

Jürgen Oelkers

Beurteilen und autonomes Lernen im Gymnasium^{)}*

„Bildung Schweiz“ ist die Zeitschrift des Schweizerischen Lehrerinnen- und Lehrerverbandes. Sie erscheint monatlich. Im November-Heft des letzten Jahres findet sich in der Rubrik „aktuell“ ein Artikel, der überschrieben ist mit: „Noten sind einfach und vertraut, aber mangelhaft.“ Den Noten wird also eine Note gegeben. Eingeleitet wird der Artikel mit folgendem Hinweis:

„Vorstösse aus rechtspolitischen Kreisen fordern in verschiedenen Kantonen die Rückkehr zu Zeugnisnoten nach traditionellem Muster. Fachleute aus Wissenschaft und Pädagogik hingegen machen seit Jahren auf die groben Fehler, auf Mängel und kritische Nebenwirkungen der Notengebung aufmerksam“ (Bildung Schweiz 2006, S. 16).

Sie haben sich einen Fachmann „aus Wissenschaft und Pädagogik“ eingeladen, der nicht aus „rechtspolitischen Kreisen“ stammt und sich gleichwohl erlaubt, von der LCH-Meinung abzuweichen. Hinweise auf die „groben Fehler, Mängel und kritischen Nebenwirkungen“ der Notengebung können im deutschen Sprachraum seit Ende des 19. Jahrhunderts nachgewiesen werden. Notenfreie Schulen gibt es heute wenn, dann im Primarschulbereich, zumeist in den unteren Klassen. In der Praxis der Sekundarschulen, in den Gymnasien oder Berufsschulen und in allen Studiengängen der Universität sind „Zeugnisnoten nach traditionellem Muster“ fest verankert. Die Promotionsfeier der Philosophischen Fakultät, der ich die Ehre habe anzugehören, verkündet in jedem Semester die Ergebnisse der Doktorprüfungen in Noten, ohne dass je ein Experte auf die kritischen Nebenwirkungen hingewiesen hätte.

Es gibt in jeder Schule eine Beurteilungskultur, die verschieden geartet ist, jedoch in aller Regel von den Beteiligten als bewährt angesehen wird. Das gilt angesichts klarer Normen und Verfahren mindestens für die Gymnasien und Berufsschulen. Von „Beurteilungskultur“ kann aber auch in einem weitergehenden Sinne gesprochen werden. Jeder Schüler und jede Schülerin beurteilt alle seine oder ihre Lehrkräfte, wenngleich weitgehend informell und 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. „Evaluation“ ist auch im Bereich der Gymnasien kein Fremdwort mehr. Die Lehrkräfte müssen sich auf eine neue Situation ihres Berufsfeldes einstellen, die sich auch im Alltag zeigen wird. Auch die Schulen sind Teil oder Objekt der Beurteilungskultur. Die Schweizer Gymnasien und Berufsschulen werden in Zukunft die Gelegenheit wahrnehmen können, sich regelmässig extern evaluieren zu lassen.

^{*)} Vortrag in der Kantonsschule Olten am 24. August 2007 in Olten.

Evaluieren kann man nur, wenn Standards gegeben sind und beachtet werden. Die politische Diskussion erweckt gelegentlich den Eindruck, dass für die Gymnasien Standards eine *neue* Forderung seien und dass Entwicklungen in diese Richtung sie bislang kaum berührt hätten. Aber das ist natürlich nicht so, Schulen und so auch Gymnasien *bestehen* aus Standards.

- Die Notenskala ist ebenso ein „Standard“ wie der Gebrauch von Lehrmitteln oder die Lehrer-Schüler-Relation.
- Ohne das gemeinsame Instrument der Notenskala könnte es in einer Schule keine Leistungsbeurteilung geben.
- Lehrmittel standardisieren in gewisser Hinsicht die Gehalte des Unterrichts
- und die Relation zwischen Lehrern und Schülern ist sogar ein statistischer Standard.

Die heutige Diskussion verwirrt diesen Tatbestand und suggeriert, dass Bildungsstandards etwas grundlegend Neues seien. Das ist nicht der Fall. Gymnasien unterscheiden sich von anderen Schulen nicht nur durch das Fächerangebot, sondern zugleich durch die Niveauforderung. Lehrkräfte des Gymnasiums benutzen den Ausdruck „gymnasiale Standards“ wie selbstverständlich, wenn sie die Besonderheit ihres Unterrichts aufzeigen wollen. Es macht für sie geradezu die Identität der Gymnasien aus, dass sie spezielle Anforderungen stellen können. Wenn der Ausdruck „Bildungsstandards“ gerade bei Gymnasiallehrkräften oft negativ besetzt ist, dann hängt das mit Befürchtungen zusammen, die eine bildungspolitische Diskussion ausgelöst haben. Der Tatbestand von Standards an sich ist davon nicht berührt.

Angesichts der gelegentlich etwas irreführenden Diskussion halte ich eingangs fest:

1. Jedes Gymnasium hat und vertritt *Standards*, versteht man darunter die Qualitätsanforderungen des Fachunterrichts.
2. *Gymnasiale Bildung* zeigt sich als fachliche und daneben auch als überfachliche Kompetenz, die im Unterricht aufgebaut wird.
3. Die Schülerinnen und Schüler beherrschen fachliche Anforderungen unterschiedlich, die Unterschiede zeigen sich in *Leistungsstufen*.
4. *Tests* und *Noten* halten auch bisher schon Fortschritte, Stagnation und Rückschritte im Leistungsverhalten fest.

Soweit gibt es nichts Neues. Gymnasien erteilen mehr als andere Schularten standardbasierten Unterricht, der sich auf Fächer bezieht und fachliches Lernniveau abverlangt. Am Ende eines jeden Lehrgangs stehen Kompetenzen, allerdings nie bei allen Schülern gleich. Das zu sagen, ist trivial und erhält nur deswegen eine herausgehobene Bedeutung, weil die Diskussion über Bildungsstandards neue Begriffe, manche sagen auch einen neuen Jargon benutzt. Statt wie früher „Fähigkeiten und Fertigkeiten“ spricht man heute von fachlichen und überfachlichen „Kompetenzen“, statt Leistungsanforderungen ist von „Standards“ die Rede, und statt „Niveauunterschied“ wird der Ausdruck „Kompetenzstufen“ benutzt. Die neue Sprache allein verweist aber nicht schon auf etwas, das in der Praxis neu wäre. Wozu dann aber die ganze Aufregung?

Noten und Zeugnisse geben tatsächlich immer wieder Anlass zur Diskussion, meistens geht es um die Objektivität der Beurteilung, also darum, welcher Massstab verwendet wurde. Massstäbe heißen heute „Standards.“ Dass heute intensiv über „Bildungsstandards“ diskutiert

wird, hat seinen Grund nicht zuletzt darin, dass im Blick auf Themen des Unterrichts und Kriterien der Notengebung Beliebigkeit oder mindestens fehlende Abstimmung unterstellt wird. Im Folgenden werde ich zunächst auf einige Befunde der Forschung eingehen, die sich auf die Praxis des Benotens beziehen (1). In einem zweiten Schritt stelle ich dar, was wir über die Beurteilungskultur in Schweizer Gymnasien wissen (2). Und abschliessend diskutiere ich Beispiele für autonomes Lernen und frage, wie sich das mit der Qualitätssicherung verträgt (3).

1. Die Praxis des Benotens

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. Zentral sind bis heute Noten, 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 Niveau- oder Kompetenzunterschieden zustande kommen. Wer sie abbilden will, muss ein gestuftes Schema verwenden.

Historisch wurde als Notensystem fast immer eine Skala von fünf oder sechs Stufen verwendet.¹ Dieser Befund ist sehr stabil. Der „Gothaer Schulmethodus“ von 1642² sah zum ersten Male im deutschen Sprachraum ein einheitliches Schema für die Bewertung der Leistungen im Blick auf *Unterrichtsgegenstände* vor. Gleichzeitig sollten *Ingenium* und *Mores*, also die Geisteszugehörigkeiten und das sittliche Verhalten der Schüler beurteilt werden. Für das *Ingenium* genühten vier Stufen, die beiden anderen Bereiche sollten mit dem Fünfersystem bewertet werden.

| <i>Ingenium</i> | <i>Unterrichtsgegenstände</i> | <i>Mores</i> |
|-----------------|-------------------------------|--------------|
| Sehr fein | fein | fromm |
| Gut | fertig | fleissig |
| Ziemlich | ziemlich | still |
| Schlecht | etwas/wenig | unfleissig |
| | schlecht | ungehorsam |

¹ Vgl. zur historischen Noten- und Zeugnisforschung: Prinz von Hohenlohe/Liedtke (1991).

