

Jürgen Oelkers

Leistungsbeurteilung und Notengebung im Gymnasium ^{)}*

1. Radikale Notenkritik

„Misstraut allen Noten!“ Dieser Ausruf verbunden mit einem Plädoyer zur Abschaffung der Noten ist am 13. Juli 2006 in der deutschen Wochenzeitung *Die Zeit* veröffentlicht worden. Verfasser war der Grundschulpädagoge Hans Brügelmann, der damals an der Universität Siegen lehrte. Er sprach sich vehement gegen eine These aus, die zwei Wochen zuvor an gleicher Stelle erschienen war. Dort hatte der Journalist Jörg Lau (2006) Ziffernoten damit gerechtfertigt, dass die Schülerinnen und Schüler wissen wollen, wo sie stehen und deswegen Noten befürworten.

Dagegen wandte sich Brügelmann mit folgendem Argument:

„Sagt eine 3 in Deutsch dem Schüler wirklich, wo er steht? Sie signalisiert ihm doch allenfalls, dass er etwa im Durchschnitt der Klasse liegt. Aber zeigt sie, dass er in geübten Diktaten nur wenige, in freien Texten dagegen viele Rechtschreibfehler hat, dass er flüssig vorlesen kann, aber Informationen aus Sachtexten nur ungenügend versteht? In der Note verdunstet die Vielfalt des individuellen Leistungsspektrums, der Rest kommt in eine Schublade“ (Brügelmann 2006).

Die Kantonsschule Frauenfeld verleiht jährlich anlässlich der Abschlussfeier den Jimmy-Bauer-Preis.¹ Der Preis wird an die bestbenoteten Schulabgänger verliehen. Die Bestnote berechnet sich aus dem Durchschnitt aller Noten in der Maturitätsprüfung. In diesem Jahr fand die Abschlussfeier am 24. Juni statt. Den Preis erhielten zwei Schülerinnen, die eine erzielte einen Notendurchschnitt von 5,84 und die andere von 5,65.² Der Notendurchschnitt war der Indikator für die Leistungsstärke.

Der Rektor der Kantonsschule verkündete die jeweilige Bestnote und verlas danach auch die Einzelnoten. Das Publikum, Eltern und Verwandte der Maturanden, Freunde, zahlreiche Schüler und viele Lehrer, reagierte auf das Ergebnis mit hörbarer Begeisterung. Es gab laute Ahs und Ohs, Bravo-Rufe und langanhaltenden Beifall für die Glanzresultate. Offensichtlich war niemand im Publikum der Meinung, dass den Noten grundsätzlich misstraut werden sollte. Das Gegenteil ist anzunehmen, Schulnoten sind anerkannte Beschreibungssysteme, die in der Gesellschaft fest verankert sind.

Niemand im Publikum störte sich auch an der Ziffernote. Gerade die Bestnote wurde als präzise Beschreibung des Leistungsverhaltens angesehen. Wer sich anstrengt und begibt

^{*)} Vortrag im Gymnasium Neufeld Bern am 24. August 2016.

¹ Jakob „Jimmy“ Bauer aus Hemmerswil bei Amriswil bestand 1893 die Maturitätsprüfung in Frauenfeld. Er war als Lehrer in London tätig und starb dort 1950. Vor seinem Tod hat er die Preisstiftung für seine alte Schule eingerichtet.

² Frauenfelder Woche vom 30. Juni 2016, S. 9.

ist, kann zu den Besten gehören und das spiegelt sich in den Noten. Die beiden Preisträgerinnen wurden also vielfach beglückwünscht und ihre Leistung erhielt grosse Anerkennung, ohne mehr zu kennen als die Zahl. Aber das genügte vollauf.

Auch unter den Schülerinnen und Schülern eines Gymnasiums sind vergleichend benotete Leistungsunterschiede nicht strittig, zumal sie ja wissen, wie die Leistungen zustande kommen. Sie wissen auch, dass die Anstrengungen je nach Lage verschieden sind, zwischen Jungen und Mädchen Unterschiede bestehen und nicht erreichte Leistungen durchaus hätten erreicht werden können, wenn die Anstrengung grösser gewesen wäre. Die Zuschreibung „Streber“ ist leicht einmal ein Indikator für eigenen Minimalismus.

Von dieser pragmatischen Sichtweise sind pädagogische Diskussionen weit entfernt. Insbesondere die Ziffernoten sind immer wieder als unzureichend oder gar schädlich kritisiert worden. Damit zusammenhängend ist die Kompetenz der Lehrerinnen und Lehrer angezweifelt worden und steht der Vorwurf im Raum, Schulnoten seien nicht objektiv, wobei nicht klar ist, was genau „objektiv“ heissen soll und ob Urteile anders als „subjektiv“ sein können.

Vor zehn Jahren hat Brügelmann sich auf eine „Vielzahl empirischer Studien“ berufen, die er an der Universität Siegen ausgewertet hat.

„Sie zeigen, dass verschiedene Lehrer Leistungen nach ganz unterschiedlichen Kriterien und anhand unterschiedlicher Massstäbe bewerten. Gibt man dieselben Arbeiten einer grösseren Zahl von Lehrern und Lehrerinnen zur Bewertung, streuen die Noten über die ganze Skala von 1 bis 6. Und das nicht nur im Aufsatz, sondern auch in Rechtschreibung und in Mathematik. Für den einen ist der Lösungsweg wichtig, für den anderen zählt nur das richtige Ergebnis. Merkmale der Person wie soziale Herkunft und ethnische Zugehörigkeit, Geschlecht und Sprachgewandtheit führen sogar zu systematischen Verzerrungen“ (ebd.).

Inzwischen ist auch ein Zusammenhang zwischen dem *Aussehen* der Schüler und der *Notengebung* konstruiert worden, der nochmals mehr zeigen soll, wie subjektiv und unfair die Noten zustande kommen. Für das Aussehen kann niemand etwas, wenn die Leistungsbeurteilung davon beeinflusst wird, dann ist das besonders unfair (Dunkake/Kiechle/Klein/Rosar 2012).³

Brügelmann legte vor zehn Jahren in seiner Schlussfolgerung nahe, künftig auf Ziffernoten zu verzichten und stattdessen mit Lernberichten zu arbeiten. Lernberichte sollen konkret beschreiben und damit erkennbar machen, wo genau die Stärken und Schwächen in einem Lernbereich liegen und vor allem, wie sich die Leistungen über einen bestimmten Zeitraum entwickeln (ebd.).

- Damit wird der Lernprozess der einzelnen Schülerinnen und Schüler in den Blick genommen und auf eine vergleichende Beurteilung verzichtet.
- Jeder lernt anders und jeder kommt voran, ohne ein Ziel zu erreichen, das für alle gleich gelten soll.
- Es gibt nur noch individuelle Ziele und keinen Vergleich der Leistungsstärke innerhalb einer Lerngruppe.

³ Allerdings ist die Datenbasis mehr als schmal, untersucht wurden drei Klassen eine grossstädtischen Gymnasiums in Deutschland. Wie zum Trost heisst es, ein „beauty is beastly-Effekt“ konnte nicht nachgewiesen werden. Damit ist gemeint, dass Schönheit hinderlich sein kann (Heilman/Saruwatari 1979).

Diese Idee hat seit den siebziger Jahren des vergangenen Jahrhunderts die Diskussion vor allem in der deutschen Grundschulpädagogik bestimmt. Wenig berücksichtigt wurden dabei die Praxis der Lernberichte, die Verständlichkeit und insbesondere die Akzeptanz bei Eltern und Schülern.

- Lernberichte verwenden eine bestimmte Sprache, oft einen Jargon, der mit Begriffen arbeitet, die für Eltern und Schüler häufig unverständlich sind.
- Die Vermeidung von impliziten Noten ist schwierig.
- Demgegenüber haben Ziffernoten den Vorteil, dass sie ad hoc verständlich sind, Akzeptanz in der Öffentlichkeit finden und ohne grosse Nachfragen kommuniziert werden können.

Auf der anderen Seite ist das „starre Ziffernotensystem“ der Lieblingsfeind vieler Schulreformer. „Motivieren ohne Noten“ war schon vor 25 Jahren ein immer wieder vorgebrachtes Stichwort der Schulkritik (Olechowski/Rieder 1990). Unterstellt wurde, dass die Schülerinnen und Schüler besser lernen, wenn sie *nicht* durch Noten geleitet werden und den eigenen Lernweg bestimmen können.

