

Noten sind praktisch – und unprofessionell

Elementarwissen zur Noten-Kontroverse – und die (bekannte) professionelle Alternative.



Ob sie Freude oder Frustr erzeugen – die messtechnische Qualität von Noten ist durchwegs miserabel.

Fotos: Peter Larson

Anton Strittmatter, Leiter der Pädagogischen Arbeitsstelle LCH

Lebewesen brauchen für Lernprozesse Feedback. Sie achten auf Folgen ihres Verhaltens für sich selbst oder in der Umgebung und passen ihr Verhalten den Rückmeldungen an. Das ist die «Ureinheit von Lernen». Erfolgt keine Rückmeldung, wird begonnenes Verhalten – weil zwecklos – nicht mehr freiwillig gezeigt. Wahrgenommene Folgen (Zielerreichung, Belohnung, Tadel etc.) sowie das hohe oder tiefe Ansehen der rückmeldenden Instanz sind wichtige Faktoren für das Auslösen von Lernprozessen. Es steht also ausser Zweifel, dass Beurteilungen für die persönliche Lernsteuerung äusserst wichtig sind. So nötig Beurteilung für die Lernsteuerung ist, so wenig zwingend ist die Form von Noten. Milliarden Menschen lernen täglich hoch wirksam durch ganz andere Formen von Rückmeldungen.

Der Gebrauch formalisierter Kurzcodes – wie unsere Ziffernoten – tritt nur bei «industrialisierten» Beurteilungsvorgängen auf, wie etwa in der modernen öffentlichen Massenbeschulung.

Bei solchen Beurteilungsveranstaltungen verbinden sich Ansprüche der Ökonomie (routiniert handhabbar) mit Ansprüchen der Gerechtigkeit. Ziffernoten sind in der Tat die am einfachsten handhabbare Lösung. Ob sie das Vergleichbarkeits-Versprechen einlösen, lässt sich wie folgt überprüfen.

Grundansprüche an die Beurteilung

Seit Jahrzehnten wird die messtechnische Leistungsfähigkeit von Ziffernoten wissenschaftlich untersucht. Die Ergebnisse fallen durchwegs miserabel aus. Folgende Grundansprüche sind an den Wert jeder Beurteilung von Lernleistungen zu stellen:

- **Validität:** Was gemessen wird, entspricht dem, was man messen will.

Und das, was man messen will, ist eine anerkannt bedeutsame Fähigkeit (Kompetenz). Die Testaufgaben bilden diese Kompetenz ausreichend ab (Konstrukt-Validität), entsprechen dem Lehrplan (curriculare Validität) bzw. erlauben eine gute Bewährungsprognose für künftige Berufs- oder Studiumssituationen (prognostische Validität).

- **Objektivität:** Die Beurteilung ist nicht willkürlich. Mehrere unabhängige Beurteilende würden zum selben Urteil gelangen, gleich korrigieren und dieselbe Note setzen. Es fließen namentlich keine Vorurteile zur Person in die Bewertung ein.
- **Reliabilität:** Die Wiederholungszuverlässigkeit einer Messung ist hoch. Es handelt sich nicht um eine flüchtige Momentaufnahme. Die Lernenden würden bei einer Wiederholung des Tests wieder ungefähr die gleichen Resultate erzielen, weil sie die gemess-



An die Stelle der alten Ablassregeln muss eine offensive Förderung treten.

tierungsstufe des Nachbarkantons Basel-Stadt. Das ist bekannt und trivial: Die den Baselbietern leistungsmässig vergleichbaren Schüler sitzen – salopp gesagt – in Basel-Stadt allesamt im Gymnasium und nicht in Berufsschulen. Häufig ist eine Vergleichbarkeit nicht einmal innerhalb derselben Schule gegeben: Eine 4 bei Lehrer X entspricht einer 3 oder einer 5 bei Parallelklassen-Lehrerin Y.

Der Unfug mit den Durchschnitten

Ein fundamentaler Irrtum der klassischen Ziffernoten-Praxis ist die Annahme, die Abstände zwischen 3 und 4 seien gleich wie jene zwischen 4 und 5. Dies ist in Wirklichkeit fast nie der Fall. Der reale Leistungsabstand von Dreier- zu Vierer-Schülern ist wahrscheinlich weit grösser als von Fünfer- zu Sechser-Schülern. (Je nach Praxis einer Lehrperson auch umgekehrt.) In der Bevölkerung ist leider der Unterschied zwischen einer Zahl und einer Ziffer unbekannt, was zu krassen Fehldeutungen beim Lesen von Noten führt.

