

Gibt es „Notenwahrheit“? Über Sinn und Unsinn von Ziffernnoten

Dr. Ferdinand Eder

Vortrag am 30.5.2019 in Krimml

Einleitung: Regierungsprogramm

Das Koalitionsabkommen der nunmehr abgewählten Bundesregierung verfolgte das Ziel „Notenwahrheit wieder herstellen“. Die Formulierung unterstellt, dass es „Notenwahrheit“ prinzipiell gebe, dass sie offenbar irgendwann verloren gegangen sei, aber doch wiederhergestellt werden könne.

„Notenwahrheit wieder herstellen“

„Überarbeitung und Präzisierung der Benotungssystematik (einschließlich verbaler Benotung) für alle Schultypen und Schulstufen. Aufbauend auf einer klaren 5-teiligen Notenskala („Sehr gut“ bis „Nicht Genügend“) für alle Schultypen erfolgt eine genaue Definition welche Note vergeben werden kann bzw. muss.“

Was signalisiert der Begriff „Wahrheit“ in diesem Zusammenhang?

Den Koalitionspartnern schwebte offenbar vor, die im Schulunterrichtsgesetz bereits bestehende Definition der Notenstufen zu präzisieren. Ausgehend von dieser wäre dann „Wahrheit“, dass die Noten wiedergeben, in welchem Ausmaß Schülerinnen und Schüler den Lehrstoff beherrschen bzw. die mit der Beherrschung des Lehrstoffs verbundenen Kompetenzen aufweisen.

Anders ausgedrückt: **Noten sollen ein „wahres“ Abbild des Wissens und Könnens sein, über das ein Schüler/eine Schülerin verfügt.**

Dies ist ein hoher Anspruch an eine Skala, die nur über 5 Ausprägungen verfügt.

Auf der anderen Seite gibt es einen schon lange anhaltenden wissenschaftlichen und zum Teil auch gesellschaftlichen Diskurs, welche Art von Leistungsrückmeldung und Leistungsbeurteilung für das Lernen in der Schule und für die gesellschaftlichen Aufgaben und Funktionen der Schule am sinnvollsten und wirksamsten ist. Dieser Diskurs weist in die Richtung, dass Noten ein wenig effizientes und vor allem verzerrendes Instrument sind, mit dem mehr negative als positive Begleitfolgen verbunden sind.

Vor diesem Hintergrund und aus diesem Anlass scheint es sinnvoll, eine Art Faktencheck zu machen, wie es um die Notenwahrheit derzeit bestellt ist, und wie es um die Möglichkeit steht, sie allenfalls wiederherzustellen.

Ziele und Inhalte des Vortrags

- 1 Kurze Anmerkungen zum Selbstverständnis der Schule in Österreich
- 2 Wie ist es zurzeit um Notenwahrheit bestellt – Analyse von Daten
 - 2.1 „Wahrheit“ im Prozess?
 - 2.2 „Wahrheit“ im Ergebnis?
- 3 Drei Fragen an die Ziffernote
 - 3.1 Fördern Noten das Lernen?
 - 3.2 Tragen Noten zur Gerechtigkeit bei?
 - 3.3 Fördern Noten die Persönlichkeitsentwicklung?
- 4 Zusammenschau und Implikationen
- 5 Alternativen?

Natürlich können die meisten dieser Fragen in einem kurzen Vortrag nur überblicksweise angesprochen werden. Insgesamt geht es aber darum, einen Beitrag für eine evidenzbasierte Diskussion über den Sinn und möglichen Wert von Noten zu führen.

1 Kurze Anmerkungen zum Selbstverständnis der Schule in Österreich



Anmerkungen zur Situation der Leistungsbeurteilung in Österreich

- Noten werden (fast) ausschließlich von den Personen vergeben, die auch vorher den Unterricht erteilt haben („teacher made assessment“)
- Definition von Notenstufen als Versuch einer groben Standardisierung
- **Selektionsorientierung** als traditionelles Merkmal des Schulsystems
- (Zu) viele schlechte Noten als Folge

Bis vor Kurzem erfolgte in der österreichischen Schule die gesamte Beurteilung der Schülerleistungen durch etwas, das man in den angloamerikanischen Ländern als „Teacher-made tests“ oder „teacher-made assessment“ bezeichnet: Alle Beurteilungen werden von jenen Personen vergeben, die vorher auch den Unterricht erteilt haben. Ein solcher Zugang hat Vor- und Nachteile, ist aber auf jeden Fall mit der Gefahr verbunden, dass Beurteilungen unterschiedlicher Lehrpersonen nicht unmittelbar miteinander vergleichbar sind. Erst mit der Einführung der standardisierten Matura gibt es ein erstes Format, das essentiell davon abgeht.

