

Jürgen Oelkers

Noten und Leistungen: Befunde und praktische Vorschläge^{)}*

1. Die Frage der Wirksamkeit

Noten für Leistungen im Unterricht von Schulen sind umstritten, und dies nicht erst seit heute. Die drei zentralen Kritikpunkte lauten so:

- Einerseits wird gesagt, dass die Noten die Leistungen nicht adäquat abbilden,
- zweitens wird die mit der Notengebung verbundene Leistungshierarchie in Frage gestellt, die die Lernfortschritte der einzelnen Schülerinnen und Schüler nicht erfasst, sondern sich selbst fortschreibt,
- und drittens wird die Objektivität bezweifelt.

Noten sind Urteile von Lehrpersonen und nicht eine urteilt so wie die andere. Bei der Notenkritik wird fast immer übersehen, wie die Leistungen zustande kommen oder anders gesagt, was im Unterricht im Blick auf Leistungen wirksam ist und was nicht. Mit dieser Frage werde ich beginnen.

John Hattie von der University of Auckland hat im Jahre 2009 die bislang aufwändigste Metaanalyse von internationalen Wirksamkeitsstudien vorgelegt, die sich auf den Bereich Schule und Unterricht beziehen. Das Buch heisst *Visible Learning* (Hattie 2009) und der Titel ist Programm. Nur die Wirkungen machen sichtbar, was es mit einem pädagogischen Konzept auf sich hat, während wir zum Beispiel an die „kindzentrierte Pädagogik“ glauben sollen. Genauer: Wir sollen an die Wirksamkeit glauben, und dies als Alternative zur herkömmlichen Schule, deren Notenpraxis dann leicht als „kinderfeindlich“ hingestellt wird.

Hattie untersuchte über 800 Metanalysen der expansiven Wirksamkeitsforschung und legte so die erste grosse Meta-Metaanalyse der internationalen Bildungsforschung vor. Der Begriff „Wirksamkeit“ bezieht sich einzig auf *achievement*, also die messbaren Leistungen von Schülerinnen und Schülern. Die Stärke der Effekte („effect size“) wird nicht umgangssprachlich beschrieben, sondern von 138 einzelnen Faktoren aus errechnet, die zu sechs Gruppen zusammengefasst werden. Gefragt wird, wie sich die Gruppen bezogen auf die Leistungen auswirken.

Hattie unterscheidet:

- Die Schule als Organisation
- den Unterricht

^{*)} Vortrag auf der LUXDIDAC am 3. Oktober 2012.

- die Curricula
- die Lehrerinnen und Lehrer
- die Schülerinnen und Schüler
- die Familie und die soziale Herkunft
(ebd., S. 31).

Vergleicht man diese sechs Gruppen, dann ergibt sich ein klares und unstrittiges Ergebnis, das man auch ohne Hattie vermutet hätte: Die grösste Effektstärke kommt den *Lehrpersonen* zu. Von ihnen hängt es primär ab, welche Leistungen die Schülerinnen und Schüler zeigen, allerdings von ihnen nicht einfach als Personen, die Stellen besetzen, sondern unter der Voraussetzung eines elaborierten beruflichen Könnens und so eines Feldes von Faktoren.

Dieses Feld der Qualität der Lehrpersonen umfasst bei Hattie acht Punkte, auf die es ankommt. Die ersten drei werden wie folgt bestimmt:

- Die Qualität des Unterrichts, so wie die *Schülerinnen und Schüler* sie wahrnehmen.
- Die Erwartungen der Lehrpersonen.
- Die Konzeptionen der Lehrpersonen über Unterricht, Leistungsbeurteilung sowie über die Schülerinnen und Schüler.

Der letzte Punkt bezieht sich auf die Sichtweisen (views) der Lehrerinnen und Lehrer, etwa ob sie glauben, dass alle Schülerinnen und Schüler Fortschritte machen können und ob die Leistungen sich ändern können oder stabil bleiben. Ein Problem ist auch, wie der Lernfortschritt von den Lehrpersonen verstanden und artikuliert wird, also wem oder was sie den Fortschritt zuschreiben und wie die Lernenden davon in Kenntnis gesetzt werden. Wenn der entscheidende Faktor etwa in der sozialen Herkunft gesehen wird, dann ziehen Lehrkräfte andere Schlüsse, als wenn sie die Begabung im Vordergrund sehen.

Die fünf weiteren Punkte für den Einfluss der Lehrkräfte auf die Lernleistungen der Schülerinnen und Schüler sehen so aus:

- Die Offenheit der Lehrkräfte oder wie sie darauf eingestellt sind, sich überraschen zu lassen.
- Das sozio-emotionale Klima im Klassenzimmer, wo Fehler und Irrtümer nicht nur toleriert werden, sondern willkommen sind.
- Die Klarheit, mit der die Lehrpersonen Erfolgskriterien und Leistungsanforderungen artikulieren.
- Die Unterstützung der Lernanstrengung.
- Das Engagement aller Schülerinnen und Schüler
(ebd., S. 34).

Aber wenn heute in der angelsächsischen Diskussion gesagt wird, „teachers make the difference“, dann ist das nur ein Mantra. Der *Unterricht* macht den Unterschied, aber unterrichtet wird verschieden und nicht alle Lehrpersonen sind gleich erfolgreich in der Beförderung des Lernens. Bestimmte Lehrkräfte erfüllen die Aufgaben besser als andere und die kritische Frage ist, *bis zu welchem Grad* das der Fall ist. Deutlich wird gesagt: „Nicht alle Lehrerinnen und Lehrer unterrichten effektiv“, nicht alle sind „Experten für Lernen“ und nicht

alle haben grossen Einfluss auf die Lernenden. Die wichtige Frage ist, „in welchem Ausmass sie Einfluss auf die Leistungen haben und was den grössten Unterschied macht“ (ebd.).¹

Geht man nicht von einer Gruppe von Faktoren aus, sondern bezieht sich auf einzelne, dann sind die drei Faktoren mit der grössten Effektstärke:

- **Self-reported grades:** Die Einschätzung ihres aktuellen Leistungsstandes durch die Schülerinnen und Schüler (ebd., S. 43).
- **Piagetian programs:** Die altersgerechte Gestaltung des Unterrichts nach den Stufen von Piaget (ebd.).
- **Providing formative evaluation of programs:** Die ständige Erhebung des Lernfortschritts und die direkte Rückmeldung an die Schülerinnen und Schüler (ebd., S. 181).