² Der heutige Schema der Schulnoten geht wesentlich auf den von Andres Reyher (1601-1673) verfassten Gothaer *Schulmethodus* zurück. Reyher war von 1639 an Rektor des Gymnasiums in Gotha. Er studierte in Leipzig, machte dort seinen Magister und lehrte an der Philosophischen Fakultät, bevor er eine Karriere als Gymnasialrektor machte. Der Herzog von Gotha, Ernst der Fromme (1601-1675), berief Reyher von Lüneburg nach Gotha, um die Reform der Volksschule voran zu bringen. „Schulmethodus“ wurde der „special- und sonderbahre Bericht“ zur neuen Volksschulordnung genannt, der 1642 veröffentlicht wurde. Diese Ordnung führte in Gotha die Schulpflicht vom fünften bis zum zwölften Lebensjahr ein, regelte die Rangfolge der Elementarfächer sowie die Einteilung der Schüler nach Klassen, schrieb die Lehrbücher vor, verfügte über Regeln der Unterrichtsmethode und der Schuldisziplin, sah Realien vor und fasste schliesslich das Schema der Notengebung. In diesem Sinne handelte es sich um die erste moderne Standardisierung des Volksschulunterrichts im deutschen Sprachraum.

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.

Zeugnisse und Noten, dazu Prüfungen und alternative Leistungsdiagnosen, sind Forschungsthemen, die sich im deutschen Sprachraum deutlich auf die zwanziger und sechziger Jahre des 20. Jahrhunderts konzentrieren. Besonders zwischen 1950 und 1970 lässt sich ein starkes Forschungsaufkommen nachweisen, das danach, im Unterschied zu den angelsächsischen Ländern, abnimmt. Die Forschung bezieht sich im Wesentlichen auf die Lehrkräfte und die Art und Weise, wie sie Noten geben. Erst neuerdings wird Interaktion zwischen Lehrern und Schülern untersucht. Die Abhängigkeit der Schule von öffentlichen Erwartungen im Blick auf Noten ist ebenso wenig untersucht worden wie die Ökonomie der Notengebung oder der Verwendungswert von Zeugnissen.

Der Nestor dieser Forschung, Karl-Heinz Ingenkamp aus Konstanz, fasste die vorliegenden Studien zusammen und beschrieb die Resultate im Blick auf vier Dimensionen:

- subjektive Fehlerquellen
- das klasseninterne Bezugssystem
- die Qualitätsunterschiede in den einzelnen Schulfächern
- den prognostischen Wert.

Die bis 1976 vorliegenden Arbeiten verweisen, so Ingenkamp, auf eine sehr weitgehende *Subjektivität* in der Zensurengebung. Zensuren und Noten seien persönlich gefärbte Urteile von Lehrkräften, nicht objektive Bewertungen, die auf einen gemeinsamen Massstab verweisen würden. Das gelte selbst für den Mathematikunterricht. Eine deutsche Studie zeigte, dass einige Mathematiklehrkräfte von den Fehlern, andere von den richtigen Lösungen ausgingen, um zur Zensur zu gelangen. Die Ansichten über die Bewertung von Faktoren wie Sauberkeit, Schrift, formal-mechanisches Rechnen im Vergleich zum Finden des Lösungsansatzes oder zum mathematischen Denken variierten erheblich (Haecker 1971).

Für die Beurteilung sei auch die „implizite Persönlichkeitstheorie“ entscheidend (Hofer 1975), mit der die Lehrkräfte ihre Verhaltens- und Charaktererwartungen an die Schülerinnen und Schüler festlegen. Diese Erwartungen werden nicht explizit formuliert und sind gleichwohl wirksam. Weitere Fehlerquellen im Urteil sind Stereotypen in der Geschlechtswahrnehmung³ oder auch einfach individuelle Vorlieben, wie schon vor Jahrzehnten in der Bewertung von Aufsätzen gezeigt wurde (Bobertag 1933, Ulshöfer 1949). Verschiedene Lehrkräfte beurteilen den gleichen Aufsatz oft verschieden, weil sie je auf andere Dinge achten.

Die drei anderen Dimensionen der Fragwürdigkeit beziehen sich auf Bedingungen und Folgen der Subjektivität. 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 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, zum Teil erheblich und manchmal auch ärgerlich. Ziffernnoten sind nie mit einem Hinweis auf den Kontext ihrer Entstehung versehen, also eine „5 ½ in Französisch“ erscheint *nicht* mit

³ Mädchen erhalten bessere Noten, als ihren in Tests erfassten Leistungen entspricht (Ingenkamp 1976, S. 71f.).

dem Signum „Kantonsschule Olten Maturitätsjahrgang 2007“, zusätzlich Angaben über Qualitätsmassstäbe, Lehrmittel und Leistungsniveaus der Schüler im Vergleich. Es sind einfach Ziffern, die einem Fach zugeordnet werden.

In verschiedenen Fächern wird mit unterschiedlicher Strenge zensiert, die Strenge nimmt mit der Bedeutung des Faches in der Studententafel zu, „stark selektive Hauptfächer“ haben die strengste Notengebung (de Groot 1971). Gleichzeitig zeigen Untersuchungen der Häufigkeitsverteilung, dass Zensuren „keine Intervallskala mit gleichen Notenabständen“ darstellen (Ingenkamp 1976, S. 232). Die Benotungsstrenge nimmt mit der Klassenstufe zu, bleibt in der Verteilung der Fächer aber gleich. Die Hauptfächer des Gymnasiums steigern die Anforderungen, die Zensurenverteilung in den Nebenfächern wird eher noch milder (Hopp/Lienert 1976). In vielen Fällen, nicht zuletzt in bestimmten Universitätsfächern, werden nur die beiden obersten Ziffern der Skala wirklich genutzt.

Der prognostische Wert dieser Praxis sei gering, schrieb Ingenkamp 1976. Zwar korrelieren innerhalb eines Schultyps die Zensuren in aufeinander folgenden Klassenstufen relativ hoch, aber das sei für die Prognose von Erfolg oder Misserfolg in nachfolgenden Schulen oder Stufen kaum von Bedeutung. Die Qualität ist zu unterschiedlich: Primarschulzeugnisse gewichten Elementarisierungseffekte wie die Rechtschreibleistungen oder auch Faktoren wie Fleiss und Wohlverhalten erheblich stärker als intellektuelle Fähigkeiten. Die Massstäbe von Aufnahmeprüfungen variieren von Jahr zu Jahr und von Ort zu Ort wiederum erheblich. Zudem hängt der Schulerfolg von der Stärke des Jahrgangs und den Klassenfrequenzen ab (Tent 1969; Sauer/Gamsjäger 1996). Der stärkste Faktor für Erfolg oder Misserfolg aber ist die Schichtzugehörigkeit, das soziale Milieu, aus dem die Schüler stammen (Ingenkamp 1976, S. 274ff; für den Kanton Zürich auch Moser/Rhyn 2000).

Der Topos der „Fragwürdigkeit“ der Zensurengebung zieht sich seitdem durch die deutschsprachige Forschungsliteratur. Die vier Dimensionen von Ingenkamp werden in einem Forschungsbericht von 1999 nochmals herausgestellt. Demnach ist die Benotung in verschiedenen Fächern unterschiedlich streng (Ziegenspeck 1999, S. 137ff.), wobei nicht nur der Rang des Faches eine Rolle spielt. In der Tendenz wird umso strenger zensiert,

- je mehr die Leistungen in schriftlichen Arbeiten überprüft werden,
- je mehr die Leistungen quantifizierbar sind
- und/oder je stärker die verbalen Anforderungen hervortreten.⁴

Mädchen erhalten dabei bessere Zeugnisnoten als Jungen, sie repetieren erheblich weniger und haben inzwischen den grösseren Schulerfolg (Richter 1996). „Erfolg“ wird in den diesbezüglichen Studien wie eine objektive Grösse betrachtet und mit Noten in Verbindung gebracht, zugleich wird der Topos der Fragwürdigkeit der Noten weiter verwendet. Der prognostische Wert von Zensuren oder Zeugnissen wird weiterhin als gering angesehen (Ziegenspeck 1999, S. 156ff.) und immer noch greift das „klasseninterne Bezugssystem“, also die Relativierung der Notenaussagen im Blick auf Ort und Kontext, in dem sie entstehen. Noten sollen ortsübergreifend vergleichbare Leistungen beschreiben, aber was sie erfassen, ist lediglich die Verteilung in einer bestimmten Leistungsgruppe.

Grossen Raum nimmt die Beschreibung der „subjektiven Fehlerquellen“ ein. Als Effekte werden unter anderem genannt:

⁴ Weiss (1962); Ingenkamp (1976).

- *Halo-Effekt*: Ein globaler Allgemeineindruck bestimmt die Wahrnehmung einzelner Merkmale.
- *Beharrlichkeitstendenz*: Lehrkräfte rücken von einem bereits gefällten Urteil bei späteren Beurteilungen nicht ab.
- *Reihungseffekt*: Unter dem Eindruck, „es können doch nicht alle gleich schlecht sein“ werden bessere Noten gegeben.
- *Kontrasteffekt*: Nach einer Serie von sehr guten Leistungen wird eine mittelmässige Leistung tendenziell als schlecht bewertet.
- *Beurteilungstendenzen*⁵: Milde oder Strenge, „zentrale Tendenz“ (Vermeidung von Extremwerten) und „motivierende“ versus „selektive“ Notengebung
- *Wissen-um-die-Folgen-Fehler*: Mildere Beurteilung bei absehbar negativen Folgen für die Schüler, nicht umgekehrt.