Um die mit Lehrerurteilen verbundene Unschärfe zu mildern, wurden die Noten erstmals im Schulunterrichtsgesetz 1974 definiert und der ganze Prozess der LF und LB in einer Verordnung geregelt.

Erst im Kontext sog. Bildungsexplosion, wo sichtbar wurde, dass Bildung und Noten in einem hohen Ausmaß bestimmen, welche gesellschaftlichen Positionen erreicht werden können, wurde die Frage nach der Gerechtigkeit virulent: Ist die schulische Leistungsbeurteilung ein Instrumentarium, das sich gegenüber allen Schichten und Sparten der Gesellschaft fair verhält, oder gibt es einen Bias, eine nicht berechnete Begünstigung bestimmter Schichten oder Gruppen. Eine der damaligen Antworten war: Die Schule ist eine Mittelschicht-Institution, die sich vor allem gegenüber Aufsteigern aus den unteren sozialen Schichten

abwehrend verhält. Damals setzten – teils gegen massiven Widerstand der Landesvertretungen – erste gesellschaftliche Gegenmaßnahmen ein: Erleichterung der Nachprüfungen, Aufsteigen mit einem Nichtgenügend können durchaus als bildungspolitische Korrekturen der Leistungsbeurteilung in den Schulen verstanden werden.

Hinter dieser abwehrenden Haltung stand und steht ein Einstellungsproblem: Schulen und Lehrpersonen sahen in schlechten Noten häufig ein Qualitätsmerkmal, - einen Hinweis, dass es sich um eine anspruchsvolle strenge Schule bzw. Lehrperson handle, die ihrem Auftrag nach Selektion der Leistungstüchtigsten in besonders effizienter Weise nachkam.

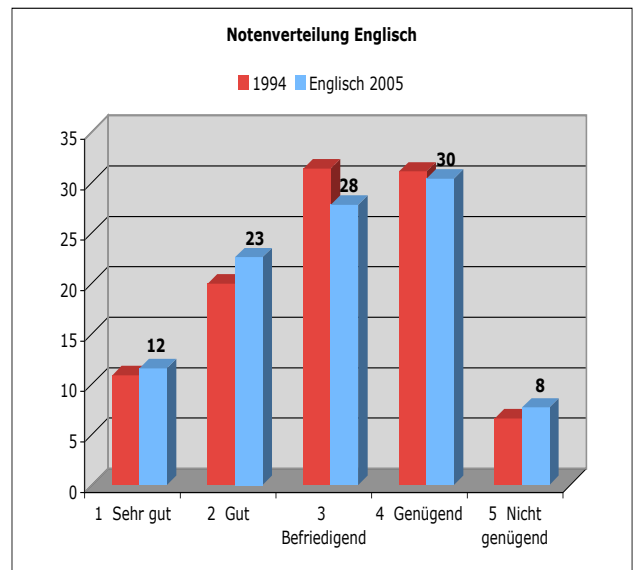
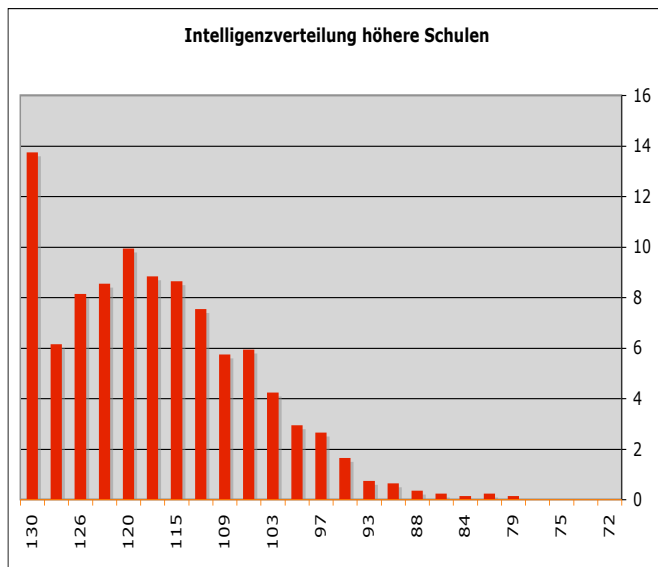
In Zeiten, in denen die höheren Schulen auf knappe Elitepositionen in der Gesellschaft vorbereiteten, war dies eine zweckrationale und nachvollziehbare Haltung.