Am anderen Ende der Skala sind Faktoren, deren Effektstärke bezogen auf die Lernleistung so schwach ist, dass sie mit grosser Wahrscheinlichkeit als wirkungslos gelten müssen. Diese Faktoren beziehen sich auf Konzepte, die in der Alternativpädagogik, aber auch in der heutigen Schulreformediskussion und nicht zuletzt in der Primarschulbildung hoch gehandelt werden, weil sie als besonders „kindgemäss“ gelten. Es handelt sich etwa um:

- **Multi-grade/multi-age classes:** Jahrgangsübergreifender Unterricht und altersdurchmischtes Lernen (ebd., S. 91ff.).
- **Student control over learning:** Selbstbestimmtes Lernen (ebd., S. 193/194).
- **Open vs. traditional:** „Offener“ versus „traditioneller“ Unterricht (ebd., S. 88/89).

Vergleicht man die Effekte von offenem und traditionellem Unterricht, dann sind die Unterschiede generell nicht sehr gross. Bezogen auf die Leistungen hat der traditionelle Unterricht leichte Vorteile, offener Unterricht wirkt eher in Bereichen wie Kreativität oder Selbstvertrauen, aber mit beiden lässt sich die Fragen nicht beantworten, was im Blick auf die Wirksamkeit den Unterschied macht (ebd., S. 89). Die Antwort ist nicht das Konzept, sondern die Qualität des Unterrichts.

Neben der Wirksamkeit von Methoden und Programmen lassen sich auch Faktoren bestimmen, die sich auf den letztendlichen Schulerfolg beziehen. Die negativsten Faktoren sind dabei

- Repetitionen von Jahrgängen (**retention**) (ebd., S. 97ff.),
- hoher Fernsehkonsum (**television**) (ebd., S. 67)
- und häufige Umzüge der Familie (**mobility**) (ebd., S. 81/82).

Im ersten Fall verliert man ein Jahr und gewinnt kaum etwas, im zweiten Fall distanziiert man sich von schulischen Lernformen, die Anstrengungsbereitschaft verlangen,

¹ „Not all teachers are effective, not all teachers are experts, and not all teachers have powerful effects on students. The important consideration is the extent to which they do have an influence on students achievements, and what it is that makes the most difference“ (Hattie 2009, S. 34).

und im dritten Fall muss Anschluss an ein fremdes System gefunden werden, das den Wechsel nicht belohnt, sondern eher bestraft. Im Übrigen haben auch Sommerferien negative Effekte, weil sie das Vergessen befördern (ebd., S. 81).

Schulaufsicht und Schulleitungen haben mittlere Effekte im Blick auf die Lehrkräfte (ebd., S. 83ff.) und das gilt auch für eine so umstrittene Grösse wie die Klassengrösse (ebd., S. 85ff.).² Die Programme der Lehrerinnen- und Lehrerbildung haben niedrige bis schwache Effekte, weil die Studierenden das Angebot nach dem Vorwissen sortieren, das sie mitbringen und sich so im Studium, wie es heute organisiert ist, selbst bestätigen (ebd., S. 109ff.). Ziele und Klarheit der Anforderungen sind Erfolgsfaktoren (ebd., S. 125ff., 163ff.), die direkte Instruktion hat nicht ausgedient (ebd., S. 204ff.) und das „problem-solving teaching“ ist effektiver als das „problem-based learning“ (ebd., S. 210ff.), weil das Problem nicht langwierig herausgefunden werden muss. Schliesslich: Kooperative Lernformen sind wirksamer als kompetitive (ebd., S. 212ff.).

Mit diesen Daten und Rangfolgen der faktoriellen Wirksamkeit lässt sich jedenfalls nicht mehr behaupten, „offene“ Lernformen seien „traditionellen“ grundsätzlich überlegen und es käme darauf an, überall jahrgangsübergreifenden Unterricht zu realisieren, der ja nur die Gruppierung der Schülerinnen und Schüler betrifft und nicht dadurch schon deren Leistung. Auch selbstbestimmtes Lernen, also Formen, in denen die Schülerinnen und Schüler über ihr Lernen in Grenzen selbst verfügen können, hat im Blick auf das Leistungsverhalten gänzlich andere Effekte, als die Alternativpädagogik annimmt.

Deutsche Studien zeigen ähnlich, dass offene Lerngelegenheiten im Blick auf individuelle Förderung wenn überhaupt, dann eher im affektiven Lernbereich wirksam sind und kognitive Kompetenzen dadurch kaum beeinflusst werden (Klieme/Warwas 2011). Fördern kann man auf verschiedene Weisen, die aber nie einfach *als Konzept* überlegen sind; auch Formen des Trainings oder adaptiver Unterricht sind im Blick auf das Lernen der Schülerinnen und Schüler nur dann wirksam, wenn sie didaktisch durchdacht und gut strukturiert sind. Das gilt auch für den Projektunterricht, der vielfach in der Realität, also in der Wahrnehmung der Lehrenden und Lernenden, gar nicht dem entspricht, was in der einschlägigen Literatur angenommen wird, nämlich „selbstgesteuertem Lernen“ (Traub 2011). Es heisst so, aber man macht es nicht.

