

Jürgen Oelkers

Das Gymnasium und seine Noten *)

John Hattie ist der Garant für eine realistische Sicht auf Schule und Unterricht. Er fragt nach der Wirksamkeit von Unterricht bezogen auf die Leistungen der Schüler in bestimmten Fächern und er sagt deutlich, was wirksam ist und was nicht. Seine zentrale These ist berühmt, am meisten wirksam unter einer Gruppe von Faktoren ist die Lehrperson, nicht Methoden des Unterrichts und seien sie noch so sehr Objekt von politischer Propaganda wie „selbstorganisiertes Lernen“ oder die „Altersdurchmischung“ der Lerngruppen. Es kommt vor allem auf die Qualität der Lehrpersonen an (Hattie 2009).

Hatties Erfolg hat seinen Grund auch darin, dass er Noten verteilt, ein Tatbestand, der zur Sprache des Schulfeldes perfekt zu passen scheint. Er unterscheidet die Wirksamkeit auf einer Skala, die die Unterschiede sichtbar macht.

- Noten sind Skalierungen für gut, weniger gut und schlecht.
- Sie setzen Leistungen voraus, die beurteilt werden.
- In diesem Sinne sind Zeugnisse und Noten Formen der Rückmeldung.
- Das Feedback der Lehrkräfte ist für Hattie eine der wirksamsten Formen der Beeinflussung des Leistungsverhaltens.

Darüber öffentlich zu diskutieren, ist nicht ganz leicht und auch heikel, denn das Thema ist besetzt und dies negativ. Zeugnisse und Noten stehen in der Kritik und mehr noch, sie gelten als gefährlich und werden als überflüssig hingestellt. Sie werden sozusagen benotet, meistens mit folgenden Argumenten:

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

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

Motivieren *mit* Noten ist die vermutlich meist verbreitete Motivationspraxis in der Schulwirklichkeit, aber die wird als anrüchig hingestellt, da sie dem Ideal des „intrinsischen“

*) Vortrag im Kollegium Spiritus Sanctus Brig am 18. September 2018.

Lernens widerspricht. Die Lehrerbildung liebt das „Intrinsische“, aber man lernt auch, wenn man motivationale Widerstände überwinden muss und man stelle sich vor, wohin man käme, wenn alles Lernen in der Schule von der Zustimmung intrinsischer Motivation abhängig wäre. Das Ideal des freien Lernens findet Zustimmung, aber die Wirklichkeit sieht anders aus.

Noten als *positive* Anreize sind auch deswegen suspekt, weil sie eine Hierarchie voraussetzen. Nur wenige Schüler können Bestnoten erreichen und das, so die Kritik, fordert die anderen nicht etwa heraus, sondern schreckt sie ab und hindert sie am Lernen. Es soll, mit anderen Worten, keinen Wettbewerb geben und niemand soll mit anderen verglichen werden. Das bekanntlich schon Jean-Jacques Rousseau 1762 in seinem Erziehungsroman *Emile ou de l'éducation* postuliert.

Es ist danach immer wieder versucht worden, Alternativen zu der Notenskala zu entwickeln. Radikale Entwürfe gehen davon aus, dass nur die Lernfortschritte des einzelnen Schülers beschrieben werden können und sich ein Vergleich in der Lerngruppe verbietet, weil der ohnehin nicht objektiv sein kann und zudem die unterschiedlichen Voraussetzungen der Lernenden ignoriert. Aber so leicht wird man die Noten nicht los.

- Noten setzen die Klassennorm voraus und basieren so auf einem Vergleich der Leistungen mit anderen.
- Diese Beschreibung hat sich bewährt, sie ist ökonomisch und vergleichsweise leicht zu handhaben.
- Zudem zeigt die Forschung, dass die Urteile der Lehrpersonen im Blick auf ihre Klasse im Allgemeinen verlässlich sind (Weinert 2001).
- Die empirische Kritik bemängelt allerdings die fehlenden Bezugsnormen (Fischer 2012, S. 50).