<p>Traditionelles Verständnis:</p> <p>Selektionsorientierung: Auf allen Stufen des Systems werden Ausleseprozesse durchgeführt durch</p> <ul style="list-style-type: none"> • Wiederholung von Schulstufen • Notenabhängige Selektionsentscheidungen (VS-AHS) <p>Ziel: Beschränkung des Zugangs auf jeweils wenige besonders Geeignete</p>	<p>Neues Verständnis:</p> <p>Förderorientierung Auf allen Stufen des Systems soll individuelle Förderung erfolgen („No Child left behind“)</p> <p>Ziel: Möglichst viele junge Menschen sollen möglichst anspruchsvolle Bildungsziele erreichen (OECD; EU-„Wissensgesellschaft“)</p>
---	--

Aus der dominierenden Selektionsorientierung resultiert:

- extrem häufige Leistungsfeststellungen und –beurteilungen
- Notwendigkeit von einfachen und pragmatischen Kalkülen
- Formalistische Handhabung (z.B. Notendurchschnitte)
- Verknappung guter Noten

Ein Abbild dieser Haltung findet man in der Notenverteilung in den höheren Schulen im Vergleich zur Intelligenz der Schülerinnen und Schüler



Grafik 1: Intelligenzverteilung im Vergleich zur Notenverteilung. Schüler von Allgemeinbildenden höheren Schulen Österreichs, 9.-12. Schulstufe. Eingetragen sind die Prozentwerte einer für Österreich repräsentativen Stichprobe (N = ca. 2400)

Allerdings: Diese Zeiten sind längst vorbei. Bildungsadministration und Gesellschaft stellten zunehmend öfter die Frage: Sind Schulen, die viel „Ausschuss“ produzieren, wirklich „gute“ Schulen?

Inzwischen gilt im Wettbewerb der Länder und Staaten: Je mehr Schülerinnen und Schüler höhere Bildungslevels erreichen, desto besser. Die Leistungsstärksten werden ja ohnehin mit dabei sein.

Dieses veränderte gesellschaftliche Anforderungsbild muss mitbedacht werden, wenn über Noten diskutiert wird. Vieles, was in den Schulen geschieht, resultiert aus dem Aufeinandertreffen alter Vorstellungen von Selektion der Besten vs neuer Anforderungen im Sinne von „No Child left behind“ .

2 Wie ist es zurzeit um Notenwahrheit bestellt – Analyse von Daten

Im Hinblick auf diese Frage kann man unterscheiden zwischen

- Wahrheit im Prozess der Notengebung, und
- Wahrheit im Ergebnis.

Ersteres zielt auf mögliche Verzerrungen bei der Vergabe von Noten, das zweite auf ihren sachlichen Informationsgehalt.

Beide Fragen sind natürlich nicht voneinander unabhängig – Verletzungen der Wahrheit im Prozess führen unweigerlich zu Beeinträchtigungen der Wahrheit im Ergebnis – letztere können aber auch noch andere Ursachen haben.

„Notenwahrheit“ im Prozess

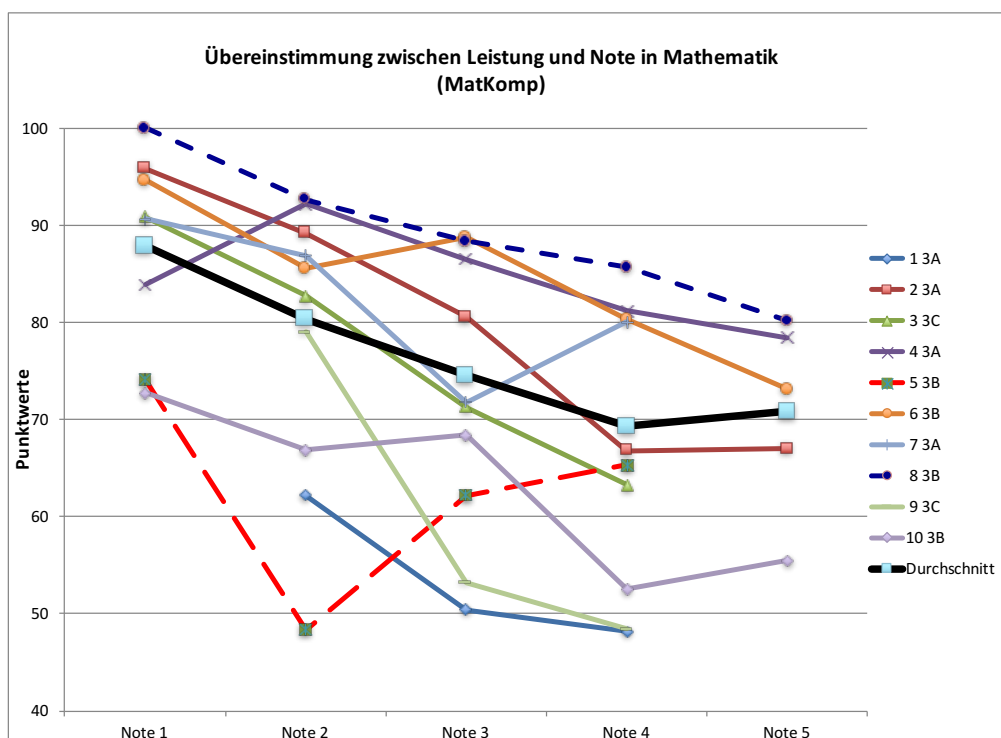
- Die Analysen zu diesem Bereich stehen unter der generellen Annahme, dass es – weil wir es fast ausschließlich mit „teacher made assessments“ zu tun haben – notwendigerweise auf allen Ebenen und in allen Sektoren des Bildungssystems zu Verzerrungen in der Notengebung kommt, die nicht den tatsächlichen Leistungen entsprechen.