Damit verbunden sind die seit Beginn des 20. Jahrhunderts immer wieder beschriebenen Beurteilungsunterschiede bei der Zensurierung zwischen verschiedenen Lehrkräften oder zwischen derselben Lehrkraft zu verschiedenen Zeitpunkten (Grey 1913; Ziegenspeck 1999, S. 188ff.). Die Streuung der Aufsatznoten ist ein bekanntes Phänomen, weniger bekannt sind Befunde, die darauf hindeuten, dass die gleichen Beurteiler die gleiche Arbeit zu verschiedenen Zeitpunkten unterschiedlich bewerten.

Zudem gibt es eine „Verlaufskurve der Bewertung“ (Weidig 1961), die sich bei langwierigen Korrekturarbeiten einstellt und im gleichen Fach individuell sehr verschieden ist. Die erste Arbeit einer Korrekturserie wird tendenziell anders bewertet als die letzte, wobei der Effekt mit der Länge der Gesamtkorrektur zunimmt. Schliesslich hat die Reihenfolge in der Qualitätswahrnehmung einen Einfluss: „Kontrasteffekt“ meint in diesem Zusammenhang, dass dieselbe Arbeit, von der Lehrkraft nach Durchsicht vieler *guter* Proben korrigiert, ganz anders ausfällt als dies nach Durchsicht vieler *schlechter* Proben geschehen würde.

Implizite Persönlichkeits- und Charakterurteile spielen eine nicht zu unterschätzende Rolle, ebenso die Einschätzung der allgemeinen Qualität und Leistungsbereitschaft. Lehrerinnen und Lehrer beziehen sich in ihrem Unterricht mehr auf die guten Schüler, sie lassen gute wie schlechte Schüler direkt oder häufiger indirekt wissen, was sie von ihrer Leistungsfähigkeit und oft damit verknüpft von ihrer Person halten, und das Selbstbild der Schüler passt sich tendenziell dem Bild an, das die Lehrer von ihnen haben und mitteilen.⁶ Die Leistungshierarchie in einer Klasse steht relativ früh fest und ist nur schwer revidierbar. Bereits bei Beginn der Schulzeit bestehen grosse Unterschiede im Lernstand, die eher zunehmen, als dass sie kleiner werden. Es gibt so immer eine Rangordnung aufgrund von Leistungsverhalten und Bewertung.

Hinzu kommen die Attribuierungen aufgrund von Erfolg oder Misserfolg. Beide können dem eigenen Vermögen oder dem eigenen Unvermögen zugeschrieben werden, wobei das schulische Selbstbild offenbar in beiden Hinsichten verstärkend wirken kann. Es scheint dann ein „circulus vitiosus“ aufzutreten; wer als Schüler ständig keinen oder nur mässigen Erfolg hat, traut sich künftig auch keinen Erfolg zu. Es gibt auch Effekte, die zeigen, wie abhängig Lehrkräfte von der sozialen Situation und dem Entgegenkommen der Klasse sind, die sie unterrichten. Lehrkräfte, denen die Schüler *positive* Erwartungen entgegenbringen, stufen sich selbst im Blick auf ihre Tüchtigkeit höher ein als Lehrkräfte, bei denen die

⁵ Nach Ulbricht (1993).

⁶ „Pygmalion-Effekt“ im Sinne von Rosenthal/Jacobson (1971).

Erwartungen der Schüler *negativ* sind. Wenn ein Schüler oder eine Schülerin konstant Leistungen zeigt, dann ist das auch eine Reaktion auf das Verhältnis zum Lehrer.

Über Lehrerkommentare zu Schülerleistungen ist Folgendes bekannt (KRAMPEN 1987):

1. *Sozial orientierte* Lehrerkommentare wirken bei leistungsschwächeren Schülern deutlich negativ, bei leistungsstärkeren neutral oder leicht positiv.
2. An einem *sachlichen Standard* orientierte Lehrerkommentare wirken in der Tendenz bei allen Schülern positiv, ohne dass eine bestimmte Leistungsgruppe deutlich von ihnen profitiert.
3. *Individuell orientierte* Lehrerkommentare wirken ebenfalls bei allen Schülern tendenziell positiv, am meisten profitieren davon die leistungsschwächeren.

Generell scheint die feste allgemeine Einschätzung eines Schülers oder einer Schülerin die Notengebung zu beeinflussen. Wer gute Leistungen zeigt, erhält und *hat* so auch Charaktervorteile. Die Noten einzelner Schüler in bestimmten Fächern sind über das Schuljahr verteilt erstaunlich konstant.⁷ Wenn die generelle Einschätzung einmal gefasst ist, scheint sie kaum korrigierbar zu sein. Tendenzuell wird dem guten Schüler eine schlechte Leistung nicht nachteilig, dem schlechten Schüler eine gute Leistung nicht positiv vermerkt. Offenbar korreliert auch Sympathie auf Seiten des Lehrers mit besseren und Antipathie mit schlechteren Noten (Ziegenspeck 1999, S. 206).

Angesichts dieser Resultate könnte man schliessen, es sei *unmöglich*, zu einer „gerechten Zensierung“ der Leistungen von Schülern zu gelangen, und dass dies der Lehrerschaft auch bewusst sei (Döring 1925, S. 177).⁸ Das schrieb der deutsche Jugendpsychologe Otto Döring im Jahre 1925. Aber das wirft dann die Frage auf, warum immer noch weitgehend Ziffernnoten die Praxis der Leistungsbeurteilung bestimmt. 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. Noten sind einfache Instrumente, deren Aufwand begrenzt ist und die leicht kommuniziert werden können. Vorteile sind nie untersucht worden, einfach weil die Forschung lange davon ausging, die Notengebung könnte keine haben.

Inzwischen sind 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 oft nicht den tatsächlichen Schülerleistungen, wenn man diese unabhängig testet. Noten beschreiben den internen Rangunterschied. Aber auch wenn Lehrkräfte gut in der Lage sind, die Schüler ihrer eigenen Klasse gemäss ihren Leistungen zu rangieren, heisst das nicht, „dass gleichen Noten in unterschiedlichen Klassen auch vergleichbare Leistungen zu Grunde liegen“ (ebd.).

Daraus ergibt sich eine dreifache Forderung:

1. Die subjektive Beurteilung muss durch eine *objektive* ersetzt werden, die Standards oder einen übergreifenden „Bezugsmassstab“ voraussetzt.

⁷ Seit Rank (1962) und Ingenkamp (1963) ist das verschiedentlich nachgewiesen worden.

⁸ Zum Problem der Gerechtigkeit bereits Wolf (1962) und Weiss (1962).

2. Das objektive Kriterium sind die vom Lehrplan geforderten *Lernziele*, unabhängig von den in einer Klasse vorhandenen Leistungsunterschieden.
3. Dafür müssen am Lehrplan orientierte, *diagnostische Instrumente* entwickelt werden, die die Noten ersetzen.
(ebd.).

Die beiden ersten Forderungen sind inzwischen in der bildungspolitischen Diskussion akzeptiert und auf dem Weg, wie das HarmoS-Kondordat zeigt. Wie sie sich angesichts des damit verbundenen Aufwandes realisieren lassen, ist eine andere Frage. Die dritte Forderung bezieht sich auf ein neues Instrument, in der Beurteilungen nicht nur in der Dimension der Lernziele möglich werden. 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.

Instrumente, mit denen mehr beurteilt werden kann als nur die Leistungen, die im Blick auf 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. Diesen Wunsch teilen viele Forscher, er ist auch in der Lehrerbildung präsent, aber er ist nie Wirklichkeit geworden. *Beurteilen* ist nach wie vor vor allem *Benoten*, und zwar Benoten von individuellen Leistungen. Wie die Beurteilungskultur an Schweizer Gymnasien aussieht und wie dazu das Leistungsverhalten passt, wird mich in einem zweiten Schritt beschäftigen. Ich werde mich dabei auf Daten aus dem Kanton Zürich beziehen, aber auch auf Befunde einer landesweiten Untersuchung, die 2004 vorgelegt wurden. Beide Untersuchungen sind umfangreicher, als ich angesichts meines Zeitrahmens darstellen kann.

2. Leistungen und Beurteilungen im Schweizer Gymnasium

Im Kanton Zürich werden die Absolventen der Gymnasien regelmässig zwei Jahre nach Abschluss der Maturitätsprüfung befragt, wie sie die Qualität ihrer Ausbildung im Blick auf fachliche und überfachliche Kompetenzen einschätzen. In der dritten Ehemaligenbefragung (2000) des Kantons Zürich, die insgesamt eine hohe Zufriedenheit ergeben hat,⁹ wird die „zeitliche Belastung durch die Schule“ zentral gewichtet, die kaum Raum lasse für die Entwicklung persönlicher Interessen (Befragung 2001, S. 22). Die Mittelschulen werden vornehmlich als Institutionen für die Vermittlung *kognitiver* Fähigkeiten wahrgenommen, wobei klare Prioritäten die Wahrnehmung und die Praxis des Verhaltens der Schülerinnen und Schüler bestimmen. Sie berechnen ihre Chancen, teilen sich ihre Zeit und wägen Aufwand und Ertrag ab, also handeln rational.