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

Der „Schulmethodus“, also der Leitfaden der Steuerung, ist von Andreas Reyher, dem Rektor des Gymnasiums in Gotha,¹ verfasst worden, erschien zuerst 1642 und erlebte zahlreiche Ausgaben. Der Methodus war Teil der ernestinischen Reformen im Fürstentum Gotha, die als erste staatliche Schulentwicklung im deutschen Sprachraum gelten können.² Verglichen mit heutigen Lehrplänen war der Methodus knapp und konzise, auf nur etwa hundert Seiten konnte alles Wesentliche gesagt werden; erst viel später wurden die Lehrplankommissionen erfunden, mit allen Risiken für die Aufblähung des Umfangs. Ich nehme nicht an, dass den vielen Fachgruppen, die für den Lehrplan 21 gearbeitet haben, der „Schulmethodus“ bekannt gewesen ist.

Seit der Ausgabe des Jahres 1662 schrieb der Methodus ein Stufenschema zur Beurteilung der Leistungen in drei Dimensionen vor. Die Beurteilung erfolgte für die jährlichen Schulexamen, die öffentlich waren und im Herbst vor der Ernte stattfanden. Die Lehrer waren der Gemeinde rechenschaftspflichtig und alle ihre Daten einschliesslich der

¹ Andreas Reyher (1601-1673) wurde 1641 als Rektor des *Gymnasiums illustre* nach Gotha berufen.

² Ernst I., der Fromme (1601-1675), war seit 1640 Herzog von Sachsen-Gotha. Auf ihn gehen umfangreiche Bildungsreformen zurück, die von der Schulpflicht ab dem 5. Lebensjahr bis zur Errichtung eines Waisenhauses reichen.

Leistungen in den Unterrichtsfächern mussten mit einem einheitlichen „Modell“ erfasst und dokumentiert werden (Methodus 1697, S. 102). Auf diese Weise waren sie leicht verständlich und stellten die Grundlage dar für die Entwicklung im nächsten Schuljahr (ebd.).

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

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

Offenbar ist dieses Schema - zeitgemäss verändert - geeignet, Leistungs- und Verhaltensdifferenzen innerhalb einer bestimmten, grösseren Lerngruppe zu beschreiben. In diesem Sinne handelt es sich um einen Standard, den nicht die Lernpsychologie, sondern die Schule selbst hervorgebracht hat und der mit einiger Wahrscheinlichkeit auch in Zukunft Verwendung finden wird.

Aber was wird dann aus der Bildungsgerechtigkeit? Wenn man als „sehr fein“, „fein“ und „fromm“ eingestuft wird, und dies auf Dauer, dann hat das ein anderes Selbstbild zur Folge, als wenn die Wertung „schlecht“, „etwas/wenig“ und „unfleissig“ lauten würde. Aber ist das ungerecht? Die Antwort lautet ja, wenn die Bewertungen *willkürlich* erfolgen würden. Ein Bewertung ist willkürlich, wenn sie

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

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

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

- Noten erfassen Leistungen im Blick auf bestimmte Aufgabenstellungen, die von Lehrkräften im Blick auf eine bestimmte Gruppe bewertet werden.

- Endnoten nach einem bestimmten Unterrichtsabschnitt, meistens am Ende eines Schulhalbjahres, werden gebildet, indem verschiedene Einzelnoten addiert und ein (gewichteter) Durchschnitt errechnet wird.

Dieses Verfahren ist leicht handhabbar und nicht so schlecht, wie die Kritik häufig annimmt. Die Lehrpersonen stellen Aufgaben und bewerten Leistungen aufgrund langjähriger Erfahrungen und handeln vermutlich in den wenigsten Fällen wirklich „willkürlich“ in dem genannten Sinne. Insofern muss man fragen, wieso sich die Notenkritik so hartnäckig immer wieder bemerkbar zu machen versteht. Wenn Medien Alternativen zur Schule propagieren, dann sind die immer notenfrei. Freilich, oft sind diese Alternativen kleine Privatschulen und nie Gymnasien.

