

Jürgen Oelkers

*Leistungsbeurteilung: Ein Widerspruch zur Bildungsgerechtigkeit? *)*

Mein Vortrag beginnt mit „Effektstärke“. Damit will ich nicht auf die Wirkung des Vortrages hinweisen und bringe mich auch nicht vorsorglich in Stellung für die Zufriedenheitsbefragung am Ende dieser Tagung, sondern beziehe mich auf das Tagungsthema. Feedback erhält bekanntlich auf John Hatties „Barometer des Lernens“ eine Spitzennote und ich verstehe, dass man sich dieses höchsten Garanten für Wirksamkeit auch in Kiel versichern will. Wobei - jedes Feedback kann am Zuhören scheitern und wer ist schon ein guter Zuhörer?!

Aber mich interessiert etwas anderes: Hatties Erfolg hat seinen Grund darin, dass er Noten verteilt, ein Tatbestand, der zur Sprache des Schulfeldes perfekt zu passen scheint. Zeugnisse und Noten generell sind Formen der Rückmeldung, aber darüber zu sprechen, ist nicht ganz leicht und auch heikel, denn Zeugnisse und Noten stehen in der Kritik und mehr noch, sie gelten als gefährlich und werden als überflüssig hingestellt.

- Noten sind unpräzise,
- ihr Zustandekommen ist intransparent,
- sie wirken als eine Art Schicksal,
- sind vor allem ein Machtfaktor
- und sicher kein Beitrag zur Bildungsgerechtigkeit.

Wäre das so, hätte sich mein Thema erledigt, ich könnte gehen und müsste dann nur noch erklären, warum ich Kosten verursacht habe, von Zürich nach Kiel gereist bin und nicht in der Jugendherberge übernachtet habe. Was mich retten könnte, wäre die Frage, wieso Noten als Teil des Sündenfalls betrachtet werden und warum eine *damnatio memoriae*¹ verlangt wird, als sei zur Lösung des Problems das Kirchenrecht zuständig. Ich werde zu zeigen versuchen, dass es sich nicht um eine rhetorische Frage handelt.

Das „starre Ziffernotensystem“ ist 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. Die Lehrerbildung liebt das „Intrinsische“, aber man lernt auch, wenn man motivationale Widerstände überwinden muss und man stelle sich vor, wohin man käme, wenn alles Lernen von der Zustimmung intrinsischer Motivation abhängig wäre. Vielleicht ist das ja der Grund, warum der frühe Schulbeginn kritisiert wird wie noch nie.

*) Vortrag auf dem 7. Kieler Schulleitungssymposium am 23. September 2016.

¹ Auslöschung aus dem Gedächtnis.

Noten als positive Anreize sind auch deswegen suspekt, weil sie eine Hierarchie voraussetzen. Nur wenige Schüler können Bestnoten erreichen und das, so die Kritik, 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. Das bekanntlich schon Jean-Jacques Rousseau 1762 in seinem Erziehungsroman *Emile ou de l'éducation* postuliert.

Es ist immer wieder versucht worden, Alternativen zu der Notenskala zu entwickeln. Radikale Entwürfe gehen 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 und zudem die unterschiedlichen Voraussetzungen der Lernenden ignoriert. Aber so leicht wird man das Problem nicht los.

- 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).

Historisch kann von einem sehr stabilen Instrument ausgegangen werden. Die gestufte Notenskala zur Beschreibung der Leistungen in einer Lerngruppe 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,² verfasst worden. erschien zuerst 1642 und erlebte zahlreiche Ausgaben. Der Methodus war Teil der ernestinischen Reformen im Fürstentum Gotha, die als erste staatliche Schulentwicklung im deutschen Sprachraum gelten können.³ Verglichen mit heutigen Lehrplänen war der Methodus knapp und konzise, auf nur etwa hundert Seiten konnte alles Wesentliche gesagt werden; erst viel später wurden die Lehrplankommissionen erfunden, mit allen Risiken für die Aufblähung des Umfangs.

Seit der Ausgabe des Jahres 1662 schrieb der Methodus ein Stufenschema zur Beurteilung der Leistungen in drei Dimensionen vor. 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.).

² Andreas Reyher (1601-1673) wurde 1641 als Rektor des *Gymnasiums illustre* nach Gotha berufen.