Hattie plädiert für die Normalform eines von der Lehrperson vorbereiteten, strukturierten und realisierten Unterrichts. Der Lehrer und die Lehrerin wird nicht in die Rolle eines konstruktivistischen Beobachters versetzt, sondern darf und soll agieren, wobei für ihn oder für sie als Qualitätsannahme gilt, dass er oder sie „with the eyes of the students“ wahrnehmen kann (Hattie 2009, S. 238) und mit seinem Unterricht ihrem Lernen dienlich ist. Auf der anderen Seite müssen sich die Schülerinnen und Schüler als die „Lehrer ihrer selbst“ verstehen und dürfen die Verantwortung für den Lernerfolg nicht abwälzen. Die Psychologie des Konstruktivismus ist für Hattie eine Form des Wissens und nicht des Unterrichts (ebd., S. 243). Wer sich darauf beruft, kommt nie weiter als bis zur Alternativpädagogik (so schon Diesbergen 2000).

Aus Hatties Untersuchung geht erneut hervor, dass gelenkter und fordernder Unterricht bezogen auf die Leistungen effektiver ist als ungelenkter und erleichternder. Und die Resultate, also die Lernstände der Schülerinnen und Schüler, sind weitaus stärker von den

² „It appears that the effects of reducing class size *may* be higher on teacher and student work-related conditions, which *may* or *may not* translate into effects on student learning“ (Hattie 2009, S. 86).

Faktoren des Unterrichts beeinflusst als von denen der Schule oder gar der Schulstruktur. Das deckt sich weitgehend mit einer Längsschnittstudie aus dem Kanton Zürich, als deren Hauptergebnis angesehen werden muss, dass der *Unterricht* und genauer: die *Qualität* des Unterrichts den Unterschied macht (Nach sechs Jahren Primarschule 2011).

2. Lernleistung und Schulerfolg

Der Alltag heutiger Schulen ist gekennzeichnet von Heterogenität, und dies von Anfang an und in vielen Fällen unausweichlich. Die bildungspolitisch zentrale Frage ist, wie sich neue Konzepte wie das in der Didaktik viel zitierte „selbstregulierte Lernen“ auf diesen Tatbestand auswirkt, also ob die Risiken des Schulbesuchs dadurch minimiert werden können oder nicht. Reformpädagogische Methoden tragen nie einen Risikostempel, sie werden in bester Absicht verwendet und stossen aber auf eine harte Wirklichkeit, die den guten Absichten oft im Wege steht. Das Problem lässt sich an einem zentralen Bereich zeigen, nämlich die Entwicklung des Lernstandes über die Schulzeit.

Die Bildungsdirektion des Kantons Zürich hat im Jahre 2003 ein Forschungsprojekt begonnen, das eine Kohorte von anfänglich rund 2.000 Schülerinnen und Schülern während ihrer gesamten Schulzeit beobachtet. Die Schüler wurden bei Schuleintritt getestet und dann nachfolgend am Ende der dritten und sechsten Klasse, ein Test am Ende der neunten Klasse folgt im nächsten Jahr. Auf diese Weise soll die Entwicklung des Lernstandes erfasst werden. Dieser Längsschnitt ist einmalig in der Schweiz, Ansätze dazu gibt es inzwischen auch im deutschsprachigen Ausland. Die PISA-Studien erfassen demgegenüber nur Momente, aber nicht Prozesse, also nicht Veränderungen an verschiedenen Zeitpunkten.

Die Ergebnisse der ersten drei Studien liegen inzwischen vor (Moser/Stamm/Hollenweger 2005; Moser/Keller/Tresch 2008; Moser/Angelone/Keller/Hollenweger/Buf 2010) und lassen sich so zusammenfassen:

- Bei Schuleintritt haben etwa vier Fünftel der Kinder den Lernstoff der ersten Klasse in Mathematik teilweise bewältigt und ein knappes Fünftel befindet sich leistungsmässig bereits in der zweiten Klasse.
- Die Lesekompetenzen liegen weiter auseinander. 5 Prozent der Kinder kennen keine Buchstaben, etwa 35 Prozent erfassen die Verbindung zwischen Buchstaben und Lauten nur im Ansatz, 15 Prozent lesen bereits schwierige Wörter und längere Sätze.

Noch disparater ist der Wortschatz, also die Fähigkeit, Bilder mit eigenen Wörtern richtig zu bezeichnen.

- Nur knapp ein Drittel der Kinder beherrscht die richtigen Bezeichnungen,
- ein weiteres Drittel überbrückt geschickt die Lücken,
- das letzte Drittel beherrscht nur ganz elementare Bezeichnungen, also kann sagen, dass ein Hammer auf einem Bild ein „Hammer“ ist, mehr jedoch nicht.
- Korrekte Bezeichnungen für andere Bilder können nicht abgerufen werden.

Die zweite Studie nach drei Jahren Schulzeit zeigt, dass der Unterricht starke, aber auch ungleiche Effekte hat. Die Schülerinnen und Schüler erreichen wie gesagt die erste Klasse der Volksschule mit grossen Unterschieden im Lernstand.

- Unterschiede konnten in den ersten drei Jahren ausgeglichen werden, sofern qualitativ guter Unterricht stattgefunden hat.
- Er macht den Unterschied, was mit der Kompetenz der Lehrkräfte, der Zusammensetzung der Klasse, den je erreichten Lernfortschritten und der Unterstützung der Eltern zu tun hat.

Die dritte Studie nach sechs Jahren Schulzeit, die gerade veröffentlicht worden ist, zeigt, dass die Schülerinnen und Schüler mit Deutsch als Zweitsprache, die ihre Schulzeit mit einem geringen Wortschatz in der Unterrichtssprache begannen, diesen Rückstand inzwischen aufgeholt haben. Die Erstsprache ist kein Hindernis mehr, das Verhalten wird gesteuert durch die Akzeptanz der Schulkultur einschliesslich des Unterrichts in Deutsch als Zweitsprache von der ersten Klasse an. Auch ein altersbedingter Vorsprung schwindet im Laufe der Primarschulzeit.

Anders steht es mit Wissensvorteilen.

- Wer schon bei Beginn der Schulzeit über ein grosses Vorwissen verfügt hat, profitiert davon über die gesamten sechs Schuljahre.
- Noch besser lassen sich die Leistungen am Ende der Primarschule mit den Leistungen am Ende der dritten Klasse vergleichen.
- Wer hier gut war, bleibt gut - und umgekehrt.