Noten sind Teil des Zweifels an der „Bildungsgerechtigkeit“ und die sind weit verbreitet. Aber meistens werden Karikaturen ausgetauscht, wenn darüber in der Öffentlichkeit diskutiert wird. Vermeidet man die Karikaturen, dann lässt sich im Blick auf die schulische Notenpraxis festhalten:

- Schulnoten bewerten Leistungen und können nur begrenzt eine Aussage darüber machen, was die tatsächlichen Kompetenzen der Schülerinnen und Schüler sind.
- Vieles, was für das Zustandekommen der Leistung mitverantwortlich ist, wird mit einer Ziffernote auch gar nicht erfasst,
- etwa die allgemeine Leistungsfähigkeit, schulisches Engagement, Vor- und Nachteile der sozialen Herkunft oder die Motivation für bestimmte Unterrichtsfächer.

Aber jede Bewertung hat Grenzen und keine umfasst alles. Noten haben den Vorteil, dass sie einfach sind, leicht zu kommunizieren und keinen übermäßigen Aufwand verlangen. Sie passen ins Arbeitsfeld der Schule, gelten als bewährt und werden als Beschreibungsform auch ausserhalb der Schule breit angewendet.

- Niemand stört sich daran, dass Fussballprofis jede Woche Noten erhalten,
- Filmkritiker vergeben Noten ebenso wie Restaurantkritiker,
- ein Hotel ohne Noten würde man kaum buchen
- und die deutsche Stiftung Warentest könnte ohne Noten ihren Betrieb einstellen.

Das Problem ist nicht das gestufte System der Leistungsbewertung, einfach weil Leistungen, wenn sie verglichen werden, immer gestuft sind, sondern was das System erfassen soll und zu welchem Zweck es eingesetzt wird. Bei der Konzentration auf das Feindbild „Ziffernoten“ ist oft übersehen worden, wie vielfältig Leistungsbewertungen in der Schule tatsächlich sind.

- Es gibt formelle und informelle Einschätzungen des Leistungsverhaltens,
- Beurteilungen der Inhalte und der Form einer Leistung,
- schriftliche und mündliche Prüfungen,
- Selbstbewertungen, Tests, Proben, Abschlussarbeiten und vieles mehr.

Allerdings ist nicht jede Bewertung gleich wichtig, die Relevanz wird von den Lernzielen und dem Design der Lernprozesse her bestimmt. Am Ende einer Zeiteinheit steht ein Ergebnis, das beurteilt worden ist. Seine Basis sind die gezeigten Leistungen und die

Beurteilung muss imstande sein, die Unterschiede zu erfassen, sofern gleiche Ziele massgebend waren.

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

Für die Schülerinnen und Schüler in Schweizer Gymnasien war die Maturaarbeit im Übrigen 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 damals nichts zählte (Oelkers 2008, S. 203ff.).

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

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

In einer deutschen Metastudie aus dem Jahre 2007 wird die Validität der Schulnoten zur Vorhersage des Studienerfolgs nochmals detailliert beschrieben (Trapmann-Hell/Weigand/Schuler 2007). Dabei wirkt sich gerade die Breite des Notensystems positiv aus. Je höher der Durchschnitt in den Fachnoten liegt, desto besser kann der Studienerfolg vorhergesagt werden

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

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

Eine jüngere Studie über den Zusammenhang von Schulangst, Schulunlust, Anstrengungsvermeidung und Schulnoten in den Fächern Mathematik und Deutsch sieht zwischen diesen Konzepten signifikante geringe bis mittlere Interkorrelationen. Am stärksten hängen Prüfungsangst und die schulbezogene Anstrengungsvermeidung mit den Schulnoten zusammen (Weber/Petermann 2016, S. 562). Das ist eigentlich trivial: Wer Angst vor Prüfungen hat, vermeidet Anstrengungen, weil die Vorstellung vorherrscht, die Prüfung sei ohnehin nicht zu bestehen. Andererseits minimiert die Anstrengungsvermeidung die Chancen des Bestehens, wenn die Prüfung nicht vermieden werden kann.