³ 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 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. In diesem Sinne handelt es sich um einen Standard, den nicht die Lernpsychologie, sondern die Schule selbst hervorgebracht hat und der mit einiger Wahrscheinlichkeit auch in Zukunft Verwendung finden wird.

Aber was wird dann aus der Bildungsgerechtigkeit? Wenn man als „sehr fein“, „fein“ und „fromm“ eingestuft wird, und dies auf Dauer, dann hat das ein anderes Selbstbild zur Folge, als wenn die Wertung „schlecht“, „etwas/wenig“ und „unfleissig“ lauten würde. Aber ist das ungerecht?

Die Antwort lautet ja, wenn die Bewertungen *willkürlich* erfolgen würden. Ein Bewertung ist willkürlich, wenn sie

- die Aufgaben und Leistungen mit verschiedenen Massstäben interpretiert,
- einzelne Schüler gegenüber anderen bevorzugt,
- ohne Kriterien erfolgt
- oder unfaire Massstäbe anwendet, also mehr oder anderes erwartet, als gelernt werden konnte.

Der Grundsatz ist, dass nur das geprüft werden darf, was unterrichtet worden ist und so gelernt werden konnte. Wenn dieser Grundsatz verletzt wird, ist das ungerecht. Wenn ein Schüler schlecht beurteilt wird oder nicht die Note erreicht, die er erreichen wollte, ist das *nicht* ungerecht, sofern die Kriterien der Benotung klar waren und keine Bevorzugung erkennbar ist.

Klagen über die Ungerechtigkeit der Notengebung müssen sich auf die gängige Praxis beziehen lassen, wenn ein realistisches Bild entstehen soll.

- 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 errechnet wird.

Dieses Verfahren ist leicht handhabbar und nicht so schlecht, wie die Kritik häufig annimmt. Die Lehrpersonen stellen Aufgaben und bewerten Leistungen aufgrund langjähriger

Erfahrungen und handeln vermutlich in den wenigsten Fällen wirklich „willkürlich“ in dem genannten Sinne.

Noten sind Teil des Zweifels an der „Bildungsgerechtigkeit“ und die sind weit verbreitet. Vor einigen Wochen ist in einem bekannten und nicht einflusslosen deutschen Nachrichtenmagazin ein Erziehungswissenschaftler mit dem Satz zitiert worden:

- „Das Einzige, was die Kinder wirklich aus der Schule mitnehmen, ist, wie man mit Ungerechtigkeit umgeht“.
- Wer dieser Kollege mit dem riskanten Satz ist, wird nicht gesagt.
- Das Zitat steht in einem Artikel, der den „grossen Frust“ der vielen schlechten Lehrer als Erklärung dafür nimmt, warum für viele Kinder die Schule „ein Ort der Demütigung“ ist.

Die Verfasserin ist eine Lehrerin mit langjähriger Erfahrung. Die Kritik kommt also aus den eigenen Reihen. Zur Abhilfe empfiehlt sie, was in Deutschland immer empfohlen wird, mehr Kreativität im Unterricht, kein direkter Weg von der Schule über die Ausbildung in die Schule, mehr Praxisbezug im „verkopften“ Studium, Aufhebung des Beamtenstatus, damit die schlechten Lehrer entlassen werden können, Veränderung der Notenpraxis, Vorbild Finnland und „niemanden zurücklassen“ (Der Spiegel Nr. 36 vom 3.9. 2016, S. 62/53).

Was genau getan werden soll, bleibt offen, aber irgendwie soll es dadurch nicht nur besser, sondern auch gerechter werden. Und tatsächlich gibt es keine Bildungsreform, die nicht mehr Gerechtigkeit versprochen hätte. Aber Lehrer ohne Beamtenstatus sind nicht gerechter als Lehrer mit; Finnland hat eine kontant hohe Jugendarbeitslosigkeit und im Blick darauf wären die Schweiz, Deutschland oder die Niederlande Vorbilder; und Noten gibt es überall, selbst in Finnland.