Meine Analysen zu diesem Bereich stehen unter der generellen Annahme, dass es – weil wir es fast ausschließlich mit „teacher made assessments“ zu tun haben, *notwendigerweise auf allen Ebenen und in allen Sektoren des Bildungssystems zu Verzerrungen in der Notengebung kommt*. Solche zu finden und aufzuzeigen wird zunehmend leichter, weil sich in den letzten Jahrzehnten durch die Teilnahme an den internationalen Vergleichsuntersuchungen und durch die Messung der Bildungsstandards die Datenlage deutlich verbessert hat.

These 1: Noten sind nur im Referenzrahmen einer Schulklasse einigermaßen valide.

Die Forschung stimmt überein, dass Lehrpersonen relativ gut in der Lage sind, durch die von Ihnen angewandten Instrumente der Leistungsmessung die Schülerinnen und Schüler in eine Rangfolge zu bringen in der Weise, dass bessere Noten tatsächlich auch mit besseren Leistungen verbunden sind.

Grafik 2 zeigt die ersten 10 Klassen aus dem Datensatz einer Untersuchung, in der einerseits mathematische Kompetenzen mit einem an TIMSS angelehnten Untersuchungsverfahren gemessen wurden (Gesamtwert) und zugleich die Noten in Mathematik erhoben wurden. Eingetragen sind die durchschnittlichen Punktwerte, die mit den entsprechenden Notenstufen verbunden sind.



Grafik 2: Mathematik-Leistungsprofile von Schulklassen, aufgeschlüsselt nach Noten.

Leseanleitung:

In Klasse 8 3B (oberste Linie) haben Schüler/innen mit „Sehr gut“ 100 Punkte erreicht, jene mit gut 93, mit Befriedigend 88, mit Genügend 85 und mit Nichtgenügend 80 Punkte.

Quelle: Datensatz Projekt MATKOMP.

Ähnliche, wenn auch nicht immer so homogene Abstufungen finden sich in fast allen Klassen. Eine wirkliche Ausnahme bildet lediglich Klasse 5 3B, wo 3 Schüler/innen mit der Note Gut die schlechtesten Durchschnittswerte überhaupt aufweisen. Aber schlechte Werte können viele Gründe haben.

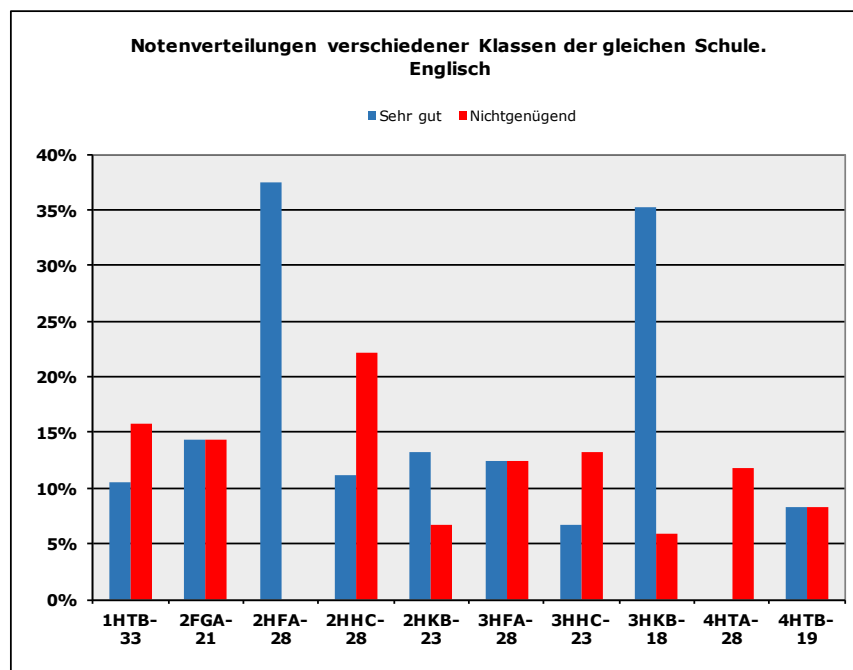
Die fett markierte Durchschnittsline zeigt die zentrale Tendenz (ohne Berücksichtigung der Klassengrößen): Im Durchschnitt gibt es eine klare Rangfolge. Lediglich der leicht gegenläufige Unterschied zwischen 4 und 5 gibt ein wenig zu denken.