Motivieren *mit* Noten ist die vermutlich meist verbreitete Motivationspraxis in der Schulwirklichkeit, aber die wird als anrühlich hingestellt, da sie dem Ideal des „intrinsischen“ Lernens widerspricht. Noten als positive Anreize sind auch deswegen suspekt, weil sie eine Hierarchie voraussetzen. Nur wenige können Bestnoten erreichen und das fordert die anderen nicht etwa heraus, sondern schreckt sie ab und hindert sie am Lernen. Es soll, mit anderen Worten, keinen Wettbewerb geben und niemand soll mit anderen verglichen werden. Rousseau hat das 1762 in seinem Erziehungsroman *Emile ou de l'éducation* postuliert.

2. Befunde und Forderungen

Schon in der Kritik vor 25 Jahren waren verbale Beurteilungen oder Rückmeldungen an die Schüler eine Alternative, bei der allerdings unklar ist, warum sie die Ziffernoten ersetzen und nicht einfach ergänzen sollen. Die Schulkritik hat auch nie bedacht, warum Ziffernoten so erfolgreich sind und warum man sie nicht längst ersetzt hat, wenn doch mit so verheerenden Wirkungen verbunden sind.

Tatsächlich ist die Kritik am Notensystem seit Beginn des 20. Jahrhunderts und so der empirischen Bildungsforschung wieder vorgebracht worden (etwa schon Schreiber 1899), ohne dass sich an der Praxis sehr viel geändert hätte. Meistens ist dabei auf die fehlende Objektivität der Lehrpersonen und verzerrende Faktoren bei ihrer Beurteilung hingewiesen worden. Die Diskussion war periodisch sehr intensiv und im Ergebnis umso erfolgloser je mehr man die Anfangsjahre der Primarschule hinter sich lässt.

Und über Jahrzehnte sind die immer gleichen Argumente ausgetauscht worden, ohne gegenteilige Thesen zu prüfen:

- Die Notengebung sei rein subjektiv und letztlich beliebig,
- ihr fehle eine objektive Referenznorm,
- zudem sei sie von Sympathiewerten bestimmt

- und für die Beurteilung sei auch die „implizite Persönlichkeitstheorie“ der Lehrkräfte massgebend.⁴

Noten werden aufgrund des klasseninternen Bezugssystems, also der beobachteten und bewerteten Verteilung in einer bestimmten Klasse, gegeben, der Wert der Noten bezieht sich jedoch immer auf das gesamte Schulsystem und auf alle Berechtigungen. Vorausgesetzt wird die Vergleichbarkeit der Notengebung an allen Orten, während die tatsächlichen Bewertungen von Schule zu Schule und von Fach zu Fach variieren können. Das könnte man auch ohne grossen Alarmismus zur Kenntnis nehmen und an der Verbesserung des Systems arbeiten, ohne eine „Objektivität“ zu fordern oder in Aussicht zu stellen, die gar nicht erreichbar ist.

„Im Durchschnitt betrachtet“, hiess es in einem einflussreichen Forschungsbericht, können die Lehrkräfte „die Rangreihe der Leistungen innerhalb ihrer Klasse recht gut einschätzen“, auch wenn mit „erheblichen Unterschieden“ zwischen den Lehrkräften gerechnet werden muss (Weinert 2001, S. 50). Die klasseninterne Rangreihe der Leistungen, also die Normalverteilung, entspricht jedoch nicht immer den tatsächlichen Schülerleistungen, wenn man diese unabhängig testet.

Doch solche Tests sind nicht der Regelfall und sollten es auch nicht werden. Die Ersetzung der Lehrerurteile durch Leistungstests reduziert nicht nur die Kompetenz der Lehrkräfte, sondern führt auch dazu, dass der Unterricht auf den Test ausgerichtet wird. Vor allem amerikanische Erfahrungen zeigen, dass es dann nur einen Gewinner gibt, nämlich die Testindustrie (Ravitch 2010).

Noten beschreiben den internen Rangunterschied. Auch wenn die Lehrkräfte gut in der Lage sind, die Schüler ihrer eigenen Klasse gemäss ihren Leistungen zu rangieren, so heisst das nicht, „dass gleichen Noten in unterschiedlichen Klassen auch vergleichbare Leistungen zu Grunde liegen“ (Weinert 2001, S. 50). Eine Note kann also an einem Ort etwas Anderes bedeuten als an einem anderen. Ziffern wie „fünf“ oder „vier“ erfassen in verschiedene Schulen und Klassen nicht dasselbe, während genau das der Eindruck ist, den die Noten vermitteln. Sie werden ohne ihren Entstehungskontext kommuniziert.

Gefordert wird daher ein erweitertes Design der Leistungsbeurteilung, das mehr umfasst als nur die Lernziele des Fachunterrichts.

- Diese Dimension wird oft „kriteriale Norm“ genannt,
- die in vielen alternativen Modellen der Leistungsbeurteilung gleichrangig ergänzt wird durch eine „Sozialnorm“ und eine „Individualnorm“,
- also den Leistungsvergleich innerhalb der jeweiligen Lerngruppe und den individuellen Fortschritt, den die Leistung für den Schüler erbracht hat.

Das Problem ist die Gleichrangigkeit der beiden anderen Normen, was so für die Gymnasien nicht anwendbar ist. Im Mittelpunkt der Beurteilung und so der Notengebung steht der Fachunterricht, die besonderen Leistungsanforderungen des Gymnasiums und der prognostische Wert für das Studium. Noten, anders gesagt, müssen auch zur Schulform und ihren Aufgaben passen.

⁴ So schon Hofer (1975).

Instrumente, mit denen mehr beurteilt werden kann als nur die Leistungen, die im Blick auf fachliche Aufgaben gezeigt werden, sind inzwischen in der einen oder anderen Form vor allem in der Volksschule gebräuchlich, ohne auch hier in den höheren Klassen die Noten ersetzt zu haben. *Beurteilen* ist nach wie vor vor allem *Benoten*, und zwar Benoten von individuellen Leistungen in Unterrichtsfächern. Das gilt besonders für die Sekundarstufe II, wo sich bislang weder eine Sozial- noch eine Individualnorm durchgesetzt haben. Allerdings gibt es einen deutlichen Trend, auch Arbeitshaltungen und Lerneinstellungen zu bewerten, was die Frage der Referenznorm nochmals anders stellt.

Niemand bestreitet mögliche Negativeinflüsse bei der Notengebung und Urteile über Leistungen können immer verbessert werden, aber wenn allein die Fehlerquellen herausgestrichen werden, dann könnte man schliessen, es sei *unmöglich*, zu einer „gerechten Zensierung“ der Leistungen von Schülern zu gelangen und dies sei der Lehrerschaft auch durchaus bewusst (Döring 1925, S. 177). Das schrieb der deutsche Jugendpsychologe Otto Wilhelm Döring schon vor mehr als neunzig Jahren. Die tatsächlichen Mühen und Anstrengungen der Lehrpersonen, zu einem gerechten Urteil zu gelangen, bleiben bei solchen Verdikten regelmässig auf der Strecke.

Aber selbst wenn die Behauptung zuträfe, die Notengebung als solche sei ungerecht und könne auch gar nicht anders sein, so bliebe dennoch die Frage bestehen, warum dann immer noch weitgehend Ziffernnoten die Praxis der Leistungsbeurteilung bestimmen.

- Es ist mindestens frappierend, wie stark die offenkundigen Vorteile einer Notenskala die Praxis bestimmt haben.
- Systeme der Leistungsbeurteilung in der Schule müssen praktikabel sein,
- also den Aufwand begrenzen, zum Schultyp oder zur Schulstufe passen und auch die Erwartungen der Abnehmer erfüllen.

Aber Noten und Zeugnisse, die die Schule vergibt und die mit Berechtigungen verbunden sind, haben in der pädagogischen Diskussion keinen guten Ruf. Und mittlerweile scheint festzustehen, dass Noten ohne Bezugsnormen subjektiv sind, „subjektiv“ als Synonym mit „beliebig“ anzusehen ist und Zeugnisse wenig aussagen.

Auf der anderen Seite ist die Beschreibung schulischer Leistungen mit einer Skala von Noten ein bewährtes Instrument, das wie gesagt in der Gesellschaft verankert ist und zu den stärksten Erwartungen an die allgemeinbildende Schule gehört. Selbst wenn man Noten umgehen will, sind sie präsent, weil unterschiedliche Leistungen im Blick auf identische Aufgaben nicht übersehen werden können.

Die Forschung hat sich über Jahrzehnte auf die Lehrkräfte und die Fehlerquellen ihrer Urteile konzentriert. Das ist erst seit einigen Jahren als Verengung wahrgenommen worden, weil damit gar nicht in den Blick kam, wie die Eltern oder die Schüler über Noten denken und was sie von ihnen erwarten.