Der Satz des unbekanntem Kollegen wird offenbar zu allzu gerne geglaubt, zumal mit der Adaption durch einen wenn auch unbekanntem Erziehungswissenschaftler. Aber er wird dadurch nicht wahrer und ist nur ein schulkritischer Reflex, der schlechte Gefühle auslöst und sich mit Appellen verbinden lässt, die abstrakt bleiben, die Moral für sich haben und nichts ändern. Zudem: Wer dieser Kritik folgt, hat Mühe, anders als mit und über Karikaturen zu diskutieren, eben „ungerechten“ Noten, suggestiven „Vorbildern“ oder „schlechten“ Lehrern, die man am besten sofort entlassen sollte. Die Arbeitsgerichte werden sich freuen.

Vermeidet man die Karikaturen, dann lässt sich im Blick auf die schulische Notenpraxis festhalten: Schulnoten bewerten Leistungen und können 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 auch gar nicht erfasst, etwa die allgemeine Leistungsfähigkeit, schulisches Engagement, Vor- und Nachteile der sozialen Herkunft oder die Motivation für bestimmte Unterrichtsfächer.

Aber jede Bewertung hat Grenzen und keine umfasst alles. Noten haben den Vorteil, dass sie einfach sind, leicht zu kommunizieren und keinen übermässigen Aufwand verlangen. Sie passen ins Arbeitsfeld der Schule, gelten als bewährt und werden als Beschreibungsform auch ausserhalb der Schule breit angewendet. Niemand stört sich daran, dass Fussballprofis jede Woche Noten erhalten, Filmkritiker vergeben Noten ebenso wie Restaurantkritiker, ein Hotel ohne Noten würde man kaum buchen und die Stiftung Warentest könnte ohne Noten ihren Betrieb einstellen.

- Das Problem ist nicht das gestufte System der Leistungsbewertung,
- einfach weil Leistungen, wenn sie verglichen werden, immer gestuft sind,
- sondern was das System 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, Selbstbewertungen, Tests, Proben, Abschlussarbeiten und vieles mehr.

Allerdings ist nicht jede Bewertung gleich wichtig, die Relevanz wird von den Lernzielen und dem Design der Lernprozesse her bestimmt. Am Ende einer Zeiteinheit steht ein Ergebnis, das beurteilt worden ist. Seine Basis sind die gezeigten Leistungen und die Beurteilung muss imstande sein, die Unterschiede zu erfassen, sofern gleiche Ziele massgebend waren.

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 Grundschulbereich. Damit reagiert die Literatur auf Vorstellungen kindlichen Lernens, die nicht von Wettbewerb und Leistungsdruck geprägt sein sollen.

Die dafür gewählte Sprache, die etwa von „Notenterror“ spricht,⁴ 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 nur am Ende Empfehlungen für die Übergänge abgeben, die aber gegenüber dem Elternwillen in den meisten Bundesländern nicht bindend sind.⁵

Auch in der Schweiz scheint es zunehmend schwer zu werden, Noten für Primarschülerinnen und Primarschüler als geeignet anzusehen. Allerdings reagieren die Behörden anders als in Deutschland, auch weil sie einen Konsens für Noten in der Bevölkerung voraussetzen können. Die Primarschule ist in der Schweiz nicht die Bastion, um langfristig überall für Notenfreiheit zu sorgen.

Im Kanton Zürich werden mit dem Ende des ersten Schuljahres Zeugnisse mit Ziffernnoten vergeben, was nach der Kritik eigentlich zu verheerenden Wirkungen führen müsste, die aber niemand bislang festgestellt hat.

- 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 Laufbahnentscheide 2013, S. 16).

⁴ Der Ausdruck stammt aus den westdeutschen Schulkämpfen der siebziger Jahre („Gegen Notenterror“: Der Spiegel Nr. 2 von 9. 1. 1978, S. 64/65).

⁵ Verbindliche Grundschulempfehlungen, die einen bestimmten Notenschnitt für den Übertritt ins Gymnasium voraussetzen, gibt es nur noch in Bayern, Brandenburg, Sachsen und Thüringen.

- 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.

Für alle Deutschschweizer Kantone gibt es erstmalig einen gemeinsamen Lehrplan für die Volksschule. Die schweizerische Erziehungsdirektorenkonferenz (EDK) hat klargestellt, dass der neue Lehrplan die Form der Leistungsbeurteilung nicht definiert. Die Lehrpersonen werden also auch mit schulisch angepassten Lernziele arbeiten deren Erreichen überprüfen können. Eine kantonale Prüfbehörde ist nirgendwo auch nur angedacht worden, weil klar ist, dass Prüfen in der Verantwortung des Lehrers bleiben muss.