Was uns die Grafik aber auch zeigt:

Leistungen, die in der einen Klasse mit einem Nichtgenügend verbunden sind, reichen in einer anderen für ein Sehr gut. Und vice versa.

Dies belegt bereits ansatzweise

These (2): Bereits innerhalb einer Schule können Noten trotz vergleichbarer Leistungen SEHR verschieden sein



Grafik 3: Notenverteilungen innerhalb der Klassen einer Schule. Datensatz Projekt „Schulprogramm an berufsbildenden Schulen“. Anzahl der Schüler/innen in der Klassenbezeichnung angeführt.

Grafik 3 zeigt die Verteilung der Noten „Sehr gut“ und „Nichtgenügend“ innerhalb der gleichen Schule. Es handelt sich um eine berufsbildende höhere Schule. Wir sehen: Es gibt eine Klasse mit fast 40% „Sehr gut“ und ohne ein Nichtgenügend, aber auch eine gleich große Klasse ohne ein einziges Sehr gut, aber mit 12% Nichtgenügend. Insgesamt variieren also die „Sehr gut“ von 9 bis fast 40%, die „Nicht genügend“ von 0 bis 22 %.

Auch wenn es dafür gute Gründe geben mag, tun sich viele Fragen auf:

In welcher Klasse bzw. bei welchem Lehrer würden Eltern ihr Kind am liebsten unterbringen wollen?

These (3) Schulnoten sind zwischen Schulen bzw. Schultypen nicht vergleichbar

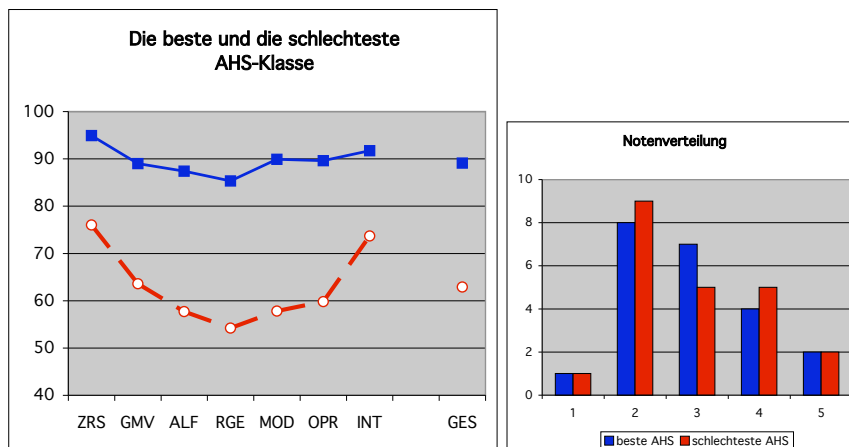
Beispiel: Mathematische Kompetenzen (MATKOMP, basierend auf TIMSS)

In einer Untersuchung zur Ausprägung mathematischer Kompetenzen an Schulen der Sekundarstufe I (7. und 8. Schulstufe) wurden die folgenden Kompetenzen erhoben:

ZRS Zahlenverständnis, Rechnen und Grundlagen der Statistik
GMV Größen, Maße und Verhältnisse
ALF Algebra und Funktionen
RGE Raumvorstellung und Grundtatsachen der Geometrie
MOD Modellbilden
OPR Operieren, Rechnen
INT Interpretieren
GES Gesamtwert

Für die einzelnen Bereiche sind Punktwerte von 0 – 100 möglich.

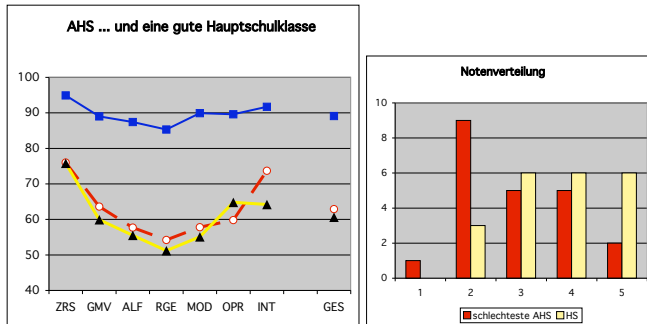
Wir haben aus dem Datensatz die von den Leistungen her beste und die schlechteste Klasse eruiert. Ihre Werte sind in Grafik 4 gegenübergestellt.



Grafik 4: Leistungs- und Notenverteilungen in zwei AHS-Klassen. Eingetragen sind die Durchschnittswerte aus Mathematik

Der Vergleich der Noten aus den beiden Klassen brachte eine Überraschung: die Notenverteilungen sind – obwohl sozusagen Welten zwischen den Leistungen liegen – beinahe identisch.