Generell wirken sich Prüfungsangst, manifeste Angst sowie Anstrengungsvermeidung negativ aus.

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

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

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

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

Die Schülerinnen und Schüler lernen auch *subversiv*, nämlich wie die Anforderungen des Unterrichts umgangen werden können, oder *strategisch*, nämlich wie sich mit einem Minimum an Aufwand ein Maximum an Ertrag erreichen lässt. Sie kalkulieren im Blick auf die Ziele den notwendigen Ressourceneinsatz und gehen keineswegs immer „intrinsisch motiviert“ vor, schon weil kaum eine Schülerin und kaum ein Schüler sich für das gesamte Angebot der Schule gleich interessiert. Die Schüler machen immer einen Unterschied, was sie gerne lernen und was nicht.

Weiter bilden die Schülerinnen und Schüler über Schule und Unterricht informelle Meinungen heraus, die das tatsächliche Lernen oft mehr beeinflussen als das offizielle

Lernsetting der Schule. Es handelt sich dabei um hoch elaborierte Kognitionen, die vor allem in der Peer-Kommunikation gebildet und stabilisiert werden (Chiapparini 2012).

Die Anstrengungsbereitschaft verteilt sich nicht einfach nur mit dem Interesse, wie oft angenommen wird, sondern reagiert auch auf Notlagen, etwa auf die Folgen der Nichterreichung von Lernzielen oder die drohenden Selektionen an den Schnittstellen. Probleme wie diese werden in der didaktischen Literatur gemieden oder normativ bestritten, obwohl sie nicht verschwinden werden und das Lernen massiv beeinflussen. Auch im Falle der Lernstrategien überwiegen die Modellannahmen, die unabhängig vom tatsächlichen Erfahrungsraum „Schule“ gedacht werden.

Die von der Kritik gewählte Sprache, die etwa von „Notenterror“ spricht,³ macht es schwer, in Noten etwas anderes zu sehen, als willkürliche Instrumente, die Kinder traumatisieren können. Darauf haben viele deutsche Grundschulen reagiert, die mit behördlicher Genehmigung „notenfrei“ ausgerichtet sind und damit den „Notenterror“ an die nächste Stufe weitergeben. Auch in der Schweiz scheint es zunehmend schwer zu sein, Noten für Primarschülerinnen und Primarschüler als geeignet anzusehen.

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

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

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

³ Der Ausdruck stammt aus den westdeutschen Schulkämpfen der siebziger Jahre („Gegen Notenterror“: Der Spiegel Nr. 2 von 9. 1. 1978, S. 64/65).

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

Die EDK hat die unterschiedlichen Regelungen zur Leistungsbeurteilung in den Kantonen erhoben. Nur die Beurteilung im Kindergarten (HarmoS 1-2) erfolgt ohne Ziffernoten, aber nicht ohne Beurteilung, für die meist vorgegebene Beobachtungsbögen eingesetzt werden. Damit werden Lernfortschritte beschrieben und Rückmeldungen an die Eltern dokumentiert. In der Primarschule und Sekundarschule I wird an der Praxis der Zifferzeugnisse festgehalten, die aber nicht einheitlich geregelt ist und kantonale Unterschiede kennt.⁴

- Primarstufe (HarmoS 3-8) und Sekundarstufe I: In allen Kantonen werden den Schülerinnen und Schülern Zeugnisse mit Noten ausgestellt, dies allerdings nicht in allen Kantonen ab dem gleichen Zeitpunkt.
- In der grossen Mehrheit der Kantone setzen Zeugnisse mit Noten zwischen dem 1. und dem 4. Schuljahr (HarmoS 3-6) ein.
- In den Schuljahren, in welchen noch keine Zeugnisse mit Noten ausgestellt werden, kommen Lernberichte ohne Noten oder andere Beurteilungsverfahren zur Anwendung.
- Lernberichte ohne Noten werden häufiger in den unteren Schuljahren ausgestellt.⁵