In den Gymnasien der Schweiz gibt es Zeugnisse praktisch überall nach dem ersten Semester.⁶ Spielraum ist nur vorhanden bei der Leistungsbeurteilung bis hin zur Zeugnisnote. Und auch in der Schweiz wird deutlich, dass sich die Primarschule stärker differenziert hat als die beiden Sekundarstufen und die Sekundarstufe I sich eher mit neuen Formen der Leistungsbeschreibung auseinandersetzt als die Sekundarstufe II, also Gymnasien und Berufsschulen. „Neu“ meint dabei ergänzend zu den Ziffernnoten.

Unter den Schülerinnen und Schülern sind vergleichend benotete Leistungsunterschiede überwiegend 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 Ziffernnoten 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.

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

⁶ http://www.edudoc.ch/static/strukturdaten/pdf_rohdaten/167.pdf

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).⁷

In der Kritik ist seit Jahren nahegelegt worden, 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.

Diese Idee hat seit den siebziger Jahren des vergangenen Jahrhunderts die Diskussion vor allem in der deutschen Grundschulpädagogik bestimmt. Aber grundsätzlich muss zwischen Noten und Lernberichten kein Gegensatz angenommen werden. Lernberichte, gut gemacht und verständlich geschrieben, können Noten ergänzen oder darauf hinführen. Die Vermeidung von impliziten Noten in Texten ist schwierig und Ziffernoten haben den Vorteil, dass sie ad hoc verständlich sind, Akzeptanz in der Öffentlichkeit finden und ohne grosse Nachfragen kommuniziert werden können.

Kritik am Notensystem ist seit Beginn des 20. Jahrhunderts immer wieder vorgebracht worden (etwa schon Schreiber 1899), ohne dass sich an der Praxis sehr viel geändert hätte. Und über Jahrzehnte sind die immer gleichen Argumente ausgetauscht worden, während versäumt wurde, gegenteilige Thesen zu prüfen. Da ist es nicht überraschend, wenn als Königsargument festgehalten wird: 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.⁸

Aber das muss vor dem Hintergrund einer Praxis verstanden werden, in der die meisten Lehrkräfte sich Mühe geben, den Leistungen gerecht zu werden und Willkür zu vermeiden. Etwas anders sollte man ernst nehmen: 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.

Abitur ist und bleibt Abitur, die Berechtigung wird nicht durch den Entstehungsort aufgewertet oder eingeschränkt. Doch jede Ortsbindung der Noten würde das System der Berechtigungen sprengen und die Zugänge verlagern. Deshalb lag die umgekehrte Reaktion nahe, nämlich das Zentralabitur, also die bürokratische Antwort auf die Kritik der unterschiedlichen Vergabebedingungen, was die Länderunterschiede aber nicht ausgleicht, sondern nur die Zuständigkeit verlagert und die Schulautonomie beschränkt.

⁷ Allerdings ist die Datenbasis mehr als schmal, untersucht wurden drei Klassen eines 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).

⁸ So schon Hofer (1975).

Zum Vergleich: Die Gymnasien im Kanton Zürich haben jeweils einen eigenen Lehrplan, organisieren die Prüfungen selbst und müssen nur gemeinsame Verfahren anwenden. Es geht also auch ohne Zentralismus, zumal schulferne Urteile kaum „objektiver“ sind als die verantwortungsvoller Lehrerinnen und Lehrer, die ihre Schüler vor Augen haben und abschätzen können, wie die Leistungen zustande gekommen sind.

- „Im Durchschnitt betrachtet“, 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 Urteile von Lehrerinnen und Lehrer durch Leistungstests reduziert nicht nur den Beurteilungsraum und so 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).

Sie suggeriert eine Objektivität der Leistungsbeurteilung, die das System nicht „gerechter“ macht, weil die Anforderungen zum Bestehen der Tests bekannt sein müssen und und so auch unterlaufen werden können, während die Nebenfolgen zur Hauptsache werden, etwa im Blick auf Leistungslöhne, die nach Testergebnissen gezahlt werden, oder bezogen auf die Bildungspolitik, die unter dem Druck der öffentlich zugänglichen Resultate „failed schools“ schliessen will oder muss. Scheitern ist das Scheitern auf einer Rangliste, also einer Notenskala, die objektiv erscheint, weil sie den Einzelfall nicht betrachten muss.