Daraufhin eruierten wir eine Hauptschulklasse mit einem Leistungsprofil, das dieser AHS-Klasse möglichst ähnlich war, und wurden entsprechend fündig: es gab eine Klasse zusammengesetzt aus Schüler/innen aller drei Leistungsgruppen – die ein praktisch identisches Profil aufwies.



Grafik 5: Notenverteilungen zweier leistungsgleicher Klassen aus AHS und Hauptschule.

Bei der Verteilung der Noten ergab sich eine weitere Überraschung: Die Beurteilung war um Vieles schlechter als jene in der AHS; kein Sehr gut, wenig Gut, dafür aber Genügend und Nicht genügend. Dazu kommt, dass diese negativen Einstufungen aus der zweiten und aus der dritten Leistungsgruppe stammen, und schulrechtlich daher noch einmal abgewertet sind.

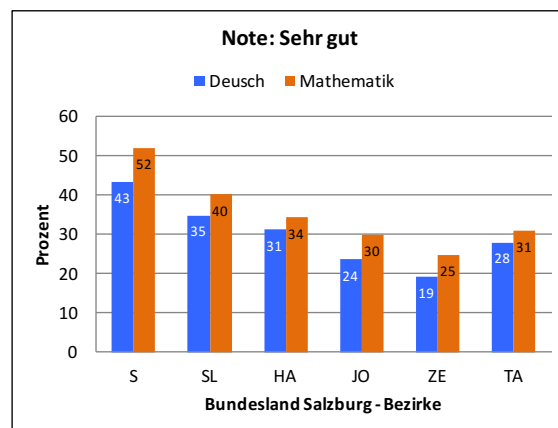
Innerhalb einzelner Einheiten des Bildungssystems bestehen oft extreme Unterschiede in den Leistungen, nicht aber in der Bewertung dieser Leistungen.

Fehlende Standards begünstigen

- Ungerechtigkeit in Hinblick auf Berechtigungen
- Unrealistische Einschätzungen der eigenen Fähigkeiten

These (4): Regionale Vergleiche von Schulnoten sind problematisch, weil sich offenbar unterschiedliche regionale Gewohnheiten des Beurteilens entwickelt haben.

Zwei Beispiele dazu:



Grafik 6: Notenverteilungen in der Volksschule, aufgeschlüsselt nach Bezirken (Regionen)

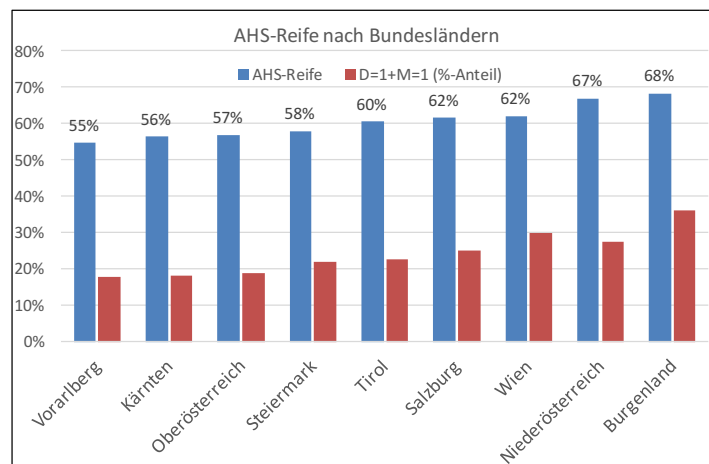
Anmerkung: S = Salzburg Stadt SL = Salzburg Land HA = Hallein JO = St. Johann TA = Tamsweg ZE = Zell am See. Quelle: Statistik des Landesschulrats für Salzburg, Schuljahr 2010/11

Im Bundesland Salzburg verhält sich die Anzahl der „Sehr gut“ in der VS fast genau direkt proportional zur Nähe eines Bezirkes zur Landeshauptstadt. Es wäre wohl schwer zu argumentieren, dass die Kinder in der Landeshauptstadt um so viel leistungsfähiger sind als

jene in Zell am See, auch wenn man berücksichtigt, dass das Bildungsniveau der Eltern in der Stadt etwas höher ist.

Salzburg ist kein Einzelfall.

Werfen wir einen Blick auf die Verteilung der sog. AHS-Reife nach Bundesländern. AHS-Reif ist, wer in Deutsch und Mathematik keine schlechtere Note als Gut hat. Grafik 7 zeigt, aufgeschlüsselt nach Bundesländern, den Prozentanteil am Schülerjahrgang 2012, für den diese Bedingung zutrifft.