Im Gymnasium gibt es Zeugnisse praktisch überall nach dem ersten Semester.⁶ Spielraum ist nur vorhanden bei der Leistungsbeurteilung bis hin zur Zeugnisnote. Deutlich wird auch, dass sich die Primarschule stärker differenziert hat als die beiden Sekundarstufen und die Sekundarstufe I eher mit neuen Formen der Leistungsbeschreibung experimentiert als die Sekundarstufe II.

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

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

⁴ <http://www.edk.ch/dyn/15780.php>

⁵ Ebd.

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

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 Masstäben 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 Schulfächer⁷ erfahren werden. Wer gute Leistungen in der Unterrichtssprache erzielt, hat in der Regel keinen Vorteil für den Erwerb der zweiten Landessprache.

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

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

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

Jahrzehntelange Erfahrungen mit Lernberichten und Ähnlichem zeigen die Steigerung der Komplexität und damit einhergehend der drohenden Unverständlichkeit.⁸ Noten müssen klar und verständlich sein, die Abstände im Leistungsverhalten wiedergeben und hohe und tiefe Grade kennen. An dieser Anforderung sind die Alternativen zu messen und so lange

⁷ Zu unterscheiden von längeren Sprachaufenthalten oder von der zweisprachigen Matur.

⁸ Das zeigen auch Schulversuche: Bos et. al. (2010).

keine besseren Alternativen zum Ziffernsystem vorliegen, wird dieses auch weiterhin die Praxis bestimmen.

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

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

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

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

Das unterscheidet die neuen 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 HarmoS-Konkordat auch „Mindeststandards“ vor; alle Schüler müssen eine *bestimmte* Stufe erreichen, nicht den gesamten Lehrplan, was weder möglich noch sinnvoll wäre.

Viele Bildungssysteme verfügen heute über Standards, Leistungstests und komplexere Formen der Evaluation. Das elaborierteste System in Europa ist seit 1988 in England aufgebaut worden, und selbst in den liberalen Gesamtschulsystemen Skandinaviens werden inzwischen Standards und Tests verwendet. Es gibt also überall mehr als nur die gewohnte Prüfungspraxis in den einzelnen Schulen.

- Die Frage ist, wie diese Praxis so verändert werden kann, dass sie zur Gesamtentwicklung passt und zugleich den einzelnen Schulen sichtbare Vorteile bringt.
- Neue Formen der Beurteilung, anders gesagt, müssen einen Beitrag zur Unterrichtsentwicklung leisten, sonst werden sie nicht akzeptiert.

Für die Schweizer Gymnasien sind Leistungsstandards nichts Neues, neu wären externe Leistungstests, die derzeit aber nicht abgesehen werden können. Sie würden die internen Leistungsbewertungen auch nicht ersetzen, sondern allenfalls ergänzen. An flächendeckende Tests ist nicht gedacht, und auch ein „High-Stakes-Testing,“ also Tests mit

Folgen für die Schule oder gar für die einzelnen Lehrkräfte,⁹ steht nicht zur Diskussion und hätte auch keine Akzeptanzchancen. Im Unterschied zu den Vereinigten Staaten kann man in der Schweiz nicht an den Lehrern vorbei Politik machen.

Aber es geht auch ganz anders: In England ist künftig für das zentrale Examen am Schluss der Sekundarstufe I (EBacc)¹⁰ eine eigene Prüfungsbehörde zuständig, schon vorher waren die Abschlussprüfungen ausgelagert, aber die Schulen konnten zwischen verschiedenen Behörden (Boards) wählen und so auf die Leistungsfähigkeit ihrer Schüler Rücksicht nehmen.