Im Blick darauf beschreiben Noten den internen Rangunterschied. Und die Lehrkräfte sind auch gut in der Lage sind, die Schüler ihrer eigenen Klasse gemäss ihren Leistungen zu rangieren, das heisst jedoch nicht, „dass gleichen Noten in unterschiedlichen Klassen auch vergleichbare Leistungen zu Grunde liegen“ (Weinert 2001, S. 50). Das trifft zu, ist aber nur vermeidbar, wenn auf Noten verzichtet wird, die ja im Anschluss an die Leistung erteilt werden und keine Aussagen darüber zulassen, was an andern Orten bei gleichen oder ähnlichen Aufgaben geschieht.

Eine vergleichende Notenbörse ist nicht in Sicht und ob das ein so gute Idee wäre, sei dahingestellt. Wenn aber auf Noten verzichtet wird, hätte man gar keinen Anhaltspunkt mehr, was die Leistungsunterschiede ausmacht und würde auf vergleichende Prognosen verzichten. Das unscharfe Verhältnis zwischen Noten und Berechtigungen, das häufig beklagt wird, wäre so gesehen das deutlich kleinere Übel.

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 bestimmen 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 wie gesagt 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. Sprichwörtlich wurde einfach die „Not mit den Noten“⁹ während sich Schüler und Eltern überwiegend Noten wünschen. Die Kritik sagt dann meist, Kinder und Eltern haben sich einfach an das falsche System gewöhnt, Noten seien nicht aussagekräftig und sollten abgeschafft werden.

Eltern sehen das anders und sind mehr denn je an etwas ganz anderem interessiert, nämlich wie die Noten praktisch zustande kommen und ob die Kriterien für sie und die schulische Öffentlichkeit transparent sind. Vor Gericht ziehen Eltern nur bei schlechten oder nicht genügend guten Noten und wer eine Flut von Klagen vermeiden will, sollte die Klarheit der Noten zum Schulprinzip erheben. Dazu gleich mehr.

Die pädagogische Notenkritik richtet sich primär auf den Aussagewert der Noten und die Beschreibung der Leistungen 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. Im Ergebnis wird festgehalten: „Besonders deutsche Durchschnittsnoten erreichen hohe Validitäten, die auch von studienfachnahen Schulnoten nicht übertroffen werden“ (ebd. S.

⁹ Die „Not mit den Noten“ ist ein viel zitierter Titel in den Medien, der schon mehr als dreissig Jahre zuvor eingängig war (Trost 1976).

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 - nicht nur den Eltern - 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 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 Schüler 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.

Die Notenfrage ist inzwischen auch in den Schweizer Medien angelangt. So fragte eine Illustrierte Anfang August dieses 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,¹⁰ 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.
- 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.¹¹

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 wissen wollen, wo sie stehen (Bos u.a. 2010, S. 13). Sie lernen im Verbund immer unter der Voraussetzung einer Leistungshierarchie und nie nur für sich.

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:

¹⁰ Die Schule hat den LISSA-Preis 2016 erhalten (<https://www.eduzis.ch/einblicke/unsere-schule/einblicke-lissa-preis-seehalde/>)

¹¹ 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).

„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 Ziffernote. 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, 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 Kind 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.

- 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 Urteils macht der Lehrpersonen minimiert.
- Sie legen dann 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. Und wenn Eltern klagen, wird das als Bestätigung der Kritik verstanden, nicht als Angriff auf die Schule.

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 von dem Ideal des „guten Schülers“ der Forschung gar nicht erfasst.

Die Nachteile von Ziffernoten sind häufig beschrieben worden, die Vorteile dagegen eher nicht, einfach weil es ausschlaggebende Vorteile gar nicht geben sollte. Das ist an der Praxis vorbeigedacht und bleibt steril. Aber ausgerechnet die Lehrplanentwicklung scheint neue Möglichkeiten zu eröffnen. Mit Einführung des neuen Lehrplans wird auch in der Schweiz über Alternativen oder die Weiterentwicklung des Systems der Ziffernoten nachgedacht.