- Bei Befragungen der Schülerinnen und Schüler wird deutlich, dass Kinder Noten wollen und wenig Zustimmung herrscht für deren Abschaffung.
- Leistungsstärkere Kinder bevorzugen ein Ziffernzeugnis, leistungsschwächere dagegen ein Verbalzeugnis.
- Ein Drittel der Befragten in einer Siegener Studie bevorzugt eine Kombination aus Verbalzeugnis und Ziffernzeugnis (Backhaus/Tzerni 2008).

Die Studie macht auch klar, dass Noten gesellschaftlich gewünscht werden. Eltern sind wohl an individuellen Informationen über die Entwicklung ihres Kindes interessiert, wünschen sich aber zugleich klare Rückmeldungen, wo ihr Kind im Bezugssystem der Schulklasse steht. Das wird festgestellt, aber nicht ernst genommen, denn danach folgt in der Studie die Notenkritik, auf die sich auch Brügelmann in dem ZEIT-Artikel bezieht. Sprichwörtlich wurde die „Not mit den Noten.“⁵

Kinder und Eltern, so die Kritik, haben sich einfach an das falsche System gewöhnt, Noten sind nicht aussagekräftig und sollten abgeschafft werden. Sonst ist in der deutschen Schule der Elternwille nahezu unantastbar, im Falle der Noten ist das anders, Eltern müssen belehrt werden, wenn sie sich für Noten einsetzen. Auch die Formel, in der Schule „vom Kinde aus“ zu denken, wird ausser Kraft gesetzt, wenn sich die Schüler entgegen der pädagogischen Erwartung für Noten aussprechen.

Die Notenkritik richtet sich primär auf das Zustandekommen der Noten und die Beschreibung in Form von Ziffern. Wenig gefragt ist der prognostische Wert von Schulnoten. Inzwischen liegen zahlreiche empirische Studien vor, die die gute prognostische Validität der Schulnoten für den späteren Studienerfolg belegen. Die Abiturdurchschnittsnote gilt als der valideste Einzelprädiktor für den Erfolg im anschliessenden Studium. Die Notenkritik übersieht diesen Zusammenhang.

In einer Metastudie aus dem Jahre 2007 wird die Validität der Schulnoten zur Vorhersage des Studienerfolgs nochmals detailliert beschrieben (Trapmann-Hell/Weigand/Schuler 2007). Dabei wirkt sich gerade die Breite des Notensystems positiv aus. Je höher der Durchschnitt liegt, desto besser kann der Studienerfolg vorhergesagt werden. Ganz so falsch kann die Kantonsschule Frauenfeld mit den Bestnoten also nicht gelegen haben.

- Im Ergebnis wird festgehalten: „Besonders deutsche Durchschnittsnoten erreichen hohe Validitäten, die auch von studienfachnahen Schulnoten nicht übertroffen werden“ (ebd. S. 25).
- Gesagt wird das in einem internationalen Vergleich.
- Für die Schweiz würden dann vermutlich noch höhere Validitäten gelten, weil das Fächerspektrum der Maturität erheblich breiter ist als im deutschen Abitur.

Eine der zentralen Begründungen für die Notenkritik bezieht sich auf die Folgen von schlechten Bewertungen, die blockierte Motivation und Schulunlust nach sich ziehen würden. Gute Noten werden akzeptiert, schlechte lösen Krisen aus. Das Prinzip der vergleichenden Graduierung in der Beschreibung des Leistungsverhaltens setzt wie gesagt voraus, dass nicht jeder gute Noten erhält und man so einen Diskriminierungseffekt auffangen muss, wenn man auf ein Ziffernotensystem setzt.

Eine jüngst erschienene Studie über den Zusammenhang von Schulangst, Schulunlust, Anstrengungsvermeidung und Schulnoten in den Fächern Mathematik und Deutsch sieht zwischen diesen Konzepten signifikante geringe bis mittlere Interkorrelationen. Am stärksten hängen Prüfungsangst und die schulbezogene Anstrengungsvermeidung mit den Schulnoten zusammen (Weber/Petermann 2016, S. 562). Das ist eigentlich trivial: Wer Angst vor

⁵ Die „Not mit den Noten“ ist ein viel zitierter Titel in den Medien, der schon mehr als dreissig Jahre zuvor eingängig war (Trost 1976).

Prüfungen hat, vermeidet Anstrengungen, weil die Vorstellung vorherrscht, die Prüfung sei ohnehin nicht zu bestehen. Andererseits minimiert die Anstrengungsvermeidung die Chancen des Bestehens, wenn die Prüfung nicht vermieden werden kann.

Generell wirken sich Prüfungsangst, manifeste Angst sowie Anstrengungsvermeidung negativ aus. „Dabei wird der Einfluss von Prüfungsangst auf die Mathematiknote und der Einfluss manifester Angst auf die Deutschnote bei Jungen vollständig über die schulbezogene Anstrengungsvermeidung erklärt. Bei Mädchen stellen sich Prüfungsangst und Anstrengungsvermeidung als gleich bedeutende Prädiktoren für die Mathematiknote dar. Auch bei Jungen bilden Prüfungsangst und Anstrengungsvermeidung vergleichbare Prädiktoren für die Deutschnote ab. Leistungssteigernde Trainingsmassnahmen wie auch präventive Massnahmen sollten daher einerseits Schulängste und Anstrengungsvermeidung bei den Schülerinnen und Schülern verringern und gleichzeitig ihre Motivation und Lernfreude steigern“ (ebd. S. 564/565).

Aber das Problem verlangt nicht lediglich eine verhaltenstherapeutische Bearbeitung von psychischen Problemen.

- Vielmehr entsteht Prüfungsangst häufig auch durch intransparente Anforderungen und diffuse Leistungserwartungen.
- Zudem sind in Fächern wie Mathematik und Deutsch unterschiedliche Vorerfahrungen vorhanden, die Schlüsse auf die eigene Kompetenz oder Nicht-Kompetenz zulassen.
- Ähnlich entstehen Versagensängste durch diffuse Selbsteinschätzung und unklare Leistungsanforderungen.

Damit ist aber nicht gesagt, wie die Schülerinnen und Schülern mit Problemen von Prüfungen und Noten umgehen, wenn „Prüfungsangst“ nicht die Standardreaktion sein kann und auch nicht überall Intransparenz herrscht. Generell nehmen die vorliegenden Studien wenig Rücksicht auf die tatsächlichen Lernstrategien und -taktiken, die heutige Schülerinnen und Schüler an den Tag legen. Anforderungen kann man unterlaufen und das eigene Leistungsverhalten abschwächen, wenn das Ziel ohnehin erreicht wird.

3. *Die Notenskala in der Schweiz*

Es ist immer wieder versucht worden, Alternativen zu der Notenskala zu entwickeln. Radikale Entwürfe gehen wie gesagt davon aus, dass nur die Lernfortschritte des einzelnen Schülers beschrieben werden können und sich ein Vergleich in der Lerngruppe verbietet, weil der ohnehin nicht objektiv sein kann. Noten setzen die Klassennorm voraus und basieren so auf einem Vergleich der Leistungen mit anderen.

- Diese Beschreibung hat sich bewährt, sie ist ökonomisch und vergleichsweise leicht zu handhaben.
- Zudem zeigt die Forschung, dass die Urteile der Lehrpersonen im Blick auf ihre Klasse im Allgemeinen verlässlich sind.
- Die empirische Kritik bemängelt allerdings die fehlenden Bezugsnormen (Fischer 2012, S. 50).
- Das war ein Hauptgrund für die Entwicklung des Lehrplans 21.

Historisch kann von einem sehr stabilen Instrument ausgegangen werden. Eine gestufte Notenskala zur Beschreibung der Leistungen geht zurück auf den sogenannten Gothaer „Schulmethodus“. So bezeichnet man das historisch wohl erste umfassende Reglement zur Steuerung von Schule und Unterricht mit verbindlichen Verfahren und Leistungsüberprüfungen.

Der „Schulmethodus“, also der Leitfaden der Steuerung, ist von Andreas Reyher, dem Rektor des Gymnasiums in Gotha,⁶ verfasst worden und erschien zuerst 1642. Der Methodus war Teil der ernestinischen Reformen im Fürstentum Gotha, die als erste staatliche Schulentwicklung im deutschen Sprachraum gelten können.⁷ Seit der Ausgabe des Jahres 1662 schrieb der Methodus ein Stufenschema zur Beurteilung der Leistungen in drei Dimensionen vor. Der Leitfaden ist bis zum Ende des 17. Jahrhunderts mehrfach neu aufgelegt worden und hat die Organisation und Verfahren der Schule massiv verändert.