- Der Grund für die Verschärfung ist die Vermeidung von Ungerechtigkeit; „gerecht“ soll sein, wenn ein einziges „examination board“ gleiche Aufgaben für alle stellt.
- Wer die zentrale Prüfung nicht schafft und sie wiederholen kann, wird aufwändig beraten, auch das soll die Bildungsgerechtigkeit erhöhen.

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

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

- In einem transparenten System der Leistungsbeurteilung wären Noten gestufte Formate mit einem gemeinsamen Massstab.
- Die Notengebung würde sich nach Kriterien vollziehen, die für alle Beteiligten klar sind und die fortlaufend kommuniziert werden
- Dann gibt nicht jede Lehrkraft „ihre“ Noten.

Vielmehr werden die Notenstufen pro Fach oder Fachbereich festgelegt und auf das eingestellt, was als Leistungsstandard gilt. Standards unterscheiden sich nach Stufen und darauf beziehen sich die Noten. Die einzelnen Stufen erhalten eine schriftliche Fassung, die 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:

⁹ Einzelne Schuldistrikte in den Vereinigten Staaten, etwa in Cincinnati, koppeln die Löhne an die Ergebnisse der Leistungstest.

¹⁰ „English Baccalaureate Certificate“. Der Abschluss ist 2010 beschlossen und ab 2012 eingeführt worden. Er umfasst fünf Kernfächer, Englisch als Muttersprache, eine Fremdsprache, Mathematik, eine Naturwissenschaft sowie Geschichte oder Geografie.

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 oder Auch verschiedener Stufen vereinbart werden.

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

Diese Kompetenzen wurden in den gültigen Empfehlungen von 2008 so gefasst:

- Kritisch-forschendes Denken als gymansiale Schlüsselkompetenz
- Selbstständigkeit und Selbstverantwortung
- Verstehen wissenschaftlicher Texte
- Abfassen schriftlicher Abhandlungen
- Beherrschung der Sprachen (D, F, E)
- Quellenkritik
- Reflexion und Kritikfähigkeit
(HSGYM 2008, S. 42-46).¹¹

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

Die Kernfrage ist, was bewertet man wie mit einer Note, vorausgesetzt die Leistungsniveaus des Gymnasiums. Das „was“ bestimmen in fachlicher Hinsicht der Lehrplan

¹¹ Siehe auch: <https://www.hsgym.ch/analyse-and-empfehlungen/ueberfachliche-kompetenzen>

und das konkrete Curriculum der Fachschaften. Eine „kriteriale Norm“ für die Noten direkt aus dem Lehrplan abzuleiten, ist weder sinnvoll noch praktikabel. Die Lehrpersonen müssen an dieser zentralen Norm ihres Unterrichts direkt beteiligt sein. Insofern ist der Bezugspunkt nicht einfach der Lehrplan, sondern die konkreten Zielsetzungen der Schule, so wie die Fachschaften sie für einen bestimmten Zeitraum - ein Semester, ein Jahr oder länger - festgelegt haben.¹²

Das „wie“ bezieht sich auf die Abfolge der Leistungen und der Leistungsbewertung während eines Zeitraumes. Leistungen setzen Aufgaben voraus, die Güte der Aufgaben ist der Schlüssel zur Leistungsbewertung und so der Noten.

- Aufgaben beziehen sich auf fachliche Probleme, setzen Vorwissen voraus, bezeichnen Anschlüsse und legen Strategien der Bearbeitung nahe.
- Die Note bewertet die Problemlösung, den Lernfortschritt gemessen an vorangehende Aufgaben und die Lernstrategien.
- Ein enges Korsett von Kompetenzstufen ist dabei nicht sehr hilfreich.