Grafik 7: Verteilung der AHS-Reife nach Bundesländern. *Quelle und Berechnung: BIFIE (B4 2010)*

Es gibt klare Hinweise, dass diese Quoten nach Bundesländern jedenfalls nicht gleich sind. Im Burgenland sind um ca. 13%-Punkte mehr Schüler/innen AHS-Reif als in Vorarlberg; dazu kommt noch, dass hier 36% die Traumkombination – „Sehr gut“ in beiden Fächern – aufweisen, gegenüber 18 %, also genau der Hälfte, in Vorarlberg.

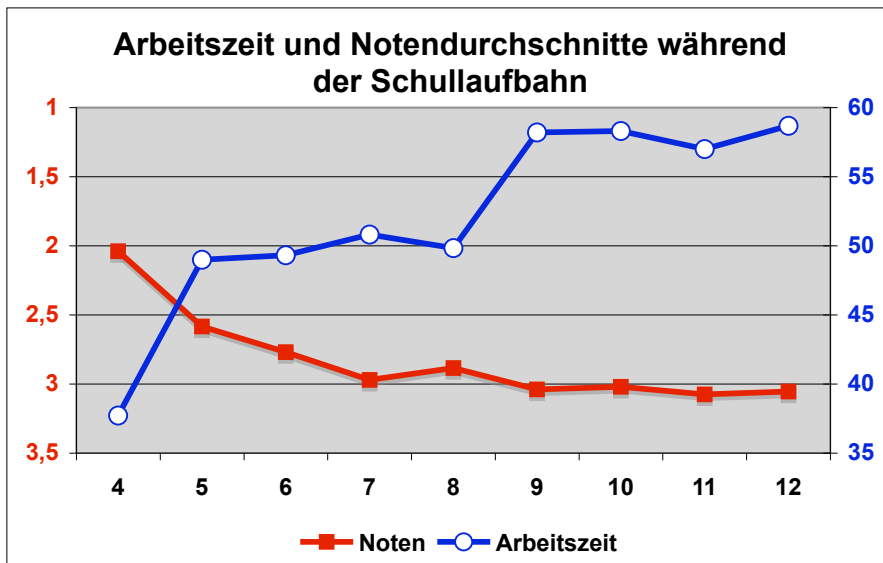
Der Anstieg der AHS-Reife und der extrem guten Beurteilungen geschieht in den Landeshauptstädten und in Wien – wohl nicht zuletzt eine Folge davon, dass im Kampf um die nicht ausreichenden Gymnasialplätze verstärkt Druck auf die Lehrpersonen ausgeübt wird, gute Noten zu vergeben.

Im historischen Vergleich lässt sich übrigens zeigen, dass die Quoten der Gymnasialreife in den letzten zwei Jahrzehnten deutlich angestiegen sind, und zwar fast ausschließlich in den Städten.

These (5): Noten bedeuten auf verschiedenen Ebenen des Schulsystems offenbar Unterschiedliches

Beispiel: Längsschnitte durch das System, Arbeitszeit

Es gibt es vereinzelt Datensätze, die zumindest im Pseudo-Längsschnitt größere Teile des Schulsystems abdecken. Aus ihnen lässt sich ablesen, dass sich das Niveau und die Verteilungen der Noten während der Schullaufbahn deutlich ändern.



Grafik 8: Arbeitszeit und Noten im Schulstufenverlauf. Eingetragen ist der Notendurchschnitt aus Deutsch, Mathematik und Englisch (in der Volksschule Deutsch, Mathematik und Sachunterricht). Quelle: Daten Befindungsuntersuchung 2005.

Der Übergang von der Volksschule in die Sekundarstufe 1 bedeutet im Mittelwert einen Absturz in den Notendurchschnitten, der sich bis an das Ende der Sekundarstufe 1 fortsetzt. Zu diesem Zeitpunkt steigen die Schüler, die in eine duale Ausbildung eintreten, aus dem Vollzeitschulsystem aus. Während der Sekundarstufe 2 bleiben dann die Notendurchschnitte auf einem niedrigen Niveau einigermaßen konstant. Im Gegensatz dazu steigt die Arbeitszeit für die Schule sprunghaft an, sodass ein eklatantes Missverhältnis zwischen Arbeitsaufwand und Arbeitsertrag entsteht, sofern man die Noten als einen Arbeitsertrag betrachtet. Unter motivationalen Gesichtspunkten ist dies mit Sicherheit keine effiziente Konstellation, und es zeigt sich immer wieder, dass die Lern- und Bildungsmotivation in der Schule einen ähnlich sinkenden Verlauf aufweist wie die Noten.

2.2 „Notenwahrheit“ im Ergebnis?

Ich komme zur zweiten Perspektive: Was sagen uns Noten über das fachliche Wissen und Können der Schülerinnen und Schüler?