Die Beurteilung erfolgte für die jährlichen Schulexamen, die öffentlich waren und im Herbst vor der Ernte stattfanden. Die Lehrer waren der Gemeinde rechenschaftspflichtig und alle ihre Daten einschliesslich der Leistungen in den Unterrichtsfächern mussten mit einem einheitlichen „Modell“ erfasst und dokumentiert werden (Methodus 1697, S. 102). Auf diese Weise waren sie leicht verständlich und stellten die Grundlage dar für die Entwicklung im nächsten Schuljahr (ebd.).

Die drei Bereiche werden so gefasst: „Ingenium“ ist Begabung, „mores“ ist Verhalten und „Unterrichtsgegenstände“ sind die curricularen Lernfelder (ebd., S. 112). Es gab noch keine Ziffernnoten, die Stufen wurden mit Adjektiven bezeichnet.

Ingenium	Unterrichtsgegenstände	Mores
Sehr fein	fein	fromm
Gut	fertig	fleissig
Ziemlich	ziemlich	still
Schlecht	etwas/wenig	unfleissig
	schlecht	ungehorsam

Offenbar ist dieses Schema - zeitgemäss verändert - geeignet, Leistungs- und Verhaltensdifferenzen innerhalb einer bestimmten, grösseren Lerngruppe zu beschreiben. Die vergleichsweise intensive Forschung hat bislang keine Alternativen geliefert, die praktikabler wären. In diesem Sinne handelt es sich um einen Standard, den die Schule selbst hervorgebracht hat und der auch in Zukunft Verwendung finden wird.

Nun kann man fragen, wieso gerade Ziffernnoten so umstritten sind und mit grossem Eifer bekämpft werden. Eine Antwort ergibt sich aus der Stufung des Schulsystems. Ziffernnoten werden in der pädagogischen Literatur generell wenig wohlwollend betrachtet, aber besonders umstritten sind sie im Primarschulbereich. Damit reagiert die Literatur auf Vorstellungen kindlichen Lernens, die nicht von Wettbewerb und Leistungsdruck geprägt sein sollen.

⁶ Andreas Reyher (1601-1673) wurde 1641 als Rektor des *Gymnasiums illustre* nach Gotha berufen.

⁷ Ernst I., der Fromme (1601-1675), war seit 1640 Herzog von Sachsen-Gotha. Auf ihn gehen umfangreiche Bildungsreformen zurück, die von der Schulpflicht ab dem 5. Lebensjahr bis zur Errichtung eines Waisenhauses reichen.

Die dafür gewählte Sprache, die etwa von „Notenterror“ spricht,⁸ macht es schwer, in Noten etwas anderes zu sehen, als willkürliche Instrumente, die Kinder traumatisieren können. Darauf haben viele deutsche Grundschulen reagiert, die mit behördlicher Genehmigung „notenfrei“ ausgerichtet sind und damit den „Notenterror“ an die nächste Stufe weitergeben. Auch in der Schweiz scheint es zunehmend schwer zu sein, Noten für Primarschülerinnen und Primarschüler als geeignet anzusehen.

Im Kanton Zürich werden mit dem Ende des ersten Schuljahres Zeugnisse mit Ziffernnoten vergeben, was nach der Kritik eigentlich zu verheerenden Wirkungen führen müsste. Mit den Noten werden fachspezifische Leistungen im Blick auf die „angestrebten Lernziele während der Zeugnisperiode“ beurteilt. Noten sind nur für die Zeugnisse vorgeschrieben, „Lehrpersonen sind nicht verpflichtet, ausserhalb des Zeugnisses Noten zu erteilen“ (Beurteilung und Laufbahntscheide 2013, S. 16). Sie können, aber sie müssen nicht.

Nicht nur in historischer Hinsicht interessant ist die Beschreibung des Notensystems (ebd.). Seit 2007 sind für die Benotung der Unterrichtsfächer wie im „Schulmethodus“ fünf Stufen vorgesehen, nicht mehr, aber auch nicht weniger, vorgegeben sind Lernziele, die für alle Schüler gleich gelten und die Leistungen können mit dem System der Noten auf unterschiedlichen Stufen erfasst werden. Die von der Kritik so schmerzlich vermissten Bezugsgrössen sind einfach die Lernziele des Lehrplans.

Note	entspricht	Bedeutung bezüglich der Lernziele
6	sehr gut	Alle Lernziele sind erreicht; zusätzlich werden regelmässig besondere Leistungen erbracht.
5	gut	Alle Lernziele sind erreicht.
4	genügend	Die wesentlichen Lernziele sind erreicht.
3	ungenügend	Die wesentlichen Lernziele sind nur teilweise Erreicht.
< 3	schwach	Die wesentlichen Lernziele sind nicht oder nur zu einem sehr geringen Teil erreicht.

Geplant ist nicht, die Ziffernnoten abzuschaffen, was generell für alle Schweizer Kantone gilt, die sich am Lehrplan 21 beteiligen. Die schweizerische Erziehungsdirektorenkonferenz hat klargestellt, dass der neue Lehrplan die Form der Leistungsbeurteilung nicht definiert. Die Lehrpersonen werden also auch weiterhin die Lernziele setzen und deren Erreichen überprüfen.

Die EDK hat auch die unterschiedlichen Regelungen zur Leistungsbeurteilung in den Kantonen erhoben. Nur die Beurteilung im Kindergarten (HarmoS 1-2) erfolgt ohne Ziffernnoten, aber nicht ohne Beurteilung, für die meist vorgegebene Beobachtungsbögen eingesetzt werden. Damit werden Lernfortschritte beschrieben und Rückmeldungen an die Eltern dokumentiert. In der Primarschule und Sekundarschule I wird an der Praxis der Ziffernzeugnisse festgehalten, die aber nicht einheitlich geregelt ist und kantonale Unterschiede kennt.⁹

⁸ Der Ausdruck stammt aus den westdeutschen Schulkämpfen der siebziger Jahre („Gegen Notenterror“: Der Spiegel Nr. 2 von 9. 1. 1978, S. 64/65).

⁹ <http://www.edk.ch/dyn/15780.php>

- Primarstufe (HarmoS 3-8) und Sekundarstufe I: In allen Kantonen werden den Schülerinnen und Schülern Zeugnisse mit Noten ausgestellt, dies allerdings nicht in allen Kantonen ab dem gleichen Zeitpunkt.
- In der grossen Mehrheit der Kantone setzen Zeugnisse mit Noten zwischen dem 1. und dem 4. Schuljahr (HarmoS 3-6) ein.
- In den Schuljahren, in welchen noch keine Zeugnisse mit Noten ausgestellt werden, kommen Lernberichte ohne Noten oder andere Beurteilungsverfahren zur Anwendung.
- Lernberichte ohne Noten werden häufiger in den unteren Schuljahren ausgestellt.¹⁰

Im Gymnasium gibt es Zeugnisse praktisch überall nach dem ersten Semester.¹¹ Spielraum ist nur vorhanden bei der Leitungsbeurteilung bis hin zur Zeugnisnote. Deutlich wird auch, dass sich die Primarschule stärker differenziert hat als die beiden Sekundarstufen und die Sekundarstufe I eher mit neuen Formen der Leistungsbeschreibung experimentiert als die Sekundarstufe II.

Einen Einblick in die Prüfungspraxis der Schweizer Gymnasien hat die Studie EVAMAR I aus dem Jahre 2004 geliefert. Die Maturantinnen und Maturanten mussten angeben, welche Noten sie im letzten Zeugnis in den einzelnen Fächern erreicht haben. Aus diesen Angaben wurde für die Fächer ein Durchschnitt gebildet.

- Im Ergebnis ist der Notendurchschnitt in den Fächern Musik und Bildnerisches Gestalten „deutlich und statistisch gesichert“ höher als der Gesamtmittelwert,
- während Physik und Anwendungen der Mathematik sowie Wirtschaft und Recht tendenziell niedriger liegen.
- Alle anderen Fächer unterscheiden sich nicht statistisch gesichert vom Gesamtmittelwert (Evaluation der Maturitätsreform 2004, S. 132).

Diese „deutlichen“ Notenunterschiede werden mit „unterschiedlichen Beurteilungspraktiken“ in Verbindung gebracht (ebd.), was auch in der einschlägigen Forschungsliteratur beschrieben wird.

Selektive Hauptfächer wie Mathematik oder Sprachen vergeben Noten nach strengeren Massstäben als etwa die musischen Fächer. Die Noten in Mathematik streuen „deutlich stärker“ als die in den anderen Fächern, die Mathematiknoten erhalten daher eine erhöhte Bedeutung bei der Versetzung, gefolgt von der zweiten Landessprache. Vornehmlich an diesen beiden Fächern entscheiden sich der Verbleib in den Maturitätsschulen und mindestens der je erreichte individuelle Notendurchschnitt.