Es gibt deutliche Risikogruppen, solche Kinder nämlich, die die Lernziele nicht erreicht haben und mit hoher Wahrscheinlichkeit im weiteren Verlauf ihrer Schulkarriere weiter zurückbleiben. Der bildungspolitische Slogan „No Child Left Behind“ erweist sich so einfach als Zweckoptimismus. Die Zürcher Studie wirft die Frage auf, wie mit der Risikogruppe umgegangen werden soll, und das verweist auf die Notwendigkeit einer konsequenten Leistungsentwicklung, für die der andere Slogan „Fördern und Fordern“ zur Verfügung steht. Der Slogan selbst hat keinen Inhalt und besagt eigentlich nur, dass irgendwie hälftig verfahren werden soll. Die wirklichen Probleme sind deswegen schwer zu bearbeiten, weil sie in den Klassen entstehen und trotz erheblichem Einsatz der Lehrkräfte auftreten.

Der Abstand zwischen den leistungsstarken und den leistungsschwachen Schülern muss im Verlauf der Schulzeit keinesfalls geringer werden, eher ist anzunehmen, dass er mit den fachlichen Anforderungen grösser wird. Bestimmte Gruppen von Schülerinnen und Schülern verbessern ihren Lernstand nicht, sondern bleiben auf dem einmal erreichten Niveau stehen.

- Die dritte Studie zeigt, dass die Zunahme der Leistungsunterschiede erst auf der Mittelstufe erfolgt, also nach der dritten Klasse.
- Bis dahin liegt der Anteil der Schülerinnen und Schüler, die die Ziele des Lehrplans in Deutsch und Mathematik nur teilweise erfüllen, bei 10 Prozent.
- Am Ende der sechsten Klasse sind das 17 Prozent für Deutsch und 18 Prozent für Mathematik, was deutlich eine Folge ist der steigenden Anforderungen.

Für die letzten Schuljahre ist eine weitere Öffnung der Leistungsschere zu erwarten. Für die leistungsschwächeren Schülerinnen und Schüler fehlen dann lohnende Aufgaben und Lernanlässe, die mehr sein müssen als die immer neue Bestätigung ihrer Schwächen. Dieses Problem der negativen Differenzierung über die Schulzeit lässt sich nicht einfach durch punktuelle Fördermassnahmen bearbeiten, sondern stellt sich im Blick auf den gesamten Verlauf einer Karriere als Schülerin und Schüler.

Wie schwer das ist, zeigen weitere Ergebnisse der dritten Studie: Die soziale Zusammensetzung der Klasse wirkt sich dann auf die Leistungen positiv aus, wenn die soziale Herkunft besonders hoch und so besonders homogen ist. Das sind 14 Prozent der Klassen im Kanton Zürich. Die Leistungsbeurteilung durch die Lehrkräfte kann diesen Vorteil durch strengere Bewertungsmaßstäbe in eine Benachteiligung verwandeln, was bei sozial belasteten Klassen umgekehrt gilt. Aus einem Bonus kann die Notengebung einen Malus machen und aus einem Malus einen Bonus.

Alle drei Studien verweisen darauf, dass die Entwicklung des Lernstandes zwischen verschiedenen Klassen auch ein- und derselben Schule höchst unterschiedlich sein kann. Das gilt ebenso für die Effekte der Förderung. In den einen Klassen gelingt es, die Leistungsschere zu verringern, in anderen jedoch nicht, was primär damit zu tun, dass sehr verschieden unterrichtet und gefördert wird.

- Verbunden damit ist allerdings auch ein strukturelles Problem, das deutlich benannt werden muss.
- Fortlaufender Unterricht, der in einem bestimmten Zeitraum und ohne zusätzliche Ressourcen Ziele erreichen soll, die für alle gleich gelten, kann nur sehr begrenzt Nachteile einzelner Schüler ausgleichen.
- Kinder bleiben auch deswegen zurück, weil erreichbare Ziele für sie gar nicht bestehen.

Herkunftsbedingte Ungleichheit können Schulen nicht einfach kurzfristig ausgleichen. Die Beherrschung der Unterrichtssprache dagegen kann durch gezielte und frühe Förderung gelernt werden. Daher ist eine konsequente Sprachförderung ein zentrales Mittel zur Gewährleistung von Erfolgen in der Schule. Das gilt ganz generell und besonders im Blick auf die zentralen Leistungsfächer Deutsch und Mathematik. Notwendig ist dafür auch eine gezielte Elternarbeit. Neuere deutsche Studien zeigen, dass generell mit der Dauer der Schulzeit die Motivation der Schülerinnen und Schüler und so ihre Leistungsbereitschaft abnimmt. Diese Tendenz wird vor allem dort gemindert, wo es einen vollumfänglichen Ganztagsbetrieb gibt. Das spricht dafür, mit der Ganztagsbeschulung weiter fortzufahren und sie auszubauen (Nachweise in Oelkers 2011).

Schulerfolg hängt wesentlich ab von der Beherrschung der Unterrichtssprache. Wer dem Unterricht sprachlich nicht folgen kann, hat in fast allen Unterrichtsfächern Nachteile. Auch die soziale Herkunft oder die Nähe bzw. Ferne zu den Bildungsanforderungen der Volksschule spielen eine wichtige Rolle. Wenn beides negativ zusammenkommt, ist trotz aller Anstrengungen die Wahrscheinlichkeit hoch, dass der Erfolg im Leistungsbereich ausbleibt. Alle Methoden müssen sich dieser Frage stellen und dürfen nicht lediglich als persönliche Vorlieben der Lehrkräfte Verwendung finden. Das gilt ebenso für Lehrpläne und Lehrmittel. Wie soll selbstreguliertes Lernen wirksam sein, wenn es für die leistungsschwächeren Schülerinnen und Schüler keine geeigneten Lehrmittel gibt?