Die „Erwartungen der Schülerschaft“ an ein Schulfach spielen eine erhebliche Rolle beim Zustandekommen der Leistung, wobei die Erwartungen sich wesentlich darauf beziehen, wann ein Fach ernst zu nehmen ist und wann nicht. Offenbar spielen neben dem Rang der Fächer im Curriculum der Gymnasien Tests und Noten eine zentrale Rolle, die von den Schülern im Blick auf Zeitaufwand und Leistungsverhalten kalkuliert werden. Die meisten haben die rote Linie des Notenkalküls vor Augen, die - wie knapp auch immer - erreicht

⁹ 87% der Ehemaligen äussern sich positiv zu ihrer Mittelschule, die kantonsrätliche Vorgabe liegt bei 85% (Befragung 2001, S. 10).

werden muss. Leistung ist für sie keineswegs die je optimale Anstrengung, sondern bezieht sich darauf, wie eine bestimmte Notenrechnung erreicht werden kann. Die Schülerinnen und Schüler setzen ihre Ressourcen gezielt ein und suchen Streueffekte zu vermeiden. Dabei spielen die Erwartungen an ein Fach und die damit verbundene Motivation eine nicht zu unterschätzende Rolle.

Trotz - oder wegen - dieses aus Sicht des gymnasialen Bildungsideals nicht gerade überzeugenden Befundes erhalten die Gymnasien in den vier bisher vorliegenden Ehemaligenbefragungen¹⁰ gute Noten im Blick auf die für das Studium bei Eingang notwendige Fachkompetenz und die darauf bezogene Studierfähigkeit. Die letzte Befragung aus dem Jahre 2003 bestätigt diesen Befund weitgehend. Die allgemeine Zufriedenheit mit den Schulen ist sogar noch leicht gestiegen, unverändert ist die gute bis sehr gute fachliche Vorbereitung auf das Studium und die Kompetenz, neues Wissen zu erwerben. Auch das Wohlbefinden und die Beurteilung durch die Lehrkräfte werden unverändert und im Schnitt positiv bis sehr positiv eingeschätzt.

Erheblich schwächer bewertet werden die Eigenständigkeit des Urteils und die Möglichkeiten, zielorientiert zu handeln. Nur 50% der Befragten stimmt der Aussage zu, im Gymnasium habe man gelernt, „selbstständig Probleme zu lösen“. Noch schwächer eingeschätzt wird der Beitrag des Gymnasiums zur Entwicklung der Sozialkompetenz und die Mitsprachemöglichkeiten der Schülerschaft (Befragung 2004, S.11ff.). Am schlechtesten beurteilt wird der Ausbildungsstand in Informatik. Noch immer, so der Befund 2003, wird das Informatik-Wissen „zum grössten Teil ausserhalb der Mittelschule erworben“ (ebd., S. 22). Interessant ist, dass Absenzen ein deutliches Konfliktthema waren und 23% der Befragten angaben, Differenzen mit Lehrkräften hätten sie belastet (ebd., S. 18). Fast ein Viertel fühlte sich von den Klassenlehrpersonen vernachlässigt und die Hälfte sah sich bei Problemen von der Schule „oft allein gelassen“ (ebd., S. 17).

Durchgehende Mängel werden aus der Sicht der Absolventen wesentlich im Bereich der *überfachlichen* Kompetenzen gesehen (ebd., S. 25). Im Blick auf diesen Befund sind zwei weitere Studien durchgeführt worden, die näher erfassen sollen, warum die Ausbildung von überfachlichen Kompetenzen in der Lernkultur des Gymnasiums offenkundig vernachlässigt oder zumindest nicht als Kernaufgabe angesehen wird (Maag Merki 2002, Maag-Merki/Leutwyler 2005). Neben positiven hat die erste Studie auch negative oder bedenkliche Resultate erbracht. Die problematischen Ergebnisse bezogen sich auf alle drei Dimensionen der Befragung, nämlich

- die Strategien des Lernens und Arbeitens,
- politische Kompetenzen
- und der Umgang mit Gesundheitsrisiken.

Die zweite Studie zeigt eine Zunahme der intrinsischen Motivation, ein positiveres Bild bei den politischen Kompetenzen, auch eine Verbesserung in der Selbstwirksamkeit sowie im Gesundheitsbereich, nicht jedoch bei den Lernstrategien (Maag Merki/Leutwyler 2004, S. 20). Lernstrategien sind keine didaktische Nebensächlichkei. Von der Beherrschung anspruchsvoller Lernstrategien wird die Leistung wesentlich mitbestimmt. Die Erwartungen an das Fach ist die eine Seite, die tatsächlich eingesetzten Strategien des Lernens die andere. Sie müssen im Unterricht verwendet werden, aber das setzt voraus, anspruchsvolle Strategien werden als eine Leistungsanforderung verstanden. Dafür müssen die Lehrkräfte von den

¹⁰ Die Befragungen werden durchgeführt im Auftrag des Kantonsrates.

Schülern auch andere als konventionelle Lernstrategien abverlangen. Das geht nur dann, wenn der tägliche Unterricht darauf eingestellt ist. Psychologische Kurse, die manche Gymnasien zu Beginn des Schuljahres durchführen, leisten dies nicht. Die Unterrichtskultur muss sich verändern, und die Lehrkräfte müssen ihr Repertoire darauf abstellen.

Einen Einblick in die Prüfungspraxis der Gymnasien hat die erste EVAMAR-Studie geliefert. Die Maturandinnen und Maturanden 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 oben zitierten Forschungsliteratur beschrieben worden sind. Selektive Hauptfächer wie Mathematik oder Sprachen vergeben Noten nach strengeren Massstäben als musische Fächer. Die Noten in Mathematik streuen „deutlich stärker“ als die in den anderen Fächern, die Mathematiknoten erhalten daher eine erhöhte Bedeutung bei der Promotion, 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 Schulfächer¹¹ 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 ebenfalls viel zitierten „Sprachbegabung“ getragen sein würde, lässt sich nicht feststellen. Wer etwa Englisch als Schwerpunktfach besucht, erzielt tiefere Noten in der Unterrichtssprache und nochmals tiefe in der zweiten Landessprache. Das gilt allerdings nicht für Latein: Wer dieses Schwerpunktfach besucht, erzielt höhere Noten in der Unterrichtssprache und fast die gleichen Noten in der zweiten Landessprache wie in Latein (ebd., S. 133).

Im Zusammenhang mit den Noten lässt sich auch ein Koppelungseffekt nachweisen:

- Wer das Schwerpunktfach Physik und Anwendungen der Mathematik gewählt hat, weist besonders hohe Noten in Mathematik und besonders

¹¹ Zu unterscheiden von längeren Sprachaufenthalten, wie sie verschiedene Gymnasien in Form der zweisprachigen Matur anbieten.

niedrige Noten in den sprachlichen Grundlagenfächern (Unterrichtssprache und zweite Landessprache) auf.

- Wer Philosophie/Pädagogik/Psychologie gewählt hat, erzielt hohe Noten in der Unterrichtssprache, erhebliche tiefere in der zweiten Landessprache und tiefe in Mathematik. Die Noten im Schwerpunktfach sind hier erheblich höher als im Fach Physik und Anwendungen der Mathematik.
- Wer von den Schülerinnen und Schülern das Schwerpunktfach Musik gewählt hat, erreicht sehr viel bessere Noten in Mathematik als diejenigen aus den Schwerpunktfächern Deutsch als Fremdsprache, Italienisch als Fremdsprache und Philosophie/Pädagogik/Psychologie.

Die Koppelungseffekte führen zu unterschiedlichen Notenprofilen und so Gewichtungen. Wer das Schwerpunktfach Deutsch als Fremdsprache besucht, erreicht mit 4.03 den tiefsten Notendurchschnitt in Mathematik überhaupt.¹² Der höchste Durchschnitt im Grundlagenfach Mathematik wird von den Absolventinnen und Absolventen des Schwerpunktfaches Physik und Anwendungen der Mathematik erreicht. Er liegt mit 4.85 sogar höher als im Schwerpunktfach, während die gleichen Schülerinnen und Schüler in der zweiten Landessprache Noten erzielen, die im Schnitt um mehr als eine halbe Note niedriger liegen als in Mathematik. Wer den Schwerpunkt Bildnerisches Gestalten besucht, erreicht in Mathematik und in der zweiten Landessprache schwache Noten, die sich aber nur um 0.16 Noten unterscheiden.

Die Bewertungen in den Grundlagenfächern sind *nicht* niedriger, wenn die Schülerinnen und Schüler eines der *neuen* Schwerpunktfächer gewählt haben. Sie können daher auch nicht als leistungsschwächer eingestuft werden, nur weil sie diese Fächer gewählt haben.