Drei Viertel derjenigen Schülerinnen und Schüler, die die Maturität nicht bestehen, haben eine ungenügende Note in Mathematik, zwei Drittel haben eine solche Note in der zweiten Landessprache. Von der Gesamtpopulation aller Maturandinnen und Maturanden hat fast ein Viertel eine ungenügende Note in Mathematik.

Mathematik und die zweite Landessprache sind die „gefährlichen Klippen“ auf dem Weg zur Maturität (ebd., S. 135). Es gibt zwischen Deutsch und Französisch als zweiter

¹⁰ Ebd.

¹¹ http://www.edudoc.ch/static/strukturdaten/pdf_rohdaten/167.pdf

Landessprache keinen Unterschied, solange beide Sprachen lediglich als Schulfächer¹² erfahren werden. Wer gute Leistungen in der Unterrichtssprache erzielt, hat in der Regel keinen Vorteil für den Erwerb der zweiten Landessprache.

Ein Transfereffekt, der von der viel zitierten „Sprachbegabung“ getragen sein würde, lässt sich nicht feststellen. Wer etwa Englisch als Schwerpunktfach besucht, erzielt tiefere Noten in der Unterrichtssprache und nochmals tiefe in der zweiten Landessprache. Das gilt allerdings nicht für Latein: Wer dieses Schwerpunktfach besucht, erzielt höhere Noten in der Unterrichtssprache und fast die gleichen Noten in der zweiten Landessprache wie in Latein (ebd., S. 133).

Diese Praxis dürfte sich in den vergangenen zehn Jahren nicht sehr weit verändert haben. Vor allem HarmoS und die der Lehrplan 21 haben die Notendiskussion neu entfacht, die Gymnasien sind davon insofern betroffen, als die Anschlüsse gesichert sein müssen, was schwierig sein dürfte, wenn auf der Sekundarstufe I ganz andere Leistungserwartungen bestehen als im Gymnasium und sich dann auch die Leistungsbewertung unterscheidet. Bereits heute unterscheidet sich der Leistungsstand bei Aufnahme in das Gymnasium erheblich.

Positiv an der Entwicklung ist der Tatbestand, dass das Zustandekommen der Schulnoten inzwischen zu einem Thema geworden ist und nach Verbesserungen gesucht wird. Bislang waren Noten Urteile der Lehrpersonen, deren Kriterien für Schüler wie Eltern weitgehend unklar waren. Weder die Art und Weise, wie die Leistungen beurteilt wurden noch die Berücksichtigung des Lernprozesses waren transparent. Noten waren auch nie wirklich Objekt von Schulentwicklung.

4. Die Suche nach Alternativen

Die Notenfrage ist inzwischen auch in den Schweizer Medien angelangt. So fragte der Schweizer Beobachter Anfang August diesen Jahres, ob eine „Schule ohne Noten“ „gut gehen“ kann. Die bange Frage wird mit John Hattie beantwortet: Es gibt in der Frage der Noten „keinen Königsweg“, entscheidend sei „die Kompetenz der Lehrperson, auf die individuellen Bedürfnisse der Schülerinnen und Schüler möglichst gut einzugehen“ (Beobachter Nr. 16 vom 5. August 2016, S. 3).

Der Bericht geht aber weiter und bezieht sich mit einem positiven Beispiel auf die Sekundarschule Niederhasli im Kanton Zürich,¹³ die mit dem Schuljahr 2013/2014 eine transparente Art der Notengebung eingeführt hat. Die Notengebung wird kombiniert mit selbstorganisiertem Lernen.

- Zu Beginn des Semesters legen die Schülerinnen und Schüler in den Kernfächern Mathematik, Deutsch, Englisch und Französisch ihre persönlichen Notenziele fest.
- Sie versuchen diese Ziele mit eigenem Lerntempo und Unterstützung durch die Lehrpersonen zu erreichen.

¹² Zu unterscheiden von längeren Sprachaufenthalten oder von der zweisprachigen Matur.

¹³ Die Schule hat den LISSA-Preis 2016 erhalten (<https://www.eduzis.ch/einblicke/unsere-schule/einblicke-lissa-preis-seehalde/>)

- Wie das geschieht, fließt in die Note ein.
- Auf „Kann-Listen“ haken sie den Stoff ab, den sie beherrschen.
- Das sorgfältige Führen dieser Listen bringt Punkte.

Weitere Punkte können sich die Schülerinnen und Schüler durch Reflexion, Planung und Fachberatung erarbeiten. Die Punkte machen zusammen bis zu 30% der Zeugnisnote aus. Die restlichen 70% bilden Leistungsnachweise durch Fachtests oder mündliche Prüfungen (Beobachter Nr. 16 vom 5. August 2016 S. 18).

Alternativen der Leistungsbeurteilung ohne Ziffernzeugnisse sind auch in verschiedenen deutschen Schulen durchgeführt worden. 2010 ist der Abschlussbericht der wissenschaftlichen Begleitforschung für vier dieser Schulen publiziert worden. Es handelt sich um Grundschulen, also nicht um weiterführende Schulen und auch nicht um Gymnasien.¹⁴

Die Schulen, heisst es in dem Bericht, machten durchgehend die Erfahrung, dass unter bestimmten Voraussetzungen auch ohne Noten motiviert und interessiert gearbeitet werden kann, dass es aber, wenn man die Ziffernnoten abschafft, andere Formen der Anerkennung von Leistungen geben muss, weil die Schüler tatsächlich wissen wollen, wo sie stehen (Bos u.a. 2010, S. 13).

Die untersuchten Schulen arbeiten mit Lernentwicklungsberichten (ebd. S. 23ff), in denen verschiedene Dimensionen des Sozialverhaltens sowie des Lern- und Arbeitsverhalten beschrieben werden. Das Ziffernzeugnis wird durch ein Pfeil- und Rasterzeugnis ersetzt. „Beide differenzieren die Fächer aus, indem sie am Lehrplan orientiert Kompetenzbereiche für die Fächer ausweisen. So wird eine spezifische Beschreibung von Lernfortschritten in Bezug auf einzelne Subdomänen möglich“ (ebd. S. 26).

Dieses Verfahren wird in einer der beteiligten Schulen wie folgt angewendet:

„Die Fächer sind hierbei unterteilt in die verschiedenen Kompetenzerwartungen des einzelnen Faches bezogen auf das vorangegangene Halbjahr. Für jede Kompetenzerwartung drückt ein Pfeil die Leistung aus. Das Pfeilzeugnis zeichnet aus, dass die Lern- und Leistungsrückmeldung durch Pfeile visualisiert wird. Durch den dazugehörigen Kommentar zur Leistungsentwicklung in jedem einzelnen Fach ist das Zeugnis individueller als ein reines Notenzeugnis.“

Weiter heisst es:

„Hinter der Pfeillänge steht eine Ziffernnote. Hier wird nach Punkten bewertet, so wie dies aus der Oberstufe bekannt ist. Für die Schülerinnen und Schüler ist diese aber nicht augenscheinlich und wird nur auf ausdrücklichen Wunsch den Eltern mitgeteilt, sodass zunächst nur die Länge des Pfeils Ausdruck der Leistung ist“ (ebd. S. 26).

Die Schüler sollen nur ihren Lernfortschritt erkennen, die Note soll ihnen erspart bleiben, nur die Eltern erfahren die Note, und dies auch nur, wenn sie einen Antrag stellen. Warum das von Vorteil sein soll, erschliesst sich nur vor dem Hintergrund der Motivationsideale des selbstorganisierten Lernens. Aber wenn die Eltern die Note erfahren,

¹⁴ Die vier beteiligten Schulen waren: Paul-Schneider-Schule (Grundschule in Münster), Wartburg-Grundschule (Gievenbeck/Münster), Rosenmaarschule (offene Ganztagsgrundschule in Köln-Höhenhaus) sowie Arnold-von-Wied -Schule (Grundschule in Bonn).

dürfte es schwer fallen, sie vor den Kindern zu verbergen und dann dürfte es auch nicht möglich sein, dass die Noten unter den Kindern kommuniziert wurden und alle Kinder wissen wollen, wo sie benotet stehen.

Ein weiteres Problem ist der gesteigerte Aufwand. Ein Rasterzeugnis etwa ist im Format DIN A3 sieben Seiten lang, enthält neben der Leistungsentwicklung auch Angaben über das Arbeits- und Sozialverhalten sowie weitere Informationen wie Absenzen und Entschuldigungen. Lesbarer sind die Zeugnisse dadurch nicht geworden, wohl aber umfangreicher und natürlich bilden sie weiterhin Graduierungen ab, hinter denen man auch Noten erkennen kann.