Die Bildungsstandards der Schweizer Volksschule sind mit Hilfe von Kompetenzstufen beschrieben worden und das hat die Frage aufgeworfen, ob man unabhängig davon noch Noten braucht, weil ja die Einstufung der Leistungen in den Lernfeldern genügt. Aber ein Notenersatz ist das nicht. Die Stufung in der Leistungsbeurteilung bleibt erhalten, damit auch die Gerechtigkeitsrisiken und die einzelnen Stufen können wie Noten gelesen werden. Auf der anderen Seite besteht die Gefahr, dass die Beschreibung der Stufen zur schematischen Einordnung der Leistungen genutzt wird. Allerdings werden Elternklagen erschwert.

Offenbar haben Frontalangriffe auf die gestufte Benotung nicht dazu geführt, das bewährte System abzuschaffen, im Gegenteil, Ziffernoten sind mit dem Bologna-Prozess auch im Tertiärbereich vorherrschend geworden. Jede Lernleistung wird bewertet, sie zählt buchstäblich für den Abschluss und muss daher ernst genommen werden. Diese enge Praxis hat Kritik ausgelöst und die Kritik wird auch für Anpassung sorgen, aber es ist kaum je eine Notenkritik. Noten gelten als Motivationsanreize und die Notenstufung an sich ist nicht strittig, nur wird sie selten voll ausgenutzt.

Notenfreie Module sind heute eine Grenzerfahrung, das kann man beklagen, aber muss auch zur Kenntnis nehmen, dass die notenfreie Universität dem Leistungsprinzip widersprechen würde. Kaum jemand fordert daher die Abschaffung der Noten, das scheint ein Grundschulphänomen zu sein. Und man stelle sich auch akademische 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, 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.

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 die Alternativen zu messen und so lange keine besseren Alternativen zum gestuften Ziffernsystem vorliegen, wird dieses auch weiterhin die Praxis bestimmen.

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.

Für die Bildungsstandards der Schweizer Volksschule sind mit den Kompetenzmodellen auch neue Aufgabenkulturen entwickelt worden, die sich von den klassischen Lehrplänen strukturell unterscheiden. Lehrpläne haben „Regelstandards“ beschrieben, die *alle* Schülerinnen und Schüler erreichen sollen, während tatsächlich immer Leistungsabstände das Verhalten bestimmen. Aus diesem Grunde werden in der Schweiz auch „Mindeststandards“ vorgeschrieben; alle Schüler müssen eine *bestimmte* Stufe erreichen, nicht den gesamten Lehrplan, was weder möglich noch sinnvoll wäre. Die Erreichbarkeit der Leistungsziele ist dann je nach Aufwand, Talent und Vorerfahrung unterschiedlich.

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 explizite 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,¹² steht nicht zur Diskussion und hätte auch keine Akzeptanzchancen. Im Unterschied zu den Vereinigten Staaten kann man in der Schweiz nicht an den Lehrern vorbei Politik machen.

Aber es geht auch ganz anders: In England ist künftig für das zentrale Examen am Schluss der Sekundarstufe I (EBacc)¹³ eine eigene Prüfungsbehörde zuständig, schon vorher waren die Abschlussprüfungen ausgelagert, aber die Schulen konnten zwischen verschiedenen Boards wählen und so auf die Leistungsfähigkeit ihrer Schüler Rücksicht nehmen. Der Grund für die Verschärfung ist die Vermeidung von Ungerechtigkeit; „gerecht“ soll sein, wenn ein einziges „examination board“ gleiche Aufgaben für alle stellt. Wer die zentrale Prüfung nicht schafft und sie wiederholen kann, wird aufwändig beraten, auch das soll die Bildungsgerechtigkeit erhöhen.

¹² Verschiedene Schuldistrikte in den Vereinigten Staaten, etwa in Cincinnati, koppeln die Löhne an die Ergebnisse der Leistungstest.

¹³ „English Baccalaureate Certificate“. Der Abschluss ist 2010 beschlossen und ab 2012 eingeführt worden. Er umfasst fünf Kernfächer, Englisch als Muttersprache, eine Fremdsprache, Mathematik, eine Naturwissenschaft sowie Geschichte oder Geografie.