- Wer das Schwerpunktfach Philosophie/Pädagogik/Psychologie gewählt hat, erreicht am Ende in Mathematik ähnlich tiefe Noten wie die Absolventinnen und Absolventen der Schwerpunktfächer Spanisch oder Italienisch als Fremdsprache.
- Wer das Schwerpunktfach Musik besucht, erhält in der Unterrichtssprache im Durchschnitt ähnliche Noten wie diejenigen, die das Schwerpunktfach Englisch oder Russisch besuchen.
- Offenbar erleichtert die Wahl der neuen Schwerpunktfächer den Weg zur Maturität nicht, wie gelegentlich vermutet wurde (Evaluation der Maturitätsreform 2004, S. 133).

Die Quote der *nicht* bestandenen Maturitätsprüfungen hat sich nach Umsetzung des MAR mit 2.4% im Jahre 2003 fast verdoppelt, liegt aber immer noch sehr tief (ebd., S. 134). Die Erfolgsquote sank vor allem in der Romandie und betrug dort 94.6% verglichen mit 98.5% in der italienischen Schweiz (ebd.). Zwischen den einzelnen Schulen gibt es zum Teil erhebliche Unterschiede. Das kann mit der Umstellung des Systems zu tun haben und ist kein Indikator für eine generelle Leistungsverschärfung.

Die weitaus meisten Jugendlichen bestehen die Prüfung, wobei Misserfolge bei den Maturanden „auffällig häufiger“ als bei den Maturandinnen zu verzeichnen sind (ebd., S. 135). Vergleichszahlen zum alten System liegen allerdings nicht vor. Der Effekt muss also keine Folge des neuen Systems sein und kann auch als Fortsetzung des langfristigen Trends

¹² Noten unter vier stehen für ungenügende Leistungen (MAR Art. 16,1).

verstanden werden, der sich mit generellen Unterschieden im schulischen Verhalten beider Geschlechter erklären liesse. Maturandinnen sind in den Abschlussklassen mit 57%¹³ weit stärker vertreten als Maturanden mit 43%, und die Maturanden bestehen häufiger die Prüfung nicht. Unterschiede zeigen sich auch in Schwerpunktfächern, nicht jedoch in den Ergänzungsfächern (ebd.).¹⁴

- Wer die Maturitätsprüfung *nicht* besteht, hat ungenügende Noten vor allem in Mathematik und der zweiten Landessprache, gefolgt mit Abstand von der zweiten Fremdsprache und dem Schwerpunktfach.
- Das gilt ähnlich für die Schülerinnen und Schüler, die die Maturität knapp bestehen. Ihre beiden schlechtesten Fächer sind Mathematik und die zweite Landessprache.
- Zwischen „nicht bestanden“ und „knapp bestanden“ ist die Differenz im Schwerpunktfach, in der zweiten Landessprache und in der Erstsprache besonders gross.
- Eine ungenügende Note im Schwerpunktfach haben 18.5% derjenigen Schülerinnen und Schüler, die die Maturitätsprüfung *knapp* bestanden, aber 42.5%, die sie *nicht* bestanden.
- Ähnlich hoch liegt die Differenz in den anderen beiden Fächern, während sie in den Naturwissenschaften vergleichsweise niedrig ist.

Das verdeutlicht die „kritische Rolle der Sprachfächer“ (ebd., S. 136). Diese These ist mit einer Modellrechnung veranschaulicht worden, die die Ergebnisse unter Veränderung der MAR-Rahmenbedingungen durchgespielt hat. Dabei ergibt sich folgendes Bild:

„Bestimmt man den Entscheid über das Bestehen der Maturität unter Ausschluss des Prinzips der doppelten Kompensation ungenügender Noten, so würde die Misserfolgsquote um fast die Hälfte zurückgehen. Zählt man die Naturwissenschaften wie drei Fächer, so würde sich die Quote um etwa einen Drittel reduzieren... Die Reduktion der Misserfolgsquote ist so eine Folge der im Vergleich zu den selektivsten Fächern insgesamt höheren Bewertung der drei Naturwissenschaften“ (ebd., S. 137).¹⁵

Die Maturaarbeit hat gemäss MAR keinen direkten Einfluss auf das Bestehen der Maturität. Gleichwohl hat die Arbeit innerhalb kurzer Zeit in der Erfahrung von Maturitätslehrgängen einen herausgehobenen Rang erhalten. Landesweit schätzen die Lehrpersonen die Maturaarbeit als ein „nützliches und interessantes“ Element der MAR-Reform, Schulleiter bezeichnen sie gar als deren „Glanzstück“, problematisch ist wenn, dann die Umsetzung. Weniger als 70% der Lehrkräfte (weniger als 60% in der Romandie) schätzen die Kriterien der Bewertung als klar ein, wobei die Einschätzung umso kritischer ausfällt, je jünger die Lehrkräfte sind. 32% der Lehrkräfte geben an, dass die Schülerinnen und Schüler nicht über die Fähigkeiten autonomen Lernens verfügen, die bei der Maturaarbeit vorausgesetzt werden. Und 52% schätzen, dass die methodischen Voraussetzungen fehlen, eine Maturaarbeit erfolgreich abfassen zu können. Vor allem aber wird beklagt, dass für die

¹³ In der Westschweiz liegt der Prozentsatz bei 60% (Evaluation der Maturitätsreform 2004, S. 146).

¹⁴ Jugendliche, die Latein oder Englisch gewählt haben, bestehen die Maturität etwas häufiger knapp, Jugendliche mit dem Schwerpunktfach Biologie und Chemie fallen häufiger durch (Evaluation der Maturitätsreform 2004, S. 135).

¹⁵ Die Schulen haben zu wenig Angaben über die Teilnoten in Chemie, Biologie und Physik gemacht. Bei der Schätzung konnte also das Prinzip der doppelten Kompensation zwischen diesen drei Fächern nicht berücksichtigt werden (Evaluation der Maturitätsreform 2004, S. 137).

Betreuung und Bewertung der Maturaarbeit die notwendige Zeit fehle (Evaluation der Maturitätsreform 2004, S. 183).

Die Mehrheit der Schülerinnen und Schüler schätzt die Bedingungen für die Abfassung der Maturaarbeit als zufrieden stellend ein, wobei vor allem die Autonomie des Arbeitens und das mit dem Thema verbundene persönliche Engagement als wertvoll bezeichnet werden (ebd., S. 180f.). Die methodische Seite der Arbeit wird dagegen wesentlich schwächer eingeschätzt. Die Betreuung durch die Lehrkräfte wird von einer grossen Mehrheit positiv eingeschätzt (ebd., S. 179). Allerdings sind 44% der Schülerinnen und Schüler der Auffassung, dass die Anweisungen der Lehrkräfte wenig klar waren. 32% waren der Meinung, die Rahmenvorgaben der Lehrkräfte seien ungenügend gewesen, wobei diese beiden Einschätzungen mit grossen regionalen Unterschieden verbunden waren (ebd., S. 183). Die Wahl der Betreuung ist unterschiedlich, bestimmte Lehrpersonen werden häufiger in Anspruch genommen als andere. Auch die Verteilung der Themen nach Fächern oder Lernbereichen ist unterschiedlich. Die meisten Maturitätsarbeiten werden landesweit in den Sprachfächern geschrieben, gefolgt von den mathematisch-naturwissenschaftlichen Fächern einschliesslich Informatik, sowie den Geistes- und Sozialwissenschaften einschliesslich Wirtschaft und Recht (ebd., S. 166).

- 40% aller 3307 in EVAMAR untersuchten Maturaarbeiten waren Theoriearbeiten,
- 33.1% wandten empirische Forschungsmethoden an,
- 12.3% stellten künstlerische Arbeiten dar
- und 14.0% waren auf soziokulturelle oder ausserschulische Themen und Aktivitäten gerichtet.

Nur im letzten Typus gibt es eine klare Geschlechtsdifferenz, 16.0% der Schülerinnen gegenüber 11.5% der Schüler wählten diesen Typus. Ansonsten ist diese Differenz minimal (ebd., S. 135).

- Gesamthaft konnten 73% der Schülerinnen und Schüler das Thema der Maturaarbeit frei wählen.
- In der deutschsprachigen Schweiz traf das sogar auf 97% zu,¹⁶ die Themenerzeugung lag also faktisch komplett in der Hand derer, die die Maturaarbeit erstellen müssen.
- Die weitaus meisten Schülerinnen und Schüler, nämlich 90%, wählten das Thema nach eigenem Interesse.
- Ein Drittel bezieht sich bei der Wahl auf individuelle Vorkenntnisse, und weniger als 10% gibt Gründe an wie die Leichtigkeit der Realisierung des Themas, eine zufällige Themenwahl oder eine Beratung durch andere Personen.
- Damit steht das Interesse im Mittelpunkt, was auch den zum Teil hohen Aufwand erklärt, mit dem diese Arbeiten durchgeführt werden.

Bewerten ist natürlich mehr als nur Benoten. Aber die Zahlen zeigen, welche herausragende Bedeutung Noten für die Praxis der Gymnasien haben und auch weiterhin haben werden. Wenn von Fördern gesprochen wird, dann bezieht sich das zumeist auf die Verbesserung der Leistung im Blick auf Noten. Die Praxis des Förderns ist individuell,

¹⁶ Im Tessin waren nur 8% der Schülerinnen und Schüler frei in der Wahl des Themas ihrer Maturaarbeit, was mit der andersartigen Organisation zusammenhängt. Die Maturaarbeit ist Teil von Seminarveranstaltungen, in deren Rahmen die Themen vergeben werden (Evaluation der Maturitätsreform 2004, S. 176).

einzelne Schülerinnen und Schüler werden mehr oder weniger gezielt angesprochen und auf Verhaltensalternativen hingewiesen. Damit können neue Möglichkeiten des Lernens gemeint sein, auch neue Ressourcen oder andere Formen der Mitarbeit.