Für das Arbeits- und Sozialverhalten gibt es vier Stufen, die wie folgt beschrieben werden:

- Erfüllt die Erwartungen im besonderen Masse
- Erfüllt die Erwartungen
- Erfüllt die Erwartungen mit Einschränkungen und
- Erfüllt die Erwartungen nicht

In dem Bereich der Fächer sind es:

- gesichert
- Weitgehend gesichert
- Teilweise gesichert und
- Übungsbedarf
(ebd. S. 27)

Ein Kernproblem auch der alternativen Ansätze geht dahin, was als Bezugsnorm gelten soll. Die bisherige Praxis der Notengebung bezieht sich auf die Klasse oder Lerngruppe als Grundeinheit. Noten geben die Leistungsdifferenzen in dieser Einheit wieder und sie vergleichen die Leistungen der Schüler untereinander.

Die Alternativen beziehen sich auf Vorgaben des Lehrplans und stellen im Blick darauf Lernfortschritte fest. Auf einen Vergleich innerhalb der Lerngruppe wird verzichtet. Der Grund ist die Vermeidung von Diskriminierungen, während tatsächlich in jeder Lerngruppe Leistungshierarchien entstehen und die Schüler nicht einfach nur wissen wollen, ob ihr Wissen „gesichert“ ist oder ob „Übungsbedarf“ besteht.

Die Politik der Diskriminierungsvermeidung hebt die Leistungsunterschiede nicht auf, sondern versucht nur, sie nicht sichtbar zu machen. Damit wird der Differenzeffekt „gute Noten - schlechte Noten“ vermieden, aber zugleich auch die Urteilsmacht der Lehrpersonen minimiert. Sie legen nur noch fest, ob ein Lernprozess zu einem wie immer selbstgesetzten oder lehrplangerechten Ziel kommt, oder noch weitere Anstrengungen nötig sind.

Die Schul- und Notenkritik sieht die Schüler eher als Opfer des Systems und nicht als Nutzniesser seiner Lücken. Subversive Lernstrategien im Leistungsverhalten sind kaum erforscht, aber gehören zum Alltag, nicht nur in den Gymnasien. Ein Grossteil der Notenkritik ist normativ angelegt und sieht Defizite vor allem bei den Lehrern und nicht so sehr bei den Schülern.

Taktisches Verhalten bei der Notengebung, die ja zum Bestehen eines Abschlusses einen bestimmten Durchschnitt voraussetzt, wird so gut wie nie untersucht und ist gleichwohl ein nicht seltener Tatbestand. Die Schülerinnen und Schüler rechnen sich aus, welchen Aufwand sie betreiben müssen, um zu einem Ergebnis zu kommen, das den Abschluss sichert und so die Berechtigung garantiert. Diese Art „Motivation durch Noten“ ist rational und setzt Risikoabwägung voraus, aber wird vom „guten Schüler“ der Forschung gar nicht erfasst.

Die Nachteile von Ziffernoten sind häufig beschrieben worden, die Vorteile dagegen eher nicht, während auffällig ist, dass erst mit Einführung des Lehrplans 21 auch in der Schweiz über Alternativen oder die Weiterentwicklung des Systems der Ziffernoten nachgedacht wird. Die gestufte Beschreibung der Standards hat die Frage aufgeworfen, ob es unabhängig davon noch Noten braucht, weil ja die Einstufung der Leistungen in den Lernfeldern genügt. Aber dann wären Standards ein Art Notenersatz.

Offenbar haben Frontalangriffe auf die Ziffernoten nicht dazu geführt, das bewährte System abzuschaffen, im Gegenteil, mit dem Bologna-Prozess sind Ziffernoten auch im Tertiärbereich vorherrschend geworden. Jede Lernleistung wird bewertet, sie zählt buchstäblich für den Abschluss und muss daher ernst genommen werden. Notenfreie Module sind eine Grenzerfahrung und überall werden Ziffernoten vergeben.

Diese enge Praxis hat Kritik ausgelöst und die Kritik wird auch für Anpassung sorgen, aber es ist nie eine Notenkritik. Noten gelten als Motivationsanreize und die Stufung an sich ist nicht strittig, nur wird sie selten voll ausgenutzt. Aber niemand fordert die Abschaffung der Noten, das scheint ein Primarschulphänomen zu sein. Und man stelle sich auch Abschlussfeiern vor, bei denen Bestnoten fehlen und stattdessen Lernberichte verlesen werden.

Die Suche nach Alternativen zu den Ziffernoten wird weitergehen. Kluge Lösungen sollten aber vom bestehenden System ausgehen und weniger die Form der Beschreibung, als die Transparenz verbessern. Ziffernoten können ergänzt werden, aber wenn sie weiterhin Verwendung finden, dann muss besser dokumentiert werden, wie sie zustande kommen und was dem Urteil zugrunde liegt. Das Problem ist dann nicht die „Subjektivität“ der Noten, sondern die Klarheit der angelegten Kriterien.

5. *Praktische Vorschläge*

Jahrzehntelange Erfahrungen mit Lernberichten und Ähnlichem zeigen die Steigerung der Komplexität und damit einhergehend der drohenden Unverständlichkeit. Noten müssen klar und verständlich sein, die Abstände im Leistungsverhalten wiedergeben und hohe und tiefe Grade kennen. An dieser Anforderung sind Alternativen zu messen und so lange keine besseren Alternativen zum Ziffernsystem vorliegen, wird dieses auch weiterhin die Praxis bestimmen.

Die neuere Kritik verweist darauf, dass Schulnoten nur begrenzt eine Aussage darüber machen, was die tatsächlichen Kompetenzen der Schülerinnen und Schüler sind. Vieles, was für das Zustandekommen der Leistung mitverantwortlich ist, wird mit einer Ziffernote gar nicht erfasst, etwa die allgemeine Leistungsfähigkeit, schulisches Engagement und die

Motivation für bestimmte Unterrichtsfächer. Aber das wertet die Note nicht ab, jede Bewertung hat Grenzen.

- Noten erfassen Leistungen im Blick auf bestimmte Aufgabenstellungen, die von Lehrkräften im Blick auf eine bestimmte Gruppe bewertet werden.
- Endnoten nach einem bestimmten Unterrichtsabschnitt, meistens am Ende eines Schulhalbjahres, werden gebildet, indem verschiedene Einzelnoten addiert und ein (gewichteter) Durchschnitt gebildet wird.

Auch dieses Verfahren ist leicht handhabbar und nicht so schlecht, wie die Kritik meistens unterstellt. Die Lehrpersonen bewerten Leistungen aufgrund langjähriger Erfahrungen und handeln keineswegs willkürlich.

Noten sind einfache Instrumente, deren Aufwand begrenzt ist und die leicht kommuniziert werden können. Sie gelten - ausser bei den Notenkritikern - als bewährt und werden als Beschreibungsform auch ausserhalb der Schule breit angewendet. Das Problem ist nicht das gestufte System der Leistungsbewertung, einfach weil die Leistungen gestuft sind, sondern was es erfassen soll und zu welchem Zweck es eingesetzt wird.

Bei der Konzentration auf das Feindbild „Ziffernnoten“ ist oft übersehen worden, wie vielfältig Leistungsbewertungen in der Schule tatsächlich sind. Es gibt formelle und informelle Einschätzungen des Leistungsverhaltens, Beurteilungen der Inhalte und der Form einer Leistung, schriftliche und mündliche Prüfungen, Tests, Proben, Abschlussarbeiten und vieles mehr. Allerdings ist nicht jede Bewertung gleich wichtig, und auch die Perspektive der Relevanz ist unterschiedlich.

Für die Schülerinnen und Schüler in Schweizer Gymnasien war die Maturaarbeit auch dann von hoher Bedeutung, als sie lediglich die Zulassungsvoraussetzung für die Maturitätsprüfung darstellte. Mit der Arbeit wurde selbständiges Lernen mit einem frei gewählten Thema realisiert, was einen Wert an sich darstellen kann. Die Beurteilung war relevant, auch wenn die Note nichts zählte (Oelkers 2008, S. 203ff.).

In der Prüfungspraxis sind bis heute Noten zentral, also die Einschätzung der Leistungen von Schülerinnen und Schüler auf einer für alle Lehrkräfte verbindlichen Skala, die die Unterschiede von Fähigkeiten in Sachgebieten erfassen soll. Trotz einer Vielzahl von scheinbaren oder tatsächlichen Alternativen ist die Notenskala das bei weitem gebräuchlichste Instrument der Leistungsbeurteilung.

Für dieses Instrument spricht, dass Leistungen in Schulklassen tatsächlich immer mit Niveauunterschieden zustande kommen. Wer sie abbilden will, *muss* daher ein gestuftes Schema verwenden, wobei das Problem nur ist, welche Stufen zur Anwendung kommen und wie die tatsächlichen Leistungsunterschiede in der Beurteilung abgebildet werden.