Hier scheint unbestritten, dass Noten etwas über das Leistungsniveau eines Kindes in der Klasse aussagen. Schüler/innen mit „Sehr gut“ gehören mit einiger Sicherheit zu den leistungsstärksten in ihrer Klasse, Schüler/innen mit Genügend oder Nichtgenügend zumindest mit einiger Wahrscheinlichkeit zu den leistungsschwachen in der Klasse. Man sollte dabei aber immer im Auge haben, dass dauerhaft gute Leistungen kein Zufall sind, sondern tatsächlich Ausdruck von Leistungsfähigkeit. Schlechte Leistungen können hingegen viele Ursachen haben, und man kann daher nie sicher sein, ob die erbrachte schlechte Leistung tatsächlich die „wahre“ Leistung ist, oder eine durch Einflussfaktoren beeinträchtigte.

Die intuitive Annahme, Noten würden eine einigermaßen zuverlässige Auskunft über das tatsächliche Wissen und Können eines Schülers oder einer Schülerin geben, kann an den Daten der Bildungsstandards überprüft werden.

Bei den in der 4. und in der 8. Schulstufe erfolgenden Messungen werden die durch die Standards geforderten Kompetenzen der Schülerinnen und Schüler in einem umfangreichen Verfahren getestet. Obwohl auch hier Messfehler und Ungenauigkeiten einzurechnen sind, kann man davon ausgehen, dass es sich um eine valide Abbildung dessen handelt, was die Schüler/innen tatsächlich können.

- **Einschränkungen:**
 - Bildungsstandards sind umfassende Einmal-Messungen und daher auch durch die Tagesform beeinflusst.
 - Noten basieren auf längeren Messzeiträumen, sind aber andererseits potentiell durch vielfältige personale und regionale Einflüsse verzerrt
- **Gemeinsamkeiten:** Die Leistungsmessungen beziehen sich auf die gleichen Lehrplaninhalte bzw. darauf basierenden Bildungsstandards.

Zur Beschreibung des erreichten Kompetenzniveaus werden vier Kategorien verwendet:

- Standards übertroffen
- Standards erreicht
- Standards teilweise erreicht
- Standards nicht erreicht

Da zugleich auch die Noten der Schüler/innen in den getesteten Fächern vorliegen, können wir überprüfen, wie aussagekräftig diese in Bezug auf die Kompetenz sind. Dazu können zwei komplementäre Fragen gestellt werden:

Überprüfungsfragen

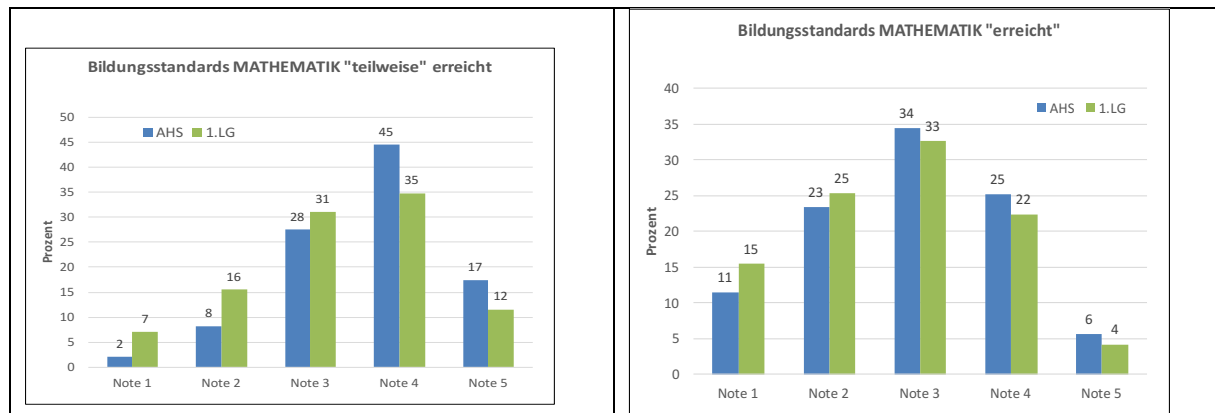
(1) Wenn ein Schüler/eine Schülerin ein bestimmtes Leistungsniveau bei den Bildungsstandards erreicht hat, können wir dann Rückschlüsse auf seine Note in diesem Fach ziehen?

bzw. komplementär dazu:

(2) Wenn ein Schüler/eine Schülerin eine bestimmte Note in einem Standards-überprüften Fach hat, können wir dann Rückschlüsse auf das Kompetenzniveau bei den Bildungsstandards ziehen?

Die Fragen sollen an den Daten der Bildungsstandards für MATHEMATIK 2012 beantwortet werden. Ein großer Teil der Schülerinnen und Schüler besuchten hier noch die Hauptschule. Kriterium ist das in Mathematik erreichte Gesamtergebnis.