In diesem System sind die Schulen also nicht mehr für die Erteilung der abschliessenden Noten verantwortlich, werden aber für die Ergebnisse zur Rechenschaft gezogen. Der Grundsatz, dass prüft, wer den Unterricht erteilt, wurde damit ausser Kraft gesetzt, was auch mit einem Misstrauen gegenüber den Lehrern und den Schulleitern erklärt werden kann. Klug ist das nicht, aber die Bürokratie freut sich über nicht limitierbare neue Aufgaben, die gerne auch als „Herausforderungen“ kommuniziert werden.

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 festlegt, was den Unterschied ausmacht. 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:

Noten:

- *Exzellent* (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 näher 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.

Die Kernfrage ist, *was* bewertet man *wie* mit einer Note oder anderen Formen der Leistungsbeurteilung. Diese Frage müssen die Schulen bearbeiten und zu überzeugenden Lösungen kommen. „Überzeugend“ ist eine Frage der Transparenz und der Kommunikation,

ausserdem der Ökonomie und des Zeitaufwandes. Gesetzliche Vorgaben dürfen den Spielraum der einzelnen Schulen nicht über Gebühr beschränken, aber klare Lernziele sind unerlässlich.

Das „was“ bestimmen in fachlicher Hinsicht der Lehrplan und das konkrete Curriculum der Fachschaften. Normen für die Noten direkt aus dem Lehrplan abzuleiten, ist weder sinnvoll noch praktikabel. Die Lehrpersonen müssen an dieser zentralen Aufgabe ihres Unterrichts direkt beteiligt sein. Insofern ist der Bezugspunkt nicht einfach der Lehrplan, sondern die konkreten Zielsetzungen der Schule, so wie die Fachschaften sie für einen bestimmten Zeitraum - ein Semester, ein Jahr oder länger - festgelegt haben.

- 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.

So können Noten brauchbare und aufschlussreiche Instrumente für das Leistungsfeedback sein. Sie sind dann weder Schicksal noch Machtdemonstration, sondern serielle Aufnahme des Leistungsstandes.

Eine Veränderung der Prüfungspraxis gelingt nur 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 blossen 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.

Die Auslagerung der Prüfungen an externe Agenturen ergibt keinen Gewinn und verlagert die Macht im Kernbereich der Schulautonomie. Solche Überlegungen liegen nahe und werden an öffentlichen Einfluss zunehmen, wenn sich die Fälle häufen, dass Noten willkürlich erteilt werden und ungerecht sind. Allein das sollte Anlass sein, die Prüfungspraxis zu verbessern und die Noten aussagekräftiger zu machen, ohne ihre Vorteile zu verspielen.

Gängige Ergänzungen zum Typus des Ziffernzeugnis sind individuelle Lernberichte, die die Lehrkräfte abfassen oder fortlaufende Portfolios, die von den Schülerinnen und

Schülern erstellt werden. Oft dokumentieren Lehrerinnen und Lehrer fortlaufend ihren Unterricht, notieren Auffälligkeiten und reflektieren ihre Massnahmen. Elaborierte Lernberichte sind aufwändig zu erstellen, sie sind zudem interpretationsabhängig und benutzen oft sprachliche Formen, die für die Lesenden gewöhnungsbedürftig sind. Sie können als zusätzliche Dokumente eingesetzt werden, allerdings muss geklärt werden, welcher rechtliche Status ihnen zu kommt.

Portfolios der Schüler dokumentieren die Themen und Gegenstände ihrer Lernarbeit und enthalten oft auch Lerntagebücher, mit denen die Arbeit reflektiert wird. Portfolios sind häufig sehr umfangreich, sie machen den Lernprozess auf Seiten der Schüler sichtbar und sollten als eigenständige Lernleistungen anerkannt werden. Als Anlage zu Zeugnissen haben sie illustrierenden Charakter, aber ohne wirkliche Benotung durch die Lehrkräfte sind sie für die Eltern, die künftigen Abnehmer und die Lernenden selbst zu 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 des eigenen Lernweges ist eine Sache, die Kriterien geleitete 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 „Klassenscockpit“ und „Stellwerk“, beides inzwischen bewährte Instrumente zu einer vergleichenden Leistungsbeurteilung, die über die Klassennorm hinausgeht und sich auf einen kantonalen Durchschnitt bezieht. Sie dienen nicht der Selektion, sondern der Verbesserung des Lehrerurteils.