- Die realistische Perspektive ist die Beibehaltung des Ziffernsystems unter der Voraussetzung, dass die Notengebung fair und transparent ist.
- Dafür müssen die einzelnen Schulen Kriterien festlegen, die für die Lehrerinnen und Lehrer verbindlich sind.
- Diese Kriterien beeinträchtigen nicht das Urteil der Lehrpersonen, sondern machen es für Schüler, Eltern und andere Lehrer transparent.

Das Prinzip der Stufung wird auch in den Bildungsstandards für die Schweizer Volksschule deutlich. Für sie wurden Kompetenzmodelle entwickelt, die wie selbstverständlich davon ausgehen, dass sich die Leistungen auf verschiedenen Stufen erfassen lassen. Die Erreichbarkeit der Leistungsziele ist dann je nach Aufwand, Talent und Vorerfahrung unterschiedlich.

Das unterscheidet die neuen Aufgabenkulturen von den klassischen Lehrplänen, die so etwas wie „Regelstandards“ beschrieben haben. Solche Vorgaben sollten *alle* Schülerinnen und Schüler erreichen können, während tatsächlich immer Leistungsabstände das Verhalten bestimmen. Aus diesem Grunde schreibt das HarmoS-Konkordat auch „Mindeststandards“ vor; alle Schüler müssen eine *bestimmte* Stufe erreichen, nicht den gesamten Lehrplan, was weder möglich noch sinnvoll wäre.

Viele Bildungssysteme verfügen heute über Standards, Leistungstests und komplexere Formen der Evaluation. Das elaborierteste System in Europa ist seit 1988 in England aufgebaut worden, und selbst in den liberalen Gesamtschulsystemen Skandinaviens werden inzwischen Standards und Tests verwendet. Es gibt also überall mehr als nur die gewohnte Prüfungspraxis in den einzelnen Schulen. Die Frage ist, wie diese Praxis so verändert werden kann, dass sie zur Gesamtentwicklung passt und zugleich den einzelnen Schulen sichtbare Vorteile bringt. Neue Formen der Beurteilung, anders gesagt, müssen einen Beitrag zur Unterrichtsentwicklung leisten, sonst werden sie nicht akzeptiert.

Für die Schweizer Gymnasien sind Leistungsstandards nichts Neues, neu wären externe Leistungstests, die derzeit aber nicht abgesehen werden können. Sie würden die internen Leistungsbewertungen auch nicht ersetzen, sondern allenfalls ergänzen. An flächendeckende Tests ist nicht gedacht, und auch ein „High-Stakes-Testing,“ also Tests mit Folgen für die Schule oder gar für die einzelnen Lehrkräfte,¹⁵ steht nicht zur Diskussion und hätte auch keine Akzeptanzchancen.

Wenn die Prüfungspraxis optimiert werden soll, muss die Schule das zu einem vorrangigen Entwicklungsziel erklären und die Implementation zur gemeinsamen Sache der Akteure machen. Das verlangt Führungsarbeit seitens der Schulleitung, die Verfahren werden von den Lehrkräften ausgearbeitet, die Notwendigkeit der Entwicklungsarbeit und der Sinn des neuen Systems muss mit allen Akteuren kommuniziert werden. Die Eltern sowie die Schülerinnen und Schüler gewinnen dann einen Einblick, nach welchen Kriterien benotet und geprüft wird. Sie können sich dann darauf einstellen, was vor allem für die Akzeptanz der Noten von grosser Bedeutung wäre.

In einem transparenten System der Leistungsbeurteilung wären Noten gestufte Formate mit einem gemeinsamen Massstab. Die Notengebung würde sich nach Kriterien vollziehen, die für alle Beteiligten klar sind und die fortlaufend kommuniziert werden. Dann gibt nicht jede Lehrkraft „ihre“ Noten, vielmehr werden die Notenstufen pro Fach oder Fachbereich festgelegt und auf das eingestellt, was als Leistungsstandard gilt. Standards unterscheiden sich nach Stufen und darauf beziehen sich die Noten. Die einzelnen Stufen erhalten eine schriftliche Fassung, die den Unterschied beschreibt. Sprachlich möglich ist das durch Graduierungen, die zu den Stufen passen.

Ein Beispiel für die Benotung einer längeren Klausur aus dem Universitätsbereich sähe so aus:

¹⁵ Einzelne Schuldistrikte in den Vereinigten Staaten, etwa in Cincinnati, koppeln die Löhne an die Ergebnisse der Leistungstest.

Noten:

- *Exzellente* (6.0): eigenständige, umfangreiche, zutreffende und mit herausragenden weiterführenden eigenen Reflexionen angereicherte Darstellung.
- *Sehr gut* (5.5.): eigenständige, zutreffende und weiterführende Darstellung.
- *Gut* (5.0): Zutreffende und weiterführende Darstellung.
- *Befriedigend* (4.5.): Zutreffende Darstellung mit leichten Mängeln.
- *Ausreichend* (4.0): Zutreffende Darstellung mit Mängeln.
- *Nicht ausreichend* (unter 4.0): Darstellung mit gravierenden Mängeln.

Das auszuhandeln und festzulegen wäre in den Schulen Aufgabe der einzelnen Fachschaften, wobei die Schulleitung ein allgemeines Format für die Beschreibung auch vorgeben kann. Die Anpassung muss auf die Besonderheiten der Fächer Rücksicht nehmen. Entscheidend ist, dass transparent gemacht wird, wann welche Note vergeben wird und was sie für ihren Referenzraum bedeutet. Identische Kriterien der Notengebung können auch zwischen den Fachschaften verschiedener Schulen vereinbart werden.

Das Zürcher Projekt HSYGM läuft auf ein System der fachlichen Abstimmung sogar zwischen den Gymnasien und Universitäten hinaus. Damit müsste sich eigentlich die Frage der gemeinsamen Kriterien für die Bewertung stellen, wenngleich das über den Projektauftrag und die Empfehlungen hinausgeht. Wichtig ist auch, dass im Projekt überfachliche Kompetenzen beschrieben wurden, die sich nicht mit den Kompetenzen des Lehrplans 21 decken, sondern Anforderungen an die Studierfähigkeit deutlich machen. Es geht nicht um Lernen im Blick auf (oft simulierte) Anwendungssituationen, sondern um einen Habitus für das Studieren.

Diese Kompetenzen wurden in den Empfehlungen von 2008 so gefasst:

- Kritisch-forschendes Denken als gymnasiale Schlüsselkompetenz
- Selbstständigkeit und Selbstverantwortung
- Verstehen wissenschaftlicher Texte
- Abfassen schriftlicher Abhandlungen
- Beherrschung der Sprachen (D, F, E)
- Quellenkritik
- Reflexion und Kritikfähigkeit
(HSGYM 2008, S. 42-46).¹⁶

Das kann zwischen den Fächern unterschiedlich verstanden und realisiert werden, sollte aber als Zielnorm des Gymnasiums bei den Aufgabenstellungen und der Beurteilung der Lernfortschritte Berücksichtigung finden.

Die Kernfrage ist, was bewertet man wie mit einer Note, vorausgesetzt die Leistungsniveaus des Gymnasiums. Das „was“ bestimmen in fachlicher Hinsicht der Lehrplan und das konkrete Curriculum der Fachschaften. Eine „kriteriale Norm“ für die Noten direkt aus dem Lehrplan abzuleiten, ist weder sinnvoll noch praktikabel. Die Lehrpersonen müssen an dieser zentralen Norm ihres Unterrichts direkt beteiligt sein. Insofern ist der Bezugspunkt nicht einfach der Lehrplan, sondern die konkreten Zielsetzungen der Schule, so wie die

¹⁶ Siehe auch: <https://www.hsgym.ch/analyse-and-empfehlungen/ueberfachliche-kompetenzen>

Fachschaften sie für einen bestimmten Zeitraum - ein Semester, ein Jahr oder länger - festgelegt haben.¹⁷

Das „wie“ bezieht sich auf die Abfolge der Leistungen und der Leistungsbewertung während eines Zeitraumes. Leistungen setzen Aufgaben voraus, die Güte der Aufgaben ist der Schlüssel zur Leistungsbewertung und so der Noten. Aufgaben beziehen sich auf fachliche Probleme, setzen Vorwissen voraus, bezeichnen Anschlüsse und legen Strategien der Bearbeitung nahe. Die Note bewertet die Problemlösung, den Lernfortschritt gemessen an vorangehende Aufgaben und die Lernstrategien. Ein enges Korsett von Kompetenzstufen ist dabei nicht sehr hilfreich.

