselbe Leistung werden in verschiedenen Klassen je nach deren Niveau unterschiedliche Noten vergeben;
- nicht informativ: differenzierte Leistungsprofile schrumpfen auf eine pauschale Ziffer.

Insofern ist die oft beschworene Erwartung, durch Noten wisse man, wo ein Kind stehe, eine Chimäre. Erst recht gilt das für die Prognose der weiteren Schullaufbahn. Tests können diese Probleme nur zum Teil verringern, haben sie doch durch die Standardisierung der Aufgaben sowie ihrer Auswertung und durch die nur punktuellen Erhebungen andere Schwächen. Insofern kommt man um das fachliche Urteil von Personen nicht herum. Dessen Grenzen können in Verbalgutachten am ehesten sichtbar und damit diskutierbar machen. Aus den internationalen Ländervergleichen wie auch aus Schulversuchen in Deutschland konnten wir überdies berichten:

- Der Verzicht auf Noten bringt keinen Leistungsabfall mit sich.

- Lehrer/innen und Eltern, die eigene Erfahrungen mit alternativen Beurteilungsformen haben, stehen dem Verzicht auf Noten deutlich positiver gegenüber als die immer noch skeptische Mehrheit der Bevölkerung.

- Einziger Grund für ein Beibehalten der Noten in der Grundschule: der teilweise noch bestehende Selektionszwang nach Klasse 4, der ein vergleichendes Ranking fordert.

Pädagogisch stellt vor allem die vergleichende Perspektive („soziale Bezugsnorm“) ein Problem dar. Schon 1968(!) hatte die KMK gefordert, Leistungen mit Bezug auf die Lernziele zu bewerten („kriteriale Bezugsnorm“). Überdies verlangen die Richtlinien für die Grundschule in vielen Bundesländern, den individuellen Lernfortschritt in die Beurteilung einzubeziehen. Die Praxis aber sieht anders aus.

Obwohl Noten angesichts der großen Heterogenität der Voraussetzungen am Schulanfang auch nicht fair sind. Denn wie eingangs erwähnt liegen schon Schulfänger drei bis vier Entwicklungsjahre auseinander: Der eine hat in der Sozialkompetenz erheblichen Nachholbedarf, die andere ist in ihren Erfahrungen mit Schrift den anderen weit voraus, ein drittes Kind kennt noch keine Zahlen. Die notwendige Öffnung des Unterrichts muss Konsequenzen haben auch für die Lernbeobachtung und Leistungsbewertung. Das galt schon immer, aber in der „inkluisiven Schule“ streuen die Unterschiede noch breiter – und ihre höhere Sichtbarkeit erschwert ein Ausweichen vor dem Anspruch, „jedem das Seine“ statt „allen dasselbe“ zu ermöglichen.

### **JedeR kann etwas**

Dritte Klasse, Dorfschule, Mitte der fünfziger Jahre: „Wie heißt der Spaßmacher im Zirkus?“, fragt Herr Dehmlow, unser neuer Lehrer, der uns bis zum Ende der vierten Klasse begleiten sollte. Volker sagt: „Klaun“. Ich, Leserratte, spreche es so aus: „Klowen“. Unser Lehrer kommentiert nicht. Wir sollen beides an die Tafel schreiben. Volker schreibt konsequent: Klaun. Ich: Clown. Herr Dehmlow lobt uns beide. „Sehr gut. Von Volker haben wir gelernt, wie die Engländer das Wort aussprechen, und C-l-o-w-n schreiben sie es.“ Eines von vielen Beispielen dafür, wie Herr Dehmlow es immer schaffte, den Schülern Selbstbewusstsein zu geben und flexibel auf sie einzugehen.

(Hajek 2013, 59)

Gleichzeitig ist die Kritik ernst zu nehmen, die an Verbalgutachten und Entwicklungsberichten geübt wird: dass sie bloße Umschreibungen von Noten und dass die Formulierungen für Eltern und Kinder oft nicht verständlich seien oder Schwächen schön geredet würden.

## Tests als Ausweg?

In den letzten 10, 15 Jahren werden uns nun zunehmend standardisierte Tests als Alternative zum fehleranfälligen Lehrerurteil empfohlen (vgl. zum Folgenden Brügelmann 2015).

Als Vorteile von Tests sind vier Potenziale zu nennen, die dieses Instrument auszeichnen:

- die Fokussierung der Datenerhebung auf vorweg begründete Leistungsaspekte;
- damit: die Transparenz der Anforderungen und Bewertung;
- die (klassenübergreifende) Kalibrierung der Maßstäbe durch Bezug auf Normstichproben;
- das Aufdecken blinder Flecken in Unterricht und Beurteilung (in einzelnen Fällen auch als Modell für Lern- oder Prüfformate, vgl. die Ablösung des lauten Vorlesens durch das Beantworten von Fragen zum Text).