Von besonderer Bedeutung für die Regelung der Schnittstelle zwischen Volksschule und Berufsbildung ist der Test „Stellwerk“. Er gibt an, wo die Schüler Mitte der achten Klasse in Kernfächern wie Mathematik und Deutsch stehen, also was sie erreicht haben und was nicht. Bis zum Ende der Schulzeit können dann Schwächen gezielt bearbeitet und die Lernziele erreicht werden. Voraussetzung ist ein direktes Feedback an Schüler und Eltern. Eine zentrale Prüfungsbehörde würde das nie können, was wiederum für die Stärkung der Verantwortung der Lehrer und die Unterstützung der Schulautonomie spricht.

In diesem Rahmen ist der Umgang mit solchen Instrumenten 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. „Effektstärken“ helfen dabei vermutlich nicht.

Eine andere Frage ist, ob nur die Schülerinnen und Schüler bewertet werden sollen. Von „Beurteilungskultur“ kann auch in einem weitergehenden Sinne gesprochen.

- Jeder Schüler und jede Schülerin beurteilt alle seine oder ihre Lehrkräfte, wengleich weitgehend informell und oft ohne Nutzen für den Unterricht.
- Studien zu den Erwartungen der Schülerschaft, darunter Absolventenbefragungen aus Zürich, zeigen, dass die Gymnasiastinnen und Gymnasiasten es inzwischen für eine professionelle Selbstverständlichkeit halten,
- dass *sie* mit einem formalisierten Verfahren ihre Lehrkräfte beurteilen müssten.

Soweit geht die Bereitschaft in vielen Schulen nicht, aber Beurteilung ist nicht mehr nur Sache der Lehrkräfte. Das lasse ich so stehen, es genügt der Hinweis, dass besonders Schweizer Berufsschulen mit „Schülerfeedback“ arbeiten und auch ohne Bestimmung der Effektstärke gute Erfahrungen machen.

Im Blick auf das Thema und die Fragestellung kann ich so schliessen: Bei der Bewertung von Leistungen in der Schule werden Noten und Zeugnisse als Formen des verbindlichen Feedbacks weiterhin eine 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.

Zur These verdichtet: Ohne Leistungsbeurteilung kann es keine Bildungsgerechtigkeit geben, weil man dann gar keine Leistungen hätte. Man kann die Frage der Bildungsgerechtigkeit aber auch noch anders verstehen, als Auftrag an die öffentliche Schule, nachhaltig Bildung vermitteln und dabei auf das Niveau zu achten. Nur das fordert und fördert wirklich.

- Der Gründer des Pädagogischen Seminars der Universität Leipzig, der Jurist und Philosoph Tuiskon Ziller (1884, S. 240),¹⁴ hatte dafür eine plausible Regel:
- Der Zögling, wie man die Schüler im 19. Jahrhundert nannte, „darf durch den Unterricht nicht geistig schwächer werden“.¹⁵

Literatur

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.

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

¹⁴ Tuiskon Ziller (1817-1882) war Gymnasiallehrer und habilitierte sich als Jurist. 1854 begann er mit Vorlesungen in Pädagogik, 1861 erfolgte die Gründung des Pädagogischen Seminars, das ein Jahr später mit einer Übungsschule verbunden wurde. Zillers *Einleitung in die Allgemeine Pädagogik* von 1856 ist die Begründungsschrift des Pädagogischen Herbartianismus.

¹⁵ Sperrung im Zitat entfällt.

- Notenvergabe durch die Lehrpersonen. In: Zeitschrift für Soziologie Band 41, Heft 2 (April 2012), S. 142-161.
- Fischer, Chr. (Hrsg.): Diagnose und Förderung statt Notengebung? Problemfelder schulischer Leistungsbeurteilung. Münster/New York/München/Berlin: Waxmann 2012.
- 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.
- 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)
- Olechowski, R./Rieder, K. (Hrsg.): Motivieren ohne Noten. Wien u.a.: Verlag Jugend und Volk 1990. (= Schule, Wissenschaft und Politik, Band 3)
- Ravitch, D.: The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education. New York: Basic Books 2010.
- 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.
- Ziller, T.: Grundlegung zur Lehre vom erziehenden Unterricht. 2. verb. Aufl. Hrsg. v. Th. Vogt. Leipzig: Verlag von Veit&Comp. 1884.