Wegen dieser Verwechslung begünstigen Ziffernoten die Anwendung primitiver, wirklichkeitsfremder Rechenverfahren. Fast alle Promotionsordnungen erlauben oder verlangen gar die Errechnung von Notendurchschnitten aus allen Prüfungsereignissen. Wie unsinnig dieses Verfahren selbst bei Vernachlässigung des Abstandsproblems ist, mögen zwei – etwas zugespitzte, aber nicht unwahrscheinliche – Fälle zeigen:

- Schülerin A hat folgendes Notenbild: 3 - 5 - 2 - 5 - 5 - 3 - 6. Dies gibt im Zeugnis nach Adam Riese mit Rundungsregeln eine klare Vier. Dies wäre in diesem Falle die grösstmögliche Falschaussage über Schülerin A. Denn sie ist alles andere als eine «Vierer-Schülerin»! Richtig wäre ein Zeugniseintrag etwa folgender Art: «Schülerin A weist grosse Leistungsschwankungen über das ganze Semester hin auf. Sie erbringt abwechslungsweise gute bis sehr gute und dann wieder ungenügende bis schlechte Leistungen.»
- Schüler B zeigt folgendes Notenbild: 3 - 2 - 4 - 5 - 6 - 5 - 5. Gäbe im Zeugnis eine Vier oder Viereinhalb. Tatsache ist aber, dass Schüler B im Zeitpunkt der Zeugniserstellung ein guter und recht stabiler Fünfer-Schüler ist.

sene Fähigkeit dauerhaft besitzen (also nicht als zufällige Tagesform) und der Test auch wirklich diesen Sachverhalt prüft.

- **Vergleichbarkeit:** Die Geprüften sollen in ihrer Leistungsfähigkeit mit anderen verglichen werden können. Die Note 5 sollte in allen parallelen Klassen oder Schulen möglichst dasselbe bedeuten.

Was sagen nun Untersuchungen über die Aussagekraft von Noten dazu?

- **Schlechte Validität:** Es ist oft unklar, was eine 3 oder eine 5 in Französisch, Geschichte oder Mathematik inhaltlich bedeutet, was ein Lernender, eine Schülerin wirklich kann und was nicht. Analysen über das Zustandekommen von Noten belegen, dass eine für Aussenstehende unbekanntes Zahl von Faktoren in die Note einfließt. Häufig ist keine der drei Validitäten nachweisbar. Die prognostische Validität, die Vorhersage von Erfolg etwa im Beruf oder in der folgenden Schul- und Studiumsstufe ist im besten Fall sehr kurzfristig gegeben, verwischt sich schon nach einem oder zwei Jahren. So gibt es etwa keinerlei Zusammenhang zwi-

schen Maturanoten und Studien-erfolg an Hochschulen.

- **Schlechte Objektivität:** Die unabhängige Nachkontrolle ergibt sogar bei Mathematikproben eine Differenz von zwei bis drei Notenpunkten. Beim Vergleich zwischen Schulnoten und den Resultaten in Tests (etwa TIMSS oder PISA) ergaben sich Differenzen von bis zu vier Notenpunkten – sogar dort, wo die Lehrpersonen die Tests zuvor als valide, also dasselbe messend, akzeptiert hatten. Bei gleichen Leistungen erzielen Knaben und Mädchen, Schüchterne und Selbstbewusste, Einheimische und Ausländer etc. chronisch unterschiedliche Noten.
- **Schlechte Reliabilität:** Die Wiederholungszuverlässigkeit bei «hausgemachten», also nicht aufgrund von Kompetenzmodellen erstellten Prüfungen ist mässig bis schlecht. Vor allem dort, wo «Teaching for the Test» praktiziert wird, führen Prüfungswiederholungen regelmässig zu viel schlechteren Resultaten.
- **Schlechte Vergleichbarkeit:** Sekundarschüler aus dem Kanton Baselland schneiden später an der Berufsschule besser ab als Absolventen der Orien-

So zwingend nötig Beurteilungen für die Lernsteuerung sind, so wenig zwingend ist die Form von Noten. Milliarden Menschen lernen täglich hoch wirksam durch ganz andere Formen von Rückmeldungen.