Eine Veränderung der Prüfungspraxis gelingt dann, wenn die Expertise der Lehrerschaft genutzt wird und auch das ist nur möglich mit Kooperationen vor Ort. Wenn die Notenformate nicht auf das Fach hin abgestimmt sind, ist mit der Präzisierung der Beschreibung wenig gewonnen. Die Notenformate werden *mit* den Fachansprüchen konkretisiert, und das verlangt den Konsens und die fortlaufende Abstimmung in einer Fachschaft. Sie ist das zentrale Medium der Entwicklung, die von der Schulleitung koordiniert und moderiert wird. Aber was Leistungen und Noten „sind,“ legt die Fachschaft fest.

Wenn gemeinsames Prüfen zu einem schulischen Standard werden soll, dann setzt das diesen Weg der Selbstverpflichtung voraus. Gemeint ist nicht eine formale Prüfung im Team oder mit einer Kommission, die es bereits heute gibt. Vielmehr geht es darum, von gemeinsamen Kriterien oder Standards auszugehen und danach zu prüfen. Ähnlich wie man in einem Fach festlegen kann, was ein Fehler ist und was nicht, so man auch bestimmen, was eine Note ist und worauf sie sich bezieht. Die berufliche Identität der Lehrkräfte ist davon nicht berührt, im Gegenteil stellt es einen professionellen Gewinn dar, wenn für mehr Transparenz in einem zentralen Geschäftsbereich der Schule gesorgt ist.

Zeugnisse basieren auf Noten. Noten sind wie gesagt in bestimmter Hinsicht prognosefähig. Wer einen hohen Leistungsschnitt erreicht, wird mit einiger Sicherheit auch in nachfolgenden Bildungsgängen gute bis sehr gute Leistungen zeigen. Das gilt mit ähnlicher Wahrscheinlichkeit auch für den umgekehrten Fall. Allerdings werden mit Noten nicht direkt Fähigkeiten und Fertigkeiten beschrieben. Sie beziehen sich auf Leistungen im Blick auf Aufgaben in bestimmten Fächern oder Lernbereichen.

Gängige Alternativen zum Typus des Ziffernzeugnis sind individuelle Lernberichte, die die Lehrkräfte abfassen oder fortlaufende Portfolios, die von den Schülerinnen und Schülern erstellt werden. Sie dokumentieren den Prozess und die Gegenstände des Lernens. Lernberichte sind aufwändig zu erstellen. Sie sind zudem interpretationsabhängig und benutzen oft sprachliche Formen, die für die Lesenden unverständlich sind. Portfolios dokumentieren einfach nur das, was gemacht worden ist. Sie sind meistens sehr umfangreich und lassen oft keinen wirklichen Schluss auf das Ergebnis oder den Lernstand zu.

Als Anlage zu Zeugnissen haben sie illustrierenden Charakter, ohne wirkliche Benotung durch die Lehrkräfte sind sie für Abnehmer und die Lernenden selbst wenig aussagekräftig. Die Schülerinnen und Schüler wollen wissen, wo sie an einem bestimmten Zeitpunkt im Vergleich mit Anderen und/oder einer Bezugsnorm stehen. Die Dokumentation

¹⁷ Im Kanton Zürich haben die Gymnasien eigene Lehrpläne, aber auch die müssen im Blick auf die Praxis konkretisiert werden.

des eigenen Lernweges ist eine Sache, die objektive Bewertung und Rangierung der Leistung durch Dritte eine andere.

Als Ergänzung zur Bewertung durch die Lehrpersonen können auch Leistungstests nützlich sein, die inzwischen auch in der Schweizer Volksschule Einzug gehalten haben. Die bekanntesten Formen sind Klassencockpit und Stellwerk, beides inzwischen bewährte Instrumente zu einer vergleichenden Leistungsbeurteilung, die über die Klassennorm hinausgeht und sich auf einen kantonalen Durchschnitt bezieht. Von besonderer Bedeutung für die Regelung der Schnittstelle zwischen Volksschule und Berufsbildung ist der Test Stellwerk.

Der Umgang mit solchen Instrumenten ist pragmatisch. Was den Schulen nutzt und einen Mehrwert verspricht, findet Verwendung, was nicht wird schnell beiseitegelegt oder gar nicht erst ausprobiert. Das gilt für Leistungstest ebenso wie für externe Evaluationen oder neue Formen der Leistungsbeschreibung. Der Charme der Ideologie verschwindet schnell, wenn die Grenzen der Wirksamkeit sichtbar werden.

Im Blick auf das Thema und die Fragestellung kann ich so schliessen: Bei der Bewertung von Leistungen in der Schule werden Noten und Zeugnisse weiterhin die zentrale Rolle spielen, die Instrumente sind bewährt und begrenzen den Aufwand. Aber die Notengebung kann verbessert werden. Ich habe vier Aufgaben näher beleuchtet, Klarheit der Kriterien, die Präzisierung der Stufung, die zur Bewertung passende Aufgabenkultur und die schulischen Lernziele als Bezugsnorm. Die bessere Berücksichtigung des Lernwegs kann fünftens durch Portfolios erreicht werden. Insgesamt also eine lösbare Aufgabe.

Literatur

Backhaus, A./Czerny, S.: Die Not mit den Noten! Beitrag zum Kongress „Herausforderungen“ Bodenseeregion Oktober 2008.

http://www.adz-netzwerk.de/files/docs/kongress2008/die_not_mit_den_noten.pdf

Beurteilungen und Schullaufbahnentscheide. Über das Fördern, Notengeben und Zuteilen.

Hrsg. Bildungsdirektion Zürich Volksschulamt. 3. überarb. Aufl. Zürich: Lehrmittelverlag des Kantons Zürich 2013.

Bos, W./Beutel, S.-I./ Berkemeyer, N./Schenk, S.: LUZI. Leistungsbeurteilung ohne Ziffernzeugnisse. Abschlussbericht der wissenschaftlichen Begleitforschung. Dortmund: IFS 2010.

Brügelmann, H.: Misstraut allen Noten! Eine Streitschrift. In: Die Zeit Nr. 29 vom 13. Juli 2016.

Döring, O.: Untersuchungen zur Psychologie des Lehrers. Leipzig: Quelle&Meyer 1925.

Dunkake, I./Kiechle, Th./Klein, M./Rosar, U.: Schöne Schüler, schöne Noten? Eine empirische Untersuchung zum Einfluss der physischen Attraktivität von Schülern auf die Notenvergabe durch die Lehrpersonen. In: Zeitschrift für Soziologie Band 41, Heft 2 (April 2012), S. 142-161.

Heilman, M.E./Saruwatari, L.R.: When Beauty is Beastly: the Effects of Appearance and Sex on Evaluations of Job Applicants for Managerial and Nonmanagerial Jobs. In: Organizational Behavior and Human Performance Vol. 23, No. 3 (June 1979), S. 360-372.

HSGYM - Hochschule und Gymnasium: Hochschulreife und Studierfähigkeit. Zürcher Analysen und Empfehlungen zur Schnittstelle. Zürich: HSGYM 2008.

- Leu, J.: Kinder wollen Noten. In: Die Zeit Nr. 27 vom 29. Juni 2006.
- Methodus Oder Bericht/Wie Nebst Göttlicher Verleyhung die Knaben und Mägdlein auf den Dorffschaften/und in den Städten/die untere Classes der Schul-Jugend im Fürstenthum Gotha/kurtz-und nuetzlich unterrichtet werden können und sollen. Auf gnädigsten Fürstlichen Befehl aufgesetzt. Gotha: Bey Christoph Reyhern 1697. (gedr. 1698)
- Oelkers, J.: Die Qualität der Gymnasien. Eine Expertise für den Kanton Zürich. Bern: h.e.p. Verlag 2008.
- Olechowski, R./Rieder, K. (Hrsg.): Motivieren ohne Noten. Wien u.a.: Verlag Jugend und Volk 1990. (= Schule, Wissenschaft und Politik, Band 3)
- Schreiber, H.: Gegen Prüfungen und Noten. In: Zeitschrift für Philosophie und Pädagogik 6 (1899), S. 31-38
- Trapmann, S./Hell, B./Weigand, S./Schuler, H.: Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - Eine Metaanalyse. In: Zeitschrift für pädagogische Psychologie Band 2, Heft 1 (2007), S. 11-27.
- Trost, G.: Die Not mit den Noten. Sind gute Schüler auch erfolgreiche Studenten? In: Die Zeit Nr. 10 vom 27. Februar 1976.
<http://pdf.zeit.de/1976/10/die-not-mit-den-noten.pdf>
- Weber, H.M./Petermann, F.: Der Zusammenhang zwischen Schulangst, Schulunlust, Anstrengungsvermeidung und den Schulnoten in den Fächern Mathematik und Deutsch. In: Zeitschrift für Pädagogik Band 62, Heft 4 (Juli/August 2016), S. 551-570.