Der bloße Wechsel der Sprache ist trügerisch, von der Rhetorik der Experten kann man nicht auf die Praxis schließen. „Kompetenz“ ist ein Modebegriff, der inzwischen für alles herhalten muss, was irgendwie innovativ klingt und doch selten das traditionelle Begriffspaar „Wissen und Können“ übersteigt. Die heutige „Kompetenzorientierung“ ist bis in die Buchtitel der Fachliteratur und die Lehrpläne der Schule vorgedrungen. Aber wenn man etwas anders bezeichnet, ist es nicht schon anders. Praktische Innovation ist weit schwieriger zu bewerkstelligen als semantischer Austausch und der Wandel der Rhetorik von Experten, mit dem oft nur Eindrucksmanagement verbunden ist. Aber was ist damit gewonnen, wenn alles, was als Zielformel gut klingen soll, „Kompetenz“ genannt wird?

Keine didaktische Theorie seit dem 19. Jahrhundert hat Unterrichtserfolg mit dem bloßen Nachvollzug von Schulwissen gleichgesetzt; es ging immer um das Verstehen, Durcharbeiten und Anwenden des Stoffes durch die Schüler, also um zunehmendes Können oder um stetig verbesserte Kompetenz.

- Der Gründer des Pädagogischen Seminars der Universität Leipzig, der Herbartianer Tuiskon Ziller (1884, S. 240),³ hatte dafür auch eine plausible Regel:
- Der Zögling, wie man die Schüler im 19. Jahrhundert nannte, „darf durch den Unterricht nicht geistig schwächer werden“.⁴

Wie oft das der Fall war, ist wiederum nicht untersucht, aber das Problem der Nachhaltigkeit des Unterrichts ist nicht neu, und auch Zillers Lösung, dass die Schulübungen verschwinden, „wenn nur die Lehrer bessere Methoden zu gebrauchen wüssten“ (ebd., S. 243),⁵ ist bis heute angesagt. Von besseren *Prüfungen* ist dabei nicht die Rede, auch die heutige Ausbildung meidet dieses Thema und gute Fachliteratur oder auch nur ein Vademecum scheint es nicht zu geben.

Unterricht und fortlaufende Leistungsüberprüfung machen das Kerngeschäft der Schule aus. In Schweizer Gymnasien gibt es im Blick auf die Prüfungspraxis wohl formal sehr genau festgelegte Verfahren, etwa bei der Verrechnung der einzelnen Noten, aber was die Noten inhaltlich bedeuten, ist weitgehend offen und wird von jeder Lehrkraft selbst festgelegt. Es gehört zur Berufsidentität von Lehrerinnen und Lehrern, über die Leistungen der Schüler in ihrem Fach autonom und nach eigenem Urteil bestimmen zu können. Absprachen über die Noten oder gar eine genaue Beschreibung der Notenstufen sind bislang weitgehend unbekannt, was sich nicht allein für die Gymnasien sagen lässt. Prüfen nach *gemeinsamen* Standards für die Noten gibt es in den wenigsten Schulen.

3. Kritik der Notengebung

Die Praxis der Leistungsüberprüfung in den Schulen ist jahrzehntelang ebenso intensiv wie erfolglos kritisiert worden. Die Argumente sind bekannt. Die Notengebung sei rein

³ Tuiskon Ziller (1817-1882) war Gymnasiallehrer und habilitiert sich als Jurist. 1854 begann er mit Vorlesungen im 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.

⁴ Sperrung im Zitat entfällt.

⁵ Sperrung im Zitat entfällt.

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 (so schon Hofer 1975). 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. Unterstellt wird die Vergleichbarkeit der Notengebung an allen Orten, während die tatsächlichen Bewertungen von Schule zu Schule und von Fach zu Fach variieren können.

Angesichts der Fehlerquellen 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 Döring (1925, S. 177) schon vor mehr als achtzig Jahren. Aber selbst wenn die Behauptung zuträfe, so bliebe die Frage bestehen, warum dann immer noch weitgehend Ziffernnoten die Praxis der Leistungsbeurteilung bestimmen. Es ist mindestens frappierend, wie vergleichbar gering die Entwicklungsarbeit gerade auf diesem Gebiet gewesen ist und wie stark die offenkundigen Vorteile der Notenskala die Praxis bestimmt haben.

Aber Noten und Zeugnisse, die die Schule vergibt und die mit Berechtigungen verbunden sind, haben in der pädagogischen Diskussion keinen guten Ruf. Schon Ende des 19. Jahrhunderts sind erste Befunde vorgelegt worden, die den Verdacht genährt haben, dass Noten ohne Bezugsnormen subjektiv und so beliebig sind und Zeugnisse wenig aussagen. Auf der anderen Seite ist die Beschreibung schulischer Leistungen mit einer Skala von Noten ein bewährtes Instrument, das in der Öffentlichkeit fest verankert ist und zu den stärksten Erwartungen an die allgemeinbildende Schule gehört. Selbst wenn man Noten umgehen will, sind sie präsent.

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 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 Kritik bemängelt nur die mangelnden Bezugsnormen (Fischer 2012, S. 50).

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 erste umfassende Reglement zur Steuerung von Schule und Unterricht mit verbindlichen Standards. Der „Schulmethodus“, also der Weg der Steuerung, ist vom Rektor des Gymnasiums in Gotha verfasst worden und erschien zuerst 1642. In der Ausgabe des Jahres 1662 wird ein Stufenschema zur Beurteilung der Leistungen in drei Bereichen vorgeschrieben. „Ingenium“ ist Begabung, „mores“ ist Verhalten und „Unterrichtsgegenstände“ sind die Lernbereiche. Allerdings gab es noch keine Ziffernnoten.

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.

Diese gestufte Form der Beschreibung hat sich in allen Schulsystemen durchgesetzt und ist mit dem Bologna-Prozess nochmals verstärkt worden. Das Prinzip der Skalierung ist gut geeignet, die tatsächlich gezeigten Leistungen in einer Lerngruppe abzubilden. Es gibt in jedem Klassenverband eine Normalverteilung mit eher geringen Spitzenwerten, einem hohen Durchschnitt und einem etwa zehnpromtigen Bestand von Leistungsschwachen. In diesem Sinne bestätigt der Pisa-Test jedes Mal die historische Form der Leistungsbeschreibung, denn auch hier wird mit gestuften Modellen der Kompetenzentwicklung gearbeitet.