Die Errechnung von Notendurchschnitten mag Gerechtigkeits-, Ansporn- oder Bestrafungsmotive haben. Sie führt aber in allzu vielen Fällen zu Leistungsbehauptungen, die keiner Wirklichkeit entsprechen. Vielmehr leisten sie Fehlprognosen und Demotivation Vorschub. Aus pädagogischer Sicht müssten an Stelle von Durchschnitten Beschreibungen des Leistungsverlaufs treten. Damit könnten Lernende, Eltern und Abnehmer mehr anfangen.

Der Unfug mit dem Ablasswesen

Die meisten heutigen Promotionsordnungen gehorchen den Regeln des vorreformatorischen Ablasshandels: «Du darfst Anforderungen missachten oder nicht erfüllen, sofern du dafür entweder mit einer schlechten Note bezahlst oder auf einem anderen Gebiet eine kompensierende gute Note erlangst. In beiden Fällen lassen wir dich dann mit deinem Nichtkönnen in Ruhe.»

Das ist das pure Gegenteil einer leistungsorientierten Schule, welche doch auf das Erreichen wichtiger Kompetenzen insistieren müsste. Dieses durch Ziffernoten begünstigte Ablasswesen behindert eine Politik des sogenannten «Zielerreichenden Lehrens und Lernens» (Mastery Learning, Pédagogie de la Maîtrise) und dürfte mit ein Grund für das mässige Leistungsniveau im internationalen Vergleich sein.

Untauglich aber erfolgreich

Das alles weiss man schon lange. Dennoch ist die Notenpraxis ungeachtet aller schlimmen Nachteile und ungerührt von allen bisherigen Reformversuchen eine grosse Erfolgsgeschichte. Die jüngsten Beispiele von Vorstössen in mehreren Kantonsparlamenten zur Wiedereinführung von Noten selbst auf der Primarstufe mögen das belegen. Der Erfolg dieses in professioneller pädagogischer Sicht heute klar untauglichen Modells liegt in verschiedenen Vorzügen begründet:

Das Notensystem ist eine einfache Konstruktion. Fünf oder sechs Ziffern,

denen ein volkstümlich leicht verständlicher Code zugeordnet werden kann: schlecht, ungenügend, genügend, gut, sehr gut.

Die Zuordnung von Ziffern zu Leistungen ist ein ökonomisches, massentaugliches Verfahren. Wortberichte für ganze Klassen oder gar viele Klassen (z.B. im Gymnasium) schreiben kostet ein Mehrfaches an Zeit und wird überdies von Zeugnislesern nicht honoriert, die eine «Kurzmitteilung auf einen Blick» bevorzugen und meist nur ein Interesse haben: «Ist mein Sohn/meine Tochter/dieser Lehrstellenbewerber genügend, ungenügend oder gut?»

Noten geniessen hohes Vertrauen. Sie erwecken – als Masszahlen gelesen – den Eindruck von Objektivität und Exaktheit. In dieser Logik erhalten Noten gar umso mehr Vertrauen, je mehr Stellen hinter dem Komma ausgewiesen werden. Wortberichte laden viel stärker zum Feilschen ein als die doch so exakten Zahlen... In Verbindung mit Angaben über den Klassendurchschnitt sowie als Zusammenrechnen aller Teilleistungen über die Periode hinweg werden Noten als ziemlich gerechte Honorierungsart angesehen – trotz des Wissens um Zufälligkeiten, Willkür etc. Folgerung daraus: Jedes Ersatzsystem muss auch Ersatz für diese unbestreitbaren Vorteile anbieten. Sonst wird es abgelehnt – von den Lehrpersonen wie auch von den Lernenden und Abnehmern. Alles wird zurückgewiesen, was kompliziert ist, zu aufwändig erscheint, wenig Vertrauen geniesst und Zweifel an der Gerechtigkeit von Urteilen weckt.

Halbbatztige Alternativen scheitern

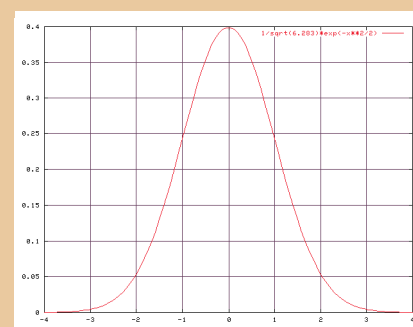
Seit den späten siebziger Jahren wurden immer wieder Alternativen zur Notengebung erprobt oder gar eingeführt. Mit mässigem Erfolg. Die Ansätze waren teils zu aufwändig, teils zwar verfahrensmässig durchdacht, aber inhaltlich schwach abgestützt (keine anerkannte Kompetenzenbeschreibung). Oder es misslang, die Abnehmersysteme (folgende Schulstufen, Eltern, Lehrbetriebe)