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

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

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

Zeugnisse basieren auf Noten. Noten sind wie gesagt in bestimmter Hinsicht prognosefähig. Wer einen hohen Leistungsschnitt erreicht, wird mit einiger Sicherheit auch in nachfolgenden Bildungsgängen gute bis sehr gute Leistungen zeigen. Das gilt mit ähnlicher Wahrscheinlichkeit auch für den umgekehrten Fall. Allerdings werden mit Noten nicht direkt

¹² Im Kanton Zürich haben die Gymnasien eigene Lehrpläne, aber auch die müssen im Blick auf die Praxis konkretisiert werden.

Fähigkeiten und Fertigkeiten beschrieben. Sie beziehen sich auf Leistungen im Blick auf Aufgaben in bestimmten Fächern oder Lernbereichen.

Gängige Ergänzungen zum Typus des Ziffernzeugnis sind individuelle Lernberichte, die die Lehrkräfte abfassen oder fortlaufende Portfolios, die von den Schülerinnen und Schülern erstellt werden. Oft dokumentieren Lehrerinnen und Lehrer fortlaufend ihren Unterricht, notieren Auffälligkeiten und reflektieren ihre Massnahmen. Elaborierte Lernberichte sind aufwändig zu erstellen, sie sind zudem interpretationsabhängig und benutzen oft sprachliche Formen, die für die Lesenden gewöhnungsbedürftig sind. Sie können als zusätzliche Dokumente eingesetzt werden, allerdings muss geklärt werden, welcher rechtliche Status ihnen zu kommt.

Portfolios der Schüler dokumentieren die Themen und Gegenstände ihrer Lernarbeit und enthalten oft auch Lerntagebücher, mit denen die Arbeit reflektiert wird. Portfolios sind häufig sehr umfangreich, sie machen den Lernprozess auf Seiten der Schüler sichtbar und sollten als eigenständige Lernleistungen anerkannt werden. Als Anlage zu Zeugnissen haben sie illustrierenden Charakter, aber ohne wirkliche Benotung durch die Lehrkräfte sind sie für die Eltern, die künftigen Abnehmer und die Lernenden selbst zu wenig aussagekräftig.

- Die Schülerinnen und Schüler wollen wissen, wo sie an einem bestimmten Zeitpunkt im Vergleich mit Anderen und/oder einer Bezugsnorm stehen.
- Die Dokumentation des eigenen Lernweges ist eine Sache, die Kriterien geleitete Bewertung und Rangierung der Leistung durch Dritte eine andere.

Als Ergänzung zur Bewertung durch die Lehrpersonen können auch Leistungstests nützlich sein, die inzwischen auch in der Schweizer Volksschule Einzug gehalten haben. Die bekanntesten Formen sind „Klassenscockpit“ und „Stellwerk“, beides inzwischen bewährte Instrumente zu einer vergleichenden Leistungsbeurteilung, die über die Klassennorm hinausgeht und sich auf einen kantonalen Durchschnitt bezieht. Sie dienen nicht der Selektion, sondern der Verbesserung des Lehrerurteils.

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

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

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

- Jeder Schüler und jede Schülerin beurteilt alle seine oder ihre Lehrkräfte, wengleich weitgehend informell und oft ohne Nutzen für den Unterricht.

- Studien zu den Erwartungen der Schülerschaft, darunter Absolventenbefragungen aus Zürich, zeigen, dass die Gymnasiastinnen und Gymnasiasten es inzwischen für eine professionelle Selbstverständlichkeit halten,
- dass *sie* mit einem formalisierten Verfahren ihre Lehrkräfte beurteilen müssten.

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

Im Blick auf das Thema und die Fragestellung kann ich so schliessen: Bei der Bewertung von Leistungen in der Schule werden Noten und Zeugnisse als Formen des verbindlichen Feedbacks weiterhin eine zentrale Rolle spielen, die Instrumente sind bewährt und begrenzen den Aufwand. Aber die Notengebung kann verbessert werden. Ich habe vier Aufgaben näher beleuchtet,

- Klarheit der Kriterien, die Präzisierung der Stufung,
- die zur Bewertung passende Aufgabenkultur und die schulischen Lernziele als Bezugsnorm.
- Die bessere Berücksichtigung des Lernwegs kann fünftens durch Portfolios erreicht werden.
- Insgesamt also eine lösbare Aufgabe.