Eine andere Frage ist, ob nur die Schülerinnen und Schüler bewertet werden sollen. Es gibt in jeder Schule eine Beurteilungskultur, die verschieden geartet ist, jedoch in aller Regel von den Beteiligten als bewährt angesehen und als Standard akzeptiert wird. Das gilt angesichts klarer Normen und Verfahren besonders für die Gymnasien und Berufsschulen. Von „Beurteilungskultur“ kann auch in einem weitergehenden Sinne gesprochen werden.

- 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 Stichwort „Evaluation“ ist inzwischen auch im Bereich der Gymnasien kein Fremdwort mehr. Die Lehrkräfte müssen sich auf eine neue Situation ihres Berufsfeldes einstellen, die sich im Alltag zeigen wird. Die Schulen selbst sind Teil und Objekt der Beurteilungskultur. Auch die Schweizer Gymnasien und Berufsschulen werden in Zukunft regelmässig extern evaluiert werden und die Daten zur Schulentwicklung nutzen.⁶ Evaluieren kann man aber nur, wenn Standards gegeben sind und beachtet werden.

Noten sind einfache Instrumente, deren Aufwand begrenzt ist und die leicht kommuniziert werden können. Sie sind gesellschaftlich fest verankert und werden als

⁶ Ein Angebot ist die *Interkantonale Fachstelle für externe Schulevaluation auf der Sekundarstufe II* (IFES) an der Universität Zürich. Das Institut führt Primär- und Metaevaluationen durch.

Beschreibungsform auch ausserhalb der Schule akzeptiert. Ihre Vorteile sind nie untersucht worden, einfach weil die Forschung lange davon ausging, die Notengebung mit einfachen Ziffern könnte gar keine Vorteile haben. Inzwischen sind aber auch andere Einschätzungen zu vernehmen. „Im Durchschnitt betrachtet“, heisst es in einem jüngeren 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. 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“ (ebd.). Eine Note kann also an einem Ort etwas Anderes bedeuten als an einem anderen. Ziffern wie „fünf“ oder „vier“ erfassen in verschiedenen Schulen und Klassen nicht dasselbe, während genau das der Eindruck ist, den die Noten vermitteln. Sie werden ohne ihren Entstehungskontext kommuniziert.

Aus diesem Befund leiten sich verschiedene Strategien ab, wie die Praxis der Notengebung verbessert werden kann.

- Die subjektive Beurteilung muss durch eine objektive ersetzt werden, die Standards oder einen übergreifenden „Bezugsmassstab“ voraussetzt.
- Das objektive Kriterium sind die vom Lehrplan geforderten Lernziele, unabhängig von den in einer Klasse vorhandenen Leistungsunterschieden.
- Dafür müssen am Lehrplan orientierte diagnostische Instrumente entwickelt werden, die die Noten ergänzen oder gar ersetzen.
(ebd.).

Die beiden ersten Forderungen liegen in der Schweiz dem so genannten HarmoS-Konkordat zugrunde, das verbindliche Bildungsstandards für die Volksschule vorsieht (Maradan/Mangold 2005). Die Forderungen sind allerdings nicht leicht umzusetzen und verlangen neuartige Strategien der Implementation, die mehr darstellen müssen als die Konstruktion eines gemeinsamen Lehrplans, etwa die Anpassung der Lehrmittel oder die Nutzung von Leistungstests für das Geschehen im Klassenzimmer (Oelkers/Reusser 2008).

Die dritte Forderung bezieht sich auf 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 gymnasiale Oberstufe kaum anwendbar wäre. Im Mittelpunkt der Beurteilung und so der Notengebung steht der Fachunterricht.

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.

Bewertungen in der Schule sind vielfältiger Art. 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.).

4. *Noten in der Praxis*

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 zustandekommen. 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.

Das Prinzip der Stufung wird auch in den Bildungsstandards für die Schweizer Volksschule deutlich. Für sie werden Kompetenzmodelle entwickelt, die davon ausgehen, dass sich die Leistungen auf verschiedenen Stufen erfassen lassen. Die Erreichbarkeit der Leistungsziele ist je nach Aufwand und Talent unterschiedlich. Das unterscheidet die neuen Aufgabekulturen 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 Harmo-Konkordat „Mindeststandards“ vor; alle Schüler müssen eine *bestimmte* Stufe erreichen, nicht den gesamten Lehrplan, was weder möglich noch sinnvoll wäre.

Wenn es in Zukunft auch in der Schweiz vergleichende Leistungstests geben wird und verbindliche Standards für die Volksschule eingeführt werden, dann stärkt das in gewisser Hinsicht den Primat des Fachunterrichts. Auch der Kompetenzaufbau wird verbessert, weil und soweit die Leistungsniveaus getestet werden. Der läuft der Alternativpädagogik zuwider, die auf jeglichen Leistungsvergleich am liebsten verzichten würde. Aber jedes Kind vergleicht sich in seinem Leistungsverhalten mit anderen Kindern.

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 Prüfungen auch nicht ersetzen, sondern allenfalls ergänzen. An flächendeckende Tests ist derzeit nicht gedacht, und auch ein „High-Stakes-Testing,“ also Tests mit Folgen für die Schule oder gar für die einzelnen Lehrkräfte,⁷ steht nicht zur Diskussion. Oft werden Leistungstest so verstanden, als würden sie wesentlich nur in dieser Funktion verwendet, was jedoch nicht der Fall ist

Einen Einblick in die Prüfungspraxis der Schweizer Gymnasien hat eine landesweite Studie 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 Landessprache keinen Unterschied, solange beide Sprachen lediglich als Schulfächer⁸ 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

⁷ Einzelne Schuldistrikte in den Vereinigten Staaten, etwa in Cincinnati, koppeln die Löhne an die Ergebnisse der Leistungstests.