Der Unfug mit der Gauss-Kurve

Viele Lehrpersonen orientieren sich bei ihrer Notenverteilung an Berechnungsmodellen, denen mehr oder weniger das Konzept der Normalverteilung (Gauss-Kurve) zu Grunde gelegt wird. Nun ist es wichtig zu wissen, wie der Mathematiker Johann Carl Friedrich Gauss (1777–1855) das Normalverteilungs-Gesetz entdeckte: nämlich beim Qualitäts-sortieren von Getreidekörnern aus unbehandelten, gewissermassen naturwüchsigen Getreidefeldern.

Folgerung: Es ist eine Schande für den «Bauern» (Lehrer), wenn bei «behandeltem Saatgut», nämlich für viel Geld beschulten Schülerinnen und Schülern, dieselbe Leistungsverteilung herauskommt wie bei unbehandelten! Zur Ehrenrettung der Zunft: Natürlich ist das selten der Fall; das Problem ist vielmehr die Verfälschung der tatsächlichen Leistungsverteilung durch gewaltsames Überstülpen der Gauss-Kurve. Erfolgreiche pädagogische Einflussnahme müsste konsequenterweise eine rechtsgipflige Verteilkurve ergeben.

A.S.



zu überzeugen. Fast durchwegs wurden gute Ansätze nach dem seriösen und motivierenden Anfangsimpuls nicht weiter gepflegt.

Im Deffekt vereinten sich häufig die Stimmen der unzufriedenen Lehrkräfte mit den Stimmen der unzufriedenen Abnehmer zu einer Retro-Welle, welche in der Wiedereinführung der klassischen Notengebung die Lösung aller Probleme wähnte – und die früheren schlechten Erfahrungen mit Noten flugs aus dem Gedächtnis tilgte.

Fortsetzung Seite 10

Die Rückkehr zu den alten Ziffernoten löst kein einziges unserer heutigen Probleme mit den Resultaten der Schulbildung.

Professionelle Alternative bekannt

Die Fachwelt ist sich in den Grundzügen einig: Ein leistungsfähiges Bildungswesen muss mit anspruchsvollen Ziel- und Beurteilungssystemen arbeiten. Herkömmliche Benotungsweisen sind längst ausserhalb dieser Diskussion. In Schweden, beispielsweise, das in den PISA-Tests gut abschneidet, beginnen Ziffernoten (in Kombination mit anderen Verfahren) erst ab dem 8. Schuljahr.

Das Modell der Zukunft – zumindest in Ansätzen in Ländern wie Schweden, Finnland, Kanada oder Belgien verwirklicht – weist folgende Koordinaten auf:

- *Kompetenzmodelle als Basis:* Die Schulen verfügen in allen Fächern über ein einheitliches Sprachsystem zur Beschreibung bedeutsamer Fähigkeiten (Kompetenzen). Als Beispiel kann der europäische Referenzrahmen (mit zugehörigem Portfolio) für das Erlernen von Fremdsprachen gelten. Zu den Kompetenzen werden verschiedene Niveaus des Erreichens beschrieben.
- *Standards:* Anhand dieser Kompetenzbeschreibungen wird den Lernenden klar gemacht, was sie schon können, welches nächste Ziele sind, wo sie die Mindeststandards erreichen und wo nicht. Die Standards sind landesweit einheitlich.
- *Illustrative Aufgaben:* Die Lehrpersonen verfügen zu den Kompetenzbeschreibungen über Aufgabensammlungen, welche eine gute Diagnose erlauben.
- *Eichungsangebote:* Lehrpersonen können den Stand ihrer Klasse mit dem Stand ähnlicher Klassen (Zusammensetzung der Schülerschaft, Schulform) vergleichen und so ihr Urteil «eichen». Solche Ansätze gibt es bereits, etwa das St.Galler «Klassencockpit» oder sogenannte «Orientierungsarbeiten».
- *Halbjährliche Standortgespräche:* Die Lernenden und ihre Eltern werden halbjährlich über den Leistungs- und Entwicklungsstand orientiert. Wo hat

der/die Lernende seine/ihre Stärken? Wo muss noch zugelegt werden? Wie sieht die Entwicklung des Kindes/Jugendlichen aus und welche Laufbahn zeichnet sich ab, falls das so bleibt? Diese Gespräche werden nach bewährten Standards geführt und periodisch in ihrer Güte evaluiert.