Die Maturaarbeit stellt eine neue Möglichkeit des Lernens im Gymnasium dar. Bei der weiteren Entwicklung der Unterrichtskultur lautet das grosse Stichwort „selbständiges“ oder „eigenverantwortliches“ Lernen. Ich gehe abschliessend darauf ein, welche Erfahrungen hier vorliegen und wie sich der Anspruch der Selbsttätigkeit mit der Qualitätssicherung im Gymnasium vereinbaren lässt. Autonome Lernformen sind natürlich auch ausserhalb des Gymnasiums entwickelt worden. Ich gehe darauf ein, sage aber zunächst etwas zur Qualitätssicherung.

3. Autonomes Lernen und Qualitätssicherung im Gymnasium

Alle Schweizer Gymnasien verwenden heute in der einen oder anderen Form Strategien der Qualitätssicherung. Die Strategien reichen von der Entwicklung und Kommunikation eigener Leitbilder (Mandel 2006) bis hin zur Beauftragung externer Evaluationen. Auch interne Evaluationen sind zu einem bestimmten Standard geworden ebenso die Kommunikation transparenter Leistungsanforderungen, etwa im Blick auf die Maturaarbeit (Oelkers 2007). Zu einem Standard geworden ist auch die Homepage jeder einzelnen Schule, die sozusagen den Markenkern darstellt und vielfach genutzt wird. Auffällig ist, dass „Bildungsstandards“ die Gymnasien bislang kaum erreicht haben.

Generell hat Qualitätssicherung im Gymnasium zu tun mit fünf einzelnen Handlungsbereichen:

- Evaluation der Schulqualität
- Transparenz der Leistungsanforderungen
- Entwicklung der Lerneinstellungen
- Analyse und ggf. Verbesserung der Prüfungs- und Notenpraxis
- Standards und Tests.

Wenn Qualitätssicherung mehr sein soll als ein unverbindliches Schlagwort und auch mehr als ein Webauftritt, dann müssen diese Bereiche bearbeitet werden. Dabei spielen neue Verfahren und Instrumente eine Rolle. Es ist sicher kein Zufall, dass in der Entwicklung zunehmend elektronische Medien verwendet werden. Auch Leistungstests gewinnen an Bedeutung. Tests sind nicht alles, sie haben auch deutliche Grenzen, aber sie werden für die Beurteilung der Leistungen immer wichtiger und sie werden ihren Platz nicht zuletzt in der Kommunikation mit Schülern und Eltern finden.

Das im Volksschulbereich am häufigsten benutzte Verfahren, das schon fast sprichwörtlich geworden ist, heisst *Klassencockpit*.¹⁷ Es handelt sich um ein Internet-Angebot für die vergleichende Bestimmung des Leistungsstandes einer Klasse. Der Test gibt Aufgaben vor, die von der Klasse bearbeitet und deren Lösungen mit einem kantonalen Schnitt verglichen werden. Die Eingabe ist anonym. Das System verlangt minimale Ausbildung, ist sofort nutzbar und erlaubt intelligente Anwendung. Klassencockpit wird inzwischen in vielen

¹⁷ <http://www.klassencockpit.ch>

Schulen erfolgreich genutzt (Moser 2003). Das zeigt sich daran, dass die Schüler zwischen den normalen Tests und Klassencockpit unterscheiden. Wenn man sie fragt, wie der Tag war, sagen sie bei Gelegenheit, heute war „Klassencockpit.“

Damit meinen sie, dass sie Testaufgaben bearbeitet haben, die *nicht* die Lehrkraft gestellt hat. Sie waren in diesem Sinne „autonom.“ Es gibt dann für ihre Leistung ein Vergleichsdatum, das früher fehlte. Viele Lehrkräfte nutzen inzwischen diese Möglichkeit, ohne einen Gedanken daran zu verschwenden, dass sie sich früher gegen jeden Vergleich gewehrt haben. Heute stellen sie Noten der einzelnen Schüler oft so dar, dass ihre Bewertung zusammen mit dem Ergebnis von Klassencockpit erscheint, um mit einem Vergleichsdatum die Transparenz zu erhöhen, nicht zuletzt im Blick auf die Elterngespräche. Das System ist nicht perfekt, stellt aber einen Schritt dar, wie Beurteilungen vom klassischen Verdacht entlastet werden können, sie seien „rein subjektiv.“ Ich bin gespannt, was die Zeugnisforschung dazu sagen wird.

Schulentwicklung ist auch unterhalb der Schwelle von Bildungsstandards, Leistungstests und einer komplexen Organisation der Qualitätssicherung möglich. Ich beschreibe im Folgenden Massnahmen, die in Schweizer Schulen entweder bereits realisiert sind oder entwickelt werden. Es sind Massnahmen zur Verbesserung vor allem der Einstellungen zum Lernen und zur Erreichung von mehr Selbständigkeit. Gemeint ist damit nicht die geheimnisvolle Mutation der Lehrkraft zum „Lerncoach,“ sondern praktikable Modelle, die für Schule und Unterricht Gewinn bringen können. Die weiterführenden Wege der Schulentwicklung werden nicht am grünen Tisch erzeugt. Wer Know How sucht, kann dies nicht aus Theorien ableiten, sondern muss sich auf die Suche nach guten Lösungen in der Praxis machen.

Ein immer wieder vorgebrachter und wie ich finde berechtigter Vorwurf, den viele Eltern äussern, ist der mangelnder Transparenz der Leistungsentwicklung der Schüler, also in einem längeren Prozess, der oft aus Sicht der Eltern verdeckt abläuft oder gut getarnt ist. Wenn Zeugnisse verteilt werden, ist es zu spät, in diese Entwicklung einzugreifen, was viele Eltern gerne täten, weil sie mit den Rückmeldungen der Kinder oft nicht zufrieden sind. Das hängt mit dem Phänomen zusammen, dass viele Schüler lieber auf schlechte Zeugnisse warten als ihr Leistungsverhalten zu verändern. Intransparenz bringt für sie kurzfristige Vorteile, wobei die Eltern oft ahnen, was auf sie zukommt.

Es gibt Schweizer Schulen, die den Eltern alle vier Wochen den Leistungsstand ihrer Kinder mitteilen. Die Schulen legen Datenbanken an, in denen alle Lehrkräfte die Noten der schriftlichen Leistungen eintragen. Die Eltern erhalten dann regelmässig einen Auszug, der sie über den Stand informiert und den sie unterschreiben müssen. Sie können dann überlegen, welche Strategien sie ergreifen, wenn ein Leistungsniveau erreicht ist, der weder sie noch ihre Kinder zufrieden stellt. Schulen, die komplett auf elektronische Formate umgestellt sind, wie die *Schule für individuelles Lernen* (SiL) in Birmensdorf im Kanton Zürich,¹⁸ kommunizieren auf diese Weise auch mit den Schülern, Sie haben jederzeit Zugriff auf den Stand ihrer Leistungsentwicklung. Aufgaben, Leistungen und Bewertungen sind gespeichert, der Lernstand ist also hoch transparent.

Es gibt andere Schulen, die den Schülerinnen und Schülern am Freitag nach der letzten Stunde schriftlich mitteilen,

¹⁸ <http://www.sil.tagesschule.ch/>

- was in der nächsten Woche an Zielen erreicht werden soll,
- welche inhaltlichen Anforderungen auf sie zukommen
- und welche zeitlichen Ressourcen sie benötigen, um gute Leistungen zu erzielen.

Die Schüler führen Lerntagebücher und können so ihren Lernfortschritt selbst kontrollieren. Die transparenten Ziele erlauben es, die Zeit einzuteilen und die notwendigen Leistungen vor sich zu sehen, ohne vom Fortgang des Unterrichts überrascht zu sein. Die Schulen verlangen so aktives Lernverhalten und Rechenschaft über den Fortgang der Entwicklung hin zum Ziel.

Eine Möglichkeit dafür sind direkte Rückmeldesysteme, die manche Schulen „Lernpass“ nennen und die mein drittes Beispiel darstellen. Einfache Rückmeldesysteme dienen der wechselseitigen Abstimmung und Kontrolle des Verhaltens.

- Die Schüler tragen jeden Tag oder jede Woche ein, welche Ziele in welchen Fächern sie an dem bestimmten Tag erreichen möchten.
- Sie protokollieren, was sie erreicht haben.
- Sie schätzen nach vorgegebenen Kategorien ihr Arbeits- und Lernverhalten ein.
- Die Lehrkraft erhält eine Rückmeldung, ob die Ressourcen und Lernhilfen ausreichend waren.
- Die Schüler erhalten abschliessend von den Lehrkräften eine Rückmeldung, wie der Verlauf des Lernprozesses und die Leistungen eingeschätzt wurden.