Insofern sind standardisierte Tests stärker als früher üblich in das Repertoire schulischer Leistungsbeurteilung einzubeziehen. Allerdings ist ihr Status zu relativieren.

Denn die richtige Antwort zu einer Aufgabe kann aus verschiedenen Gründen gewählt worden sein, z. B.

- weil Schüler/innen wie erwartet das von den Test-Autor/inn/en avisierte Wissen/ Können zur aktuellen Problemlösung genutzt haben, aber auch,
- weil sie diese oder eine ähnliche Aufgabe vor kurzem bearbeitet und die verfügbare Lösung nur abgerufen haben,
- weil sie dank ihrer allgemeinen Intelligenz durch Nachdenken einen eigenen Lösungsweg gefunden haben,
- weil sie unter Zeitdruck rasch noch ein Kreuz (zufällig richtig) gesetzt haben,
- weil sie die Lösung aus Hinweisen im Text und den verfügbaren Alternativen als plausibel erraten haben.

Umgekehrt gilt auch für Fehler, dass sie unterschiedliche Gründe haben können:

Unaufmerksamkeit, anderer Lösungsweg, ... Vor allem aber gibt es bessere und schlechtere Fehler – und falsche Ergebnisse, die auf ein höheres Kompetenzniveau verweisen als eine richtige Lösung aus falschem Grund. Ein Beispiel aus dem Anfangsunterricht, also einer vergleichsweise einfachen Aufgabe. Fünf Kinder haben geschrieben:

KINO

KINNO

KINO

KIENO

KINO

Bei der Auswertung eines standardisierten Tests würde herauskommen: zwei Kinder haben das Wort falsch geschrieben und drei richtig. Objektiv und reliabel, d. h. von beliebigen Tester/inne/n reproduzierbare Aus- und Bewertung. Aber ist sie auch valide?

Personen- und situationsunabhängig ist nur eine Oberflächen-deutung möglich. Um die Kompetenz der Kinder beurteilen zu können muss man aber wissen, wie sie sonst schreiben und wie sie vorher geschrieben haben, d. h. auf welcher Stufe der Schreibeentwicklung sie sich befinden. Aufgrund einer solchen – theoriegeleitet begleitenden - Lernbeobachtung könnten sich ganz andere Ein-schätzungen ergeben:

- Tom, das erste Kind, schreibt KINO (richtig), weil er in einer Straße wohnt, in der es beim täglichen Gang zum Kindergarten an einem Filmpalast vorbeikommt, auf dessen Schild ihn seine Mutter mehrfach hingewiesen hat. Auf ähnliche Weise hat er ein Dutzend Wörter gelernt, die er als Buchstabenfolge reproduzieren kann, ohne zu wissen, warum man z. B. KINO gerade mit diesen Buchstaben schreibt.
- Nora schreibt mit Hilfe einer Anlauttabelle (HUNT, RATFARA usw.). Sie hat das Wort nach ihrer Aussprache verschriftet – und Glück gehabt, dass diese Strategie im konkreten Fall zum richtigen Ergebnis führt. Auch sie schreibt also aus falschem Grunde richtig.
- Carl ist schon einen Schritt weiter als Nora, weil er weiß, dass es nicht reicht, laut für Laut zu verschriften: Man muss beispielsweise aufpassen, ob ein Selbstlaut „lang“ oder „kurz“ gesprochen wird, und dass dann „da etwas Besonderes ist“. Aber ob man einen Buchstaben verdoppelt – und ggf. welchen – das hat er noch nicht verstanden. Carl schreibt also aus halbrichtigem Grund falsch.
- Jule weiß dagegen schon: Wenn man ein langes /i:/ hört und das Wort nicht kennt, dann ist es fast immer richtig und deshalb klug <ie> zu schreiben. Diese statistisch begründete Faustregel hilft ihr Fehler zu vermeiden. Sie schreibt also aus richtigem Grund falsch.
- Leonie kennt diese Faustregel auch. Aber sie darüber hinaus gelernt, dass es Ausnahmen gibt wie MASCHINE, APFELSINE und eben KINO, die man sich als Einzelwörter merken muss. Nur sie schreibt aus richtigem Grund richtig.

Kinder können schon die Aufgabe unterschiedlich verstehen – und die Auswertung kann leicht verkennen, was sich hinter der Lösung verbirgt. Wir können den subjektiven Faktor also nicht ausschalten. Aber er muss transparent und verhandelbar werden – durch dialogische Formen der Rückmeldung und Beurteilung.

[...]