⁸ Zu unterscheiden von längeren Sprachaufenthalten oder von der zweisprachigen Matur.

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).

Wohin die Entwicklung gehen könnte, lässt sich am Projekt „Hochschule:Gymnasium“ (HSGYM) im Kanton Zürich zeigen. Arbeitsgruppen der Zürcher Gymnasien und der beiden Universitäten haben diejenigen Fachinhalte bestimmt, die bei Aufnahme eines Studiums vorausgesetzt werden müssen. Es handelt sich nicht um „Bildungsstandards“ im Sinne von HarmoS, aber es sind doch klare Festlegungen der Gehalte, die der gymnasiale Unterricht vermitteln soll. Tätig war keine staatliche Kommission, sondern die direkt Betroffenen. Damit wird die allgemeine Hochschulreife als Universitätszugang gewahrt, die Verantwortung für Inhalte aber immens gesteigert. Letztlich entsteht so ein gemeinsamer Lehrplan von Oberstufe und Eingangssemester.⁹

Ein zentrales Stichwort lautet „accountability“, die Verantwortung der Schule für das Zustandekommen und die Bewertung der Leistungen ihrer Schüler. Natürlich hängt der Erfolg des Unterrichts auch davon ab, wie die Schülerinnen und Schüler ihre Lernressourcen einsetzen und wie sie das schulische Angebot nutzen. In diesem Sinne ist die Verantwortung immer geteilt; wenn Schüler eine minimalistische Strategie der Leistungserfüllung wählen, ist das ihre Entscheidung und nicht die der Schule. Sie muss nur klar machen, was die Konsequenzen einer solchen Lernstrategie sind, nämlich schlechtere Bewertungen und gegebenenfalls das Nichterreichen des Leistungsziels.

In diesem Sinne sind „eigenverantwortete Schulen“ immer solche, die die Verantwortung teilen und zwischen den verschiedenen Akteuren festlegen. Nicht teilbar ist die Verantwortung der Schule für die Leistungsziele und die Art und Weise der Leistungsüberprüfung. Beides zusammen stellt das Angebot der Schule dar, das möglichst transparent dargestellt und kommuniziert werden muss. Grundlegend sind klare Ziele und durchschaubare Kriterien, die eindeutig beschrieben werden und so auch Geltung finden. Die Schulen müssen sich auf Ziele beziehen, die sie auch erreichen können, und sie müssen ihre Entwicklungsprozesse regelmässig bilanzieren, damit sie wissen, wo sie stehen.

5. Zur Entwicklung des Prüfungswesens

Wenn die Prüfungspraxis optimiert werden soll, dann muss die Schule das zu einem vorrangigen Entwicklungsziel erklären und die Implementation zur gemeinsamen Sache der Akteure machen. Das verlangt Führungsarbeit seitens der Schulleitung, das System selbst wird 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

⁹ <http://www.educ.ethz.ch/hsgym/index>

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:

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 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. HSYGM läuft auf ein solches System der fachlichen Abstimmung hinaus; wenn mit der Umsetzung begonnen wird, stellt sich unmittelbar die Frage der gemeinsamen Kriterien auch für die Bewertung. Eine „kriteriale Norm“ für die Noten direkt aus dem Lehrplan abzuleiten, ist weder möglich noch praktikabel. Insofern ist der Bezugspunkt nicht einfach der Lehrplan, sondern der gegebene schulische Rahmen, also die Unterrichtspraxis. Im Kanton Zürich haben die Gymnasien eigene Lehrpläne, aber auch die müssen im Blick auf die Praxis konkretisiert werden.

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 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 kann 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 neuere Kritik verweist aber zu Recht 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 fachübergreifende Kompetenzen wie allgemeine Leistungsfähigkeit, schulisches Engagement und die Motivation für bestimmte Unterrichtsfächer.

- 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.

Zeugnisse basieren auf Noten. Noten sind 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. Hier kommt es darauf an, dass Mindeststandards erreicht werden; nach oben sind die Leistungen immer offen. Allerdings sind mit Noten nicht bestimmte Fähigkeiten und Fertigkeiten beschrieben. Sie beziehen sich auf Leistungen im Blick auf Aufgaben in bestimmten Fächern oder Lernbereichen.

Eine Note etwa in Englisch fasst die innerhalb einer bestimmten Zeit gezeigten Leistungen zusammen. Bezogen auf die Sprachfächer sind die meisten Schweizer Kantone im Blick auf die Volksschule dazu übergegangen, die Noten in Zeugnissen mit Fähigkeiten in Einzelbereichen zu verknüpfen. In den Sprachfächern wird unterschieden zwischen

- Hörverstehen
- Leseverstehen
- Sprechen
- Schreiben

Für andere Volksschulfächer sind Differenzierungen schwieriger und auch in den Sprachfächern wird letztlich eine Note erteilt. Ohne verbindliche Operationalisierung sind die vier Bewertungsbereiche der Sprachfächer zudem schwer handhabbar. Und es besteht die Gefahr, kommunikative Fähigkeiten gleich oder höher zu bewerten wie den korrekten Ausdruck.

Das geltende Zeugnis für die Sekundarstufe des Kantons Zürich basiert auf einer Operationalisierung. Das Zeugnis unterscheidet - wie der Gothaer-Schulmethodus - fünf Stufen, die wie folgt gefasst werden.

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

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

Als Ergänzung zur Bewertung durch die Lehrpersonen empfehlen sich Leistungstests, 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. Von besonderer Bedeutung für die Regelung der Schnittstelle zwischen Volksschule und Berufsbildung ist der Test Stellwerk.

In einer im Dezember 2009 vorgelegten Evaluation der Nutzung des Tests im Kanton St. Gallen werden Ergebnisse vorgestellt, die auf eine deutliche Qualitätsverbesserung hinweisen. Schulleitungsmitglieder und Lehrpersonen geben an:

„Am häufigsten wird Stellwerk eingesetzt zur Messung des Leistungsstandes in den Fachbereichen, zur standardisierten Beurteilung von aussen, als Leistungsvergleich innerhalb des Kantons und der Klasse und als ... Basis für Elterngespräche (vor allem Stellwerk 8 bei den Lehrpersonen)“ (Goetze, Denzler, Wissler 2009, S. 7).