Solche Verfahren sind elektronisch relativ leicht umzusetzen, sie sind bislang allerdings nicht unter gymnasialen Voraussetzungen erprobt worden. Rückmeldesysteme, die die Schüler benutzen können und die ihr Lernen steuern, sind sehr effizient, verlangen aber eine hohe Autonomie und sind nur erfolgreich, wenn sie als Lernstrategie akzeptiert sind und ständig genutzt werden. Ob das mit grossen Klassen und unter der Voraussetzung von differenziertem Fachunterricht möglich ist, müsste erprobt werden. Die Schule in Birmensdorf bietet Fachunterricht auf Sekundarstufe I, aber sie ist eine kleine Privatschule.

Mein viertes Beispiel betrifft Form und Inhalt der Zeugnisse, die oft als das genaue Gegenteil von „mehr Selbständigkeit“ hingestellt werden. Zeugnisse sind oft Ärgernisse, nicht wegen der Noten, sondern weil diese zu wenig aussagen. Als Reaktion darauf ist im Kanton Zürich das Volksschulzeugnis gründlich verändert worden, um realistischere Aussagen zu erreichen. Die wesentliche Änderung geht dahin, in bestimmten Fächern Kompetenzniveaus zu unterscheiden, also etwa in Deutsch nicht mehr pauschal Noten zu geben, sondern mit vier Notenniveaus erreichte oder nicht erreichte Kompetenzen zu bewerten. Noten werden dann vergeben in Hörverstehen, Leseverstehen, Sprechen und Schreiben, und nicht mehr „in Deutsch.“

In Zukunft werden auch überfachliche Kompetenzen bewertet. Neu wird mit einem Viererschema von „trifft zu“ bis „trifft nicht zu“ das Arbeits- und Lernverhalten sowie das Sozialverhalten der Schülerinnen und Schüler bewertet. Im Bereich „Arbeits- und Lernverhalten“ gibt es drei Kategorien, die in sich differenziert bewertet werden, nämlich Lernmotivation und Einsatz, Aufgabenbearbeitung und Selbsteinschätzung. Wenn mehr autonomes Lernen angestrebt wird, dann muss der Effekt in irgendeiner Form dokumentiert werden, am besten im Zeugnis. Lehrerverbände und auch Elternorganisationen haben darauf

positiv reagiert, vor allem weil mehr Transparenz erreicht wird und aussagekräftiger kommuniziert werden kann.

Mein letztes Beispiel ist das inzwischen viel diskutierte und von den Medien positiv herausgestellte „Selbstlernsemester“ der Kantonsschule Zürcher Oberland in Wetzikon. Es handelt sich um einen Schulversuch, der im Herbstsemester 2004/2005 erstmalig realisiert und extern evaluiert wurde (Binder/Feller-Länzlinger 2005). Während eines gesamten Halbjahres erhielten die Schülerinnen und Schüler in den Grundlagenfächern Deutsch, Mathematik, Französisch und Englisch, im Schwerpunktfach (Griechisch, Latein oder Physik) sowie im Fach Sport keinen auf Lektionen bezogenen Unterricht. Sie mussten ohne tägliche und direkte Anleitung durch die Lehrkräfte lernen.

„Anstelle der wöchentlichen Anzahl von Lektionen für jedes der Fächer“ erhielten die Schülerinnen und Schüler „einen Semesterauftrag mit Aufgaben und definierten Lernzielen, die selbständig oder in Gruppen erarbeitet werden mussten. Die Begleitung und Betreuung durch die Lehrpersonen erfolgte im Rahmen von wöchentlich stattfindenden Sprechstunden oder im direkten persönlichen Kontakt“, zum Teil auch per E-Mail, zwischen Lehrpersonen und Schülern. „Die Überprüfung und Beurteilung der Lernzielerreichung erfolgte mittels verschiedener Prüfungsformen“ (ebd., S. 7).

Das erste Semester ist evaluiert worden, weitere Erhebungen liegen noch nicht vor. Die Ergebnisse des ersten Semester lassen sich im Vergleich mit einer Kontrollgruppe so darstellen.

1. Die vorgegebenen Lernziele wurden in allen acht Fächern im gleichen Ausmass wie im Normalunterricht erreicht. Einzig in zwei Klassen mit Sprachprofil hatten die Schüler mehr Schwierigkeiten mit Mathematik.
2. Die Semesteraufträge waren in allen Klassen als Lernvorgaben geeignet.
3. Die Lernzielüberprüfung war für die Lehrkräfte in der zweiten Semesterhälfte eine starke Belastung.
4. Die Formen der Lernbegleitung mussten von den Lehrkräften aufwändig entwickelt werden.
5. In der Einschätzung aller Beteiligten haben die Schüler nicht nur fachliche, sondern vor allem auch überfachliche Kompetenzen wie anspruchsvollere Lernstrategien ausbilden können.
6. Alle Beteiligten beurteilen das Selbstlernsemester positiv. Knapp 70 Prozent der Schüler geben an, sie hätten mit dieser Form besser gelernt als im gewohnten Unterricht.
(ebd., S. 4/5).

Auf dieser Linie sagten Lehrkräfte in den Interviews:

- „Es war ernüchternd für mich. Die Schülerinnen und Schüler brauchten mich nicht ...
- Oft wollten (sie) nicht, dass man sich darum kümmert, wie sie lernen ...
- Sie wollten nicht, dass ich als Lehrperson in ihre Welt der Lerngruppe eindringe“ (ebd., S. 26).

Was mit solchen Versuchen angestrebt wird, ist eine aktive Schülerschaft, der zugetraut wird, dass sie im Rahmen präziser Vorgaben aktiv handeln kann, ohne ständig Anleitung zu benötigen. Es geht um eine effektive Partizipation der Schülerinnen und Schüler, die Verantwortung für ihren Lernerfolg übernehmen, ohne darauf zu warten, motiviert zu werden. Für das oft missverständliche Thema der „Motivation“ gibt es eine Faustformel:

- Der Erfolg des Unterrichts hängt ganz wesentlich davon ab,
- dass die Schülerinnen und Schüler wissen und möglichst auch einsehen,
- warum sie lernen, was sie lernen.

Die Motivation der Schüler ist nicht zuletzt eine Folge transparenter und nachvollziehbarer Leistungserwartungen, für die gute Gründe und nicht Allerweltserklärungen zur Verfügung stehen müssen. Oft verstehen die Schüler nicht, was die Ziele des Unterrichts sind, und oft wissen sie nicht, ob überhaupt Ziele verfolgt werden, zu schweigen davon, ob am Ende die Ziele erreicht wurden oder nicht.

Transparenz setzt zweierlei voraus, nämlich klare Ziele und durchschaubare Kriterien. Bislang erheben Schulen über sich kaum Daten. Der wesentliche Grund besteht darin, dass sie nicht auf Ziele ausgerichtet sind, die in regelmässigen Abständen bilanziert werden müssen. Die Entwicklung von Schulen kann nicht einfach darin bestehen, wohlmeinende Absichten in Hochglanzbroschüren drucken zu lassen und sie einem unspezifischen Publikum mitzuteilen, oder Webauftritte zu organisieren, die gelegentlich etwas eitel als „Schulprofil“ bezeichnet werden. Vielmehr müssen sich Schulen auf echte, nämlich erreichbare Ziele beziehen, auf die ihre Entwicklung innerhalb eines bestimmten Zeitraums ausgerichtet wird. Und sie muss von dem lernen, was sie erreichen oder nicht erreichen. In der Ökonomie spricht man etwas unschön von „Outputsteuerung.“

Meine Zusammenfassung lautet so: Wenn es um die Entwicklung der Unterrichtskultur der Gymnasien geht, dann müssen die Ziele der Schulen und die verschiedenen Lernformen in ein vernünftiges Verhältnis gesetzt werden. Was immer etwas abfällig „Frontalunterricht“ genannt wird, ist eine sehr ökonomische Form der Zielerreichung, die ergänzt und auch als Form weiterentwickelt werden kann, ohne dass es klug wäre, damit einen unüberbrückbaren Gegensatz zur Projektmethode oder zum „Selbstlernsemester“ zu verbinden. Der Einsatz der Lernformen erfolgt pragmatisch, und er muss zielverträglich sein. Im Gymnasiallehrgang sind nur begrenzt Phasen des Ausprobierens möglich. Neue elektronische Formate des Unterrichts habe ich angedeutet. Für diese gilt dasselbe wie für die erweiterten Lehr- und Lernformen des Gymnasiums. Sie müssen sich im Blick auf Gütekriterien als sinnvoll herausstellen und sind kein Selbstzweck. Mathematik ist über Jahrhunderte erfolgreich mit den Medien Tafel, Kreide, Schwamm unterrichtet worden. Wenn es heute bessere Möglichkeiten gibt, wird die Schule sie nutzen.