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

- Der Gründer des Pädagogischen Seminars der Universität Leipzig, der Jurist und Philosoph Tuiskon Ziller (1884, S. 240),¹³ hatte dafür eine plausible Regel:
- Der Zögling, wie man die Schüler im 19. Jahrhundert nannte, „darf durch den Unterricht nicht geistig schwächer werden“.¹⁴

Literatur

¹³ Tuiskon Ziller (1817-1882) war Gymnasiallehrer und habilitierte sich als Jurist. 1854 begann er mit Vorlesungen in Pädagogik, 1861 erfolgte die Gründung des Pädagogischen Seminars, das ein Jahr später mit einer Übungsschule verbunden wurde. Zillers *Einleitung in die Allgemeine Pädagogik* von 1856 ist die Begründungsschrift des Pädagogischen Herbartianismus.

¹⁴ Sperrung im Zitat entfällt.

- Beurteilungen und Schullaufbahnentscheide. Über das Fördern, Notengeben und Zuteilen. Hrsg. Bildungsdirektion Zürich Volksschulamt. 3. überarb. Aufl. Zürich: Lehrmittelverlag des Kantons Zürich 2013.
- Bos, W./Beutel, S.-I./ Berkemeyer, N./Schenk, S.: LUZI. Leistungsbeurteilung ohne Ziffernzeugnisse. Abschlussbericht der wissenschaftlichen Begleitforschung. Dortmund: IFS 2010.
- Chiapparini, E.: Ehrliche Unehrlichkeit. Eine qualitative Untersuchung der Tugend Ehrlichkeit bei Jugendlichen an der Zürcher Volksschule. Opladen/ Berlin/Toronto: Budrich 2012.
- Dunkake, I./Kiechle, Th./Klein, M./Rosar, U.: Schöne Schüler, schöne Noten? Eine empirische Untersuchung zum Einfluss der physischen Attraktivität von Schülern auf die Notenvergabe durch die Lehrpersonen. In: Zeitschrift für Soziologie Band 41, Heft 2 (April 2012), S. 142-161.
- 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.
- Hattie, J.: Visible Learning. A Synthesis of Over 800 Meta-Analyses Relating to Achievement. London/New York: Routledge 2009.
- HSGYM - Hochschule und Gymnasium: Hochschulreife und Studierfähigkeit. Zürcher Analysen und Empfehlungen zur Schnittstelle. Zürich: HSGYM 2008.
- Methodus Oder Bericht/Wie Nebst Göttlicher Verleyhung die Knaben und Mägdlein auf den Dorffschaften/und in den Städten/die untere Classes der Schul-Jugend im Fürstenthum Gotha/kurtz-und nuetzlich unterrichtet werden können und sollen. Auf gnädigsten Fürstlichen Befehl aufgesetzt. Gotha: Bey Christoph Reyhern 1697. (gedr. 1698)
- Oelkers, J.: Die Qualität der Gymnasien. Eine Expertise für den Kanton Zürich. Bern: h.e.p. Verlag 2008.
- Olechowski, R./Rieder, K. (Hrsg.): Motivieren ohne Noten. Wien u.a.: Verlag Jugend und Volk 1990. (= Schule, Wissenschaft und Politik, Band 3)
- Trapmann, S./Hell, B./Weigand, S./Schuler, H.: Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - Eine Metaanalyse. In: Zeitschrift für pädagogische Psychologie Band 2, Heft 1 (2007), S. 11-27.
- Weber, H.M./Petermann, F.: Der Zusammenhang zwischen Schulangst, Schulunlust, Anstrengungsvermeidung und den Schulnoten in den Fächern Mathematik und Deutsch. In: Zeitschrift für Pädagogik Band 62, Heft 4 (Juli/August 2016), S. 551-570.
- 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.