Im Blick auf die Schülerinnen und Schüler wird festgehalten:

„Auch hier ergibt sich ein klares Bild: Die Lernenden sehen Stellwerk in erster Linie als Möglichkeit, die eigenen Kenntnisse in den wichtigsten Fächern zu prüfen und zu sehen, in welchen Fächern sie sich verbessern können“ (ebd., S. 8).

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 Leistungstests 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:

- Es wird auf ein gemischtes Instrumentarium hinauslaufen,
- bei dem Noten und Zeugnisse weiterhin eine zentrale Rolle spielen,
- *ergänzt* um Daten aus Leistungstests und Demonstrationen des fachlichen und überfachlichen Könnens der einzelnen Schülerinnen und Schüler,
- das sich an der Beschreibung ihres Lernweges in Portfolios aufzeigen lässt.

Ganz so beruhigend endet mein Vortrag aber nicht. Heute kündigt sich ein grundlegender Wandel der schulischen Lernkultur an, bei dem das Internet der Treiber sein wird. Dieser Wandel hat Auswirkungen sowohl auf die Lernzeit als auch auf die Erfassung der Leistungen. Die heutige Diskussion über Tages- oder Ganztagschulen geht davon aus, dass Unterricht in der Form von Lektionen erteilt und dann für den Tagesbetrieb sinnvoll ergänzt werden muss. Aber das selbstorganisierte Lernen mit dem Laptop und gesteuert durch schnell erneuerbare Aufgabekulturen stellt genau diese Prämisse in Frage, einfach weil die Lernzeit individuell werden kann und die Leistungsentwicklung transparent ist. Aber das wäre der nächste Vortrag.

Literatur

- Diesbergen, C.: Radikal-konstruktivistische Pädagogik als problematische Konstruktion. Eine Studie zum Radikalen Konstruktivismus und seiner Anwendung in der Pädagogik. 2. Aufl. Bern et. al.: Peter Lang Verlag 2000. (= Explorationen. Studien zur Erziehungswissenschaft, hrsg. v. J. Oelkers, Band 22)
- Döring, O.: Untersuchungen zur Psychologie des Lehrers. Leipzig 1925.
- Evaluation der Maturitätsreform 1995 (EVAMAR). Im Auftrag der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) und des Bundesamtes für Bildung und Wissenschaft (BBW). Neue Fächerstruktur - Pädagogische Ziele - Schulentwicklung. Schlussbericht der Phase 1. Bern; EDK/BBW 2004.
- Fischer, Chr. (Hrsg.): Diagnose und Förderung statt Notengebung? Problemfelder schulischer Leistungsbeurteilung. Münster/New York/München/Berlin: Waxmann 2012.
- Goetze, W./Denzler, N./Wissler, P.: Evaluation Stellwerk. Thalwil: Büro für Bildungsfragen 2009.
- Hattie, J.: Visible Learning. A Synthesis of Over 800 Meta-Analyses Relating to Achievement. London/New York: Routledge 2009.
- Hofer, M.: Die Validität der impliziten Persönlichkeitstheorie von Lehrern. In: Unterrichtswissenschaft Heft 2 (1975), S. 5-18.
- Klieme, E./Warwas, J.: Konzepte der individuellen Förderung. In: Zeitschrift für Pädagogik Beiheft (2011).
- Maradan, O./Mangold, M.: Bildungsstandards in der Schweiz. Das Projekt HarmoS. In: ph/akzenteHeft 2 (2005), S. 3-7.
- Moser, U./ Stamm, M./Hollenweger, J.: Für die Schule bereit? Lesen, Wortschatz, und soziale Kompetenzen bei Schuleintritt. Oberentfelden: Sauerländer Verlage AG 2005.

- Moser, U./Keller, F./ Tresch, S.: Schullaufbahn und Leistung. Bildungserfolg und Lernverlauf von Zürcher Schülerinnen und Schülern am Ende der dritten Volksschulklasse. Bern: h.e.p. Verlag 2008.
- Moser, U./Angelone, D./Keller, F./Hollenweger, J./Buff, A.: Lernstandserhebung am Ende der 6. Klasse. Schlussbericht zuhanden der Bildungsdirektion. Unveröff. Ms. Zürich: Institut für Bildungsevaluation/Pädagogische Hochschule Zürich 2010.
- Nach sechs Jahren Primarschule. Lernstand der Schulanfängerinnen und Schulanfänger von 2003 vor ihrem Übertritt in die Sekundarstufe I. Zürich: Bildungsdirektion Kanton Zürich 2011.
- Oelkers, J.: Die Qualität der Gymnasien. Eine Expertise für den Kanton Zürich. Bern: h.e.p. Verlag 2008.
- Oelkers, J.: Expertise zum Thema: „Ganztagsschule“ in der Ausbildung der Professionen. Zuhanden der deutschen Kinder- und Jugendstiftung. Zürich: Institut für Erziehungswissenschaft 2011.
- Oelkers, J./Reusser, K.: Expertise „Qualität entwickeln - Standards sichern - mit Differenz umgehen.“ Berlin: Bundesministerium für Bildung und Technologie 2008.
- Traub, S.: Selbstgesteuert lernen im Projekt? Anspruch an Projektunterricht und dessen Bewertung aus Sicht von Lehrenden und Lernenden. In: Zeitschrift für Pädagogik Band 57, Heft 1 (2011), S. 93-113.
- Tresch, S.: Potenzial Leistungstest. Wie Lehrerinnen und Lehrer Ergebnisrückmeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen. Bern: h.e.p. Verlag 2007.
- Weinert, F.E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim/Basel: Beltz Verlag 2001.
- Ziller, T.: Grundlegung zur Lehre vom erziehenden Unterricht. 2. verb. Aufl. Hrsg. v. Th. Vogt. Leipzig: Verlag von Veit&Comp. 1884.