Und noch etwas sei deutlich gesagt: Kein Lehrer und keine Lehrerin lässt sich abstrakt über seinen oder ihren Unterricht belehren. Guter Unterricht entsteht nicht einfach „in Anwendung“ didaktischer Kriterien, hinter jeder Qualität steht eine persönliche Erfahrung, die für die Akzeptanz von Vorschlägen massgebend ist. Kriterien müssen mit einer überzeugenden Praxis verbunden sein und in das individuelle Know How der Lehrkräfte eingehen. Die Praxis des Unterrichts ist komplex, durchaus störungsanfällig und abhängig von der jeweiligen Zusammensetzung der Lerngruppe. Das gilt für jede Lernform, keine verfügt über eine eingebaute Erfolgsgarantie.

Literatur

- Befragung ehemaliger Zürcher Mittelschülerinnen und Mittelschüler. Erstellt vom Statistischen Amt des Kantons Zürich. Zürich: Statistisches Amt des Kantons Zürich 2001..
- Befragung ehemaliger Zürcher Mittelschülerinnen und Mittelschüler. Erstellt vom Statistischen Amt des Kantons Zürich. Zürich: Statistisches Amt des Kantons Zürich 2004.
- Bildung Schweiz Heft 11 (2006).
- Bobertag, O.: Leistungsschätzung und Leistungsmessung in der Volksschule. In: Zeitschrift für Pädagogische Psychologie 34 (1933), S. 377-393.
- De Groot, A.D.: Fünfen und Sechsen. Weinheim 1971.
- 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.
- Grey, C.F.: Variations in the Grades of High School Pupils. Baltimore 1913.
- Grünig, B. u.a.: Leistung und Kontrolle. Die Entwicklung von Zensurenggebung und Leistungsmessung in der Schule. Weinheim/München 1999. (= Erziehung im Wandel. Hrsg.v. H. Rauschenberger, Bd. 4)
- Haecker, H.: Subjektive Faktoren im Leistungsurteil der Lehrer. In: Schule und Psychologie 18 (1971), S. 74-84.
- Hofer, M.: Die Validität der impliziten Persönlichkeitstheorie von Lehrern. In: Unterrichtswissenschaft Heft 2 (1975), S. 5-18.
- Hopp, A.-D./Lienert, G.A.: Eine Verteilungsanalyse von Gymnasialzensuren. In: Kh. Ingenkamp (Hrsg.): Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte. 6., überarb. u. erw. Aufl. Weinheim/Basel 1976, S. 250-263.
- Huberman, M.: Das Selbstkonzept. Eine Untersuchung über die Wirkung von Noten, Ranglisten und Preisen auf Kinder der Genfer Primarschule. Genève: FAPSE 1980.
- Ingenkamp, Kh.: Zur Problematik der Auslese und ihrer Bewährungskontrolle. In: Kh. Ingenkamp (Hrsg.): Pädagogisch-psychologische Untersuchungen zum Übergang auf weiterführende Schulen. Weinheim 1963, S. 7-54.
- Ingenkamp, Kh. (Hrsg.): Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte. 6., überarb. u. erw. Aufl. Weinheim/Basel 1976. (erste Aufl. 1971)
- Jäger, R.S.: Von der Beobachtung zur Notenggebung. Ein Lehrbuch. Diagnostik und Benotung in der Aus-, Fort- und Weiterbildung. M. e. Beitrag v. U. Lissmann. Landau 2000.
- Krampen, G.: Effekte von Lehrercommentaren zu Noten bei Schülern. In: R. Olechowski/E. Persy (Hrsg.): Fördernde Leistungsbeurteilung. Wien/München 1987, S. 297-227.
- Langfeldt, H.P./Tent, L.. Pädagogisch-psychologische Diagnostik. Band 2: Anwendungsbereiche und Praxisfelder. Göttingen/Bern/Toronto/Seattle 1999.
- Maag Merki, K.: Evaluation Mittelschulen - Überfachliche Kompetenzen. Schlussbericht der ersten Erhebung 2001. Universität Zürich, Pädagogisches Institut, Vervielf. Ms. Zürich: Forschungsbereich Schulqualität&Schulentwicklung 2002..
- Maag Merki, K./Leutwyler, B.: Evaluation Mittelschulen. Überfachliche Kompetenzen. Zwischenbericht der zweiten Erhebung 2004. Vervielf. Ms. Zürich: Forschungsbereich Schulqualität&Schulentwicklung 2004.
- Maag Merki, K./Leutwyler, B.: Die Entwicklung überfachlicher Kompetenzen im Gymnasium. Eine Längsschnittstudie zwischen dem 10. und 12. Schuljahr auf der

- Sekundarstufe II. Vervielf. Ms. Zürich: Forschungsbereich Schulqualität&Schulentwicklung 2005.
- Mandel, S.: Wirksamkeit der Leitbilder von Schweizer Gymnasien. Das Leitbild als Mittel zur Qualitätssicherung, Qualitätsentwicklung und Qualitätsevaluation. Bern et. al.: Peter Lang Verlag 2006. (= Explorationen. Studien zur Erziehungswissenschaft, hrsg. v. J. Oelkers, Bd. 52)
- Moser, U./Rhy. H.: Lernerfolg in der Primarschule. Eine Evaluation der Leistungen am Ende der Primarschule. Aarau 2000.
- Oelkers, J.: Zum Problem von Standards aus historischer Sicht. In: Neue Sammlung 44 (2004), S. 179-200.
- Oelkers, J.: Expertise Mittelschulen. Zuhanden der Bildungsdirektion des Kantons Zürich. Ms. Zürich 2006.
- O'Neil, H.F./Abedi, J./Miyoshi, J./Mastergeorge, A.: Monetary Incentives for Low-Stakes Tests. In: Th. Fitzner Hrsg.): Bildungsstandards. Internationale Erfahrungen - Schulentwicklung - Bildungsreform. Bad Boll: Evangelische Akademie 2004, S. 164-201.
- Prinz von Hohenzollern, J.G./Liedtke, M. (Hrsg.): Schülerbeurteilungen und Schulzeugnisse. Historische und systematische Aspekte. Bad Heilbrunn/Obb. 1991. (= Schriftenreihe zum Bayerischen Schulmuseum Ichenhausen, hrsg. v. Bayerischen Nationalmuseum, Bd. 10)
- Rank, Th.: Schulleistung und Persönlichkeit. München 1962.
- Rheinberg, F.: Leistungsmessung und Lernmotivation. Göttingen 1980.
- Richter, S.: Unterschiede in den Schulleistungen von Mädchen und Jungen. Geschlechtsspezifische Aspekte des Schriftsprachenerwerbs und ihre Berücksichtigung im Unterricht. Regensburg 1996.
- Rosenthal, R./Jacobson, L.: Pygmalion im Unterricht. Lehrererwartungen und Intelligenzentwicklung der Schüler. Weinheim 1971.
- Sacher, W.: Praxis der Notengebung. Hilfen für den Schulalltag. Bad Heilbrunn/Obb. 1984.
- Sauer, J./Gamsjäger, E.: Ist Schulerfolg vorhersehbar? Die Determinanten der Grundschulleistung und ihr prognostischer Wert für den Sekundarschulerfolg. Göttingen/Bern/Toronto/Seattle 1996.
- Schrader, F.W.: Lern- und Leistungsdiagnostik im Unterricht. In: F.E. Weinert (Hrsg.): Psychologie des Unterrichts und der Schule. Göttingen/Bern/Toronto/Seattle 1997, S. 659-699.
- Schreiber, H.: Gegen Prüflingen und Noten. In: Zeitschrift für Philosophie und Pädagogik 6 (1899), S. 31-38.
- Tent, L.: Die Auslese von Schülern für weiterführende Schulen. Göttingen 1969.
- Tresch, S.: Potenzial Leistungstest. Wie Lehrerinnen und Lehrer Ergebnismeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen. Diss. Phil. Universität Zürich (Fachbereich Allgemeine Pädagogik). Ms. Zürich 2006.
- Ulshöfer, R.: Zur Beurteilung von Reifeprüfungsaufsätzen. In: Der Deutschunterricht 1 (1949), S. 84-102.
- Ulbricht, H.: Wortgutachten auf dem Prüfstand. Eine empirische Untersuchung zur verbalen Beurteilung in der 1. und 2. Klasse der Grundschule mittels Elternbefragung und Zeugnisanalyse. Münster/New York 1993.
- Weidig, E.R.: Die Bewertung von Schülerleistungen. Weinheim 1961.
- Weinert, F.E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim/Basel 2001.
- Weise, G.: Leistungsmessung. In: J. Petersen/G.-B. Reinert (Hrsg.): Pädagogische Positionen. Ein Leitfaden für Lehrer aller Schulen. 2. Aufl. Donauwörth 1990, S. 216-230.
- Weiss, R.: Zensur und Zeugnis. Beiträge zu einer Kritik der Zuverlässigkeit und Zweckmässigkeit der Ziffernbenotung. Linz 1965. (= Wissenschaftliche Veröffentlichungen des Bundes in Oberösterreich, Bd. 3)
- Wolf, K.: Die Gerechtigkeit des Erziehers. München 1962.

Ziegenspeck, J.: Handbuch Zensur und Zeugnis in der Schule. Historischer Rückblick, allgemeine Problematik, empirische Befunde und bildungspolitische Implikationen. Ein Studien- und Arbeitsbuch. Bad Heilbrunn/Obb. 1999.