- *Ergänzende «Noten» für die wichtigsten Kompetenzbereiche:* Weil viele Schulpartner ein aufwändiges Interpretieren von Portfolios nicht leisten können, werden diese durch Kurzcodes ergänzt, z.B. die Kompetenz ist «sehr gut», «gut», «genügend», «ungenügend» erreicht. Die meisten fortgeschrittenen Länder verwenden einen Buchstabencode: A, B, C, D (E). Wer behauptet, das sei nur Augenwischerei und könne ebenso gut durch die Ziffern 1, 2, 3, 4 etc. ausgedrückt werden, vergisst die geschilderte Problematik der Verwechslung von Ziffern und Zahlen!
- *Offensives Reagieren auf Defizite:* An Stelle der heutigen «Ablassregeln» tritt eine Förderpolitik für Lernende, bei denen das Erreichen der Mindeststandards gefährdet ist. Sie werden nicht mit einer schlechten Note entlassen, sondern kommen rechtzeitig in ein Förderprogramm, z.B. mit Zusatzunterricht. Nötigenfalls geschieht dies auf Kosten anderer Lernbereiche, in denen keine Mindeststandards gefährdet sind.

Erfolgsbedingungen

Für alle genannten Ansätze gibt es Beispiele, auch in der Schweiz. Nur wurden sie nie in dieser Kombination praktiziert – und wir sind weit weg von Arbeitsbedingungen, die eine solche Praxis seriös ermöglichen. Wer eine Leistungsschule mit guten Resultaten in Form von nachhaltig gelernten Fähigkeiten will, muss den Preis dafür kennen und bezahlen wollen. Dieser kann heute verlässlich beziffert werden:

Es braucht Lehrpläne mit fachlich hochstehenden Kompetenzmodellen und damit koordinierte Mindeststandards, Lehrmittel, Aufgabensammlun-

gen und Eichungstests. Das kostet viele Millionen und kann und muss nicht kantonale finanziert werden, sondern braucht die Leistungsfähigkeit der interkantonalen Kräfte- und Finanzienzusammenlegung. Die Lehrpersonen brauchen eine permanente intensive Schulung in der Handhabung dieser Instrumente, in fachdidaktischer Förderdiagnostik und in der Führung von Standort- und Perspektivengesprächen mit den Lernenden und Eltern.

Die Schulen brauchen zusätzliche Ressourcen für gezielte Förderarbeit bei Lernenden, welche die Mindeststandards im Klassenunterricht nicht erreichen können. Konkret: Es braucht zusätzliches Förderpersonal. Die Schweiz liegt heute mit nur 1,2 Lehrpersonen pro Primarklasse am unteren Ende vergleichbarer Länder, welche ein teils doppelt so gutes Betreuungsverhältnis ausweisen. Die Lehrkräfte müssen einen weit höheren Teil ihrer Arbeitszeit für die Förderdiagnostik, für differenzierte Beurteilungen und Gespräche einsetzen können. Bei den – im internationalen Vergleich höchsten – Pflichtlektionenzahlen von 28–30 geht das unmöglich. In den PISA-Spitzenländern liegen die Pflichtlektionenzahlen eher bei 20.

Die Handhabung von Kompetenzmodellen und Portfolios erträgt keine Lehrkräfte, welche das gleichzeitig in sechs oder zehn verschiedenen Fächern tun müssen. Das Fach- bzw. Fächergruppen-System der Ausbildung und des Einsatzes von Lehrpersonen muss rasch vorangetrieben werden. Die Unterstützung und Steuerung einer solchen Qualitätspolitik erfordert voll ausgebaute Schulleitungen.

Die Rückkehr zu den alten Ziffernoten löst kein einziges unserer heutigen Probleme mit den Resultaten der Schulbildung. Das Herumbasteln an seitenlangen Beurteilungslisten vom Typ Tugendkatalog auch nicht. Wer wirklich eine leistungsfähigere Schule will, muss in das zu investieren bereit sein, was heute als erfolgbringend bekannt ist. Oder das Maul halten und sich mit Mittelmässigkeit abfinden. Die Lehrerschaft ist nicht länger bereit, den Schwarzen Peter für das chronische Misslingen halbbatzig durchgeführter Reformen zu übernehmen.

Weiter im Text

http://home.t-online.de/home/c.ammann/paed/sub_inhalt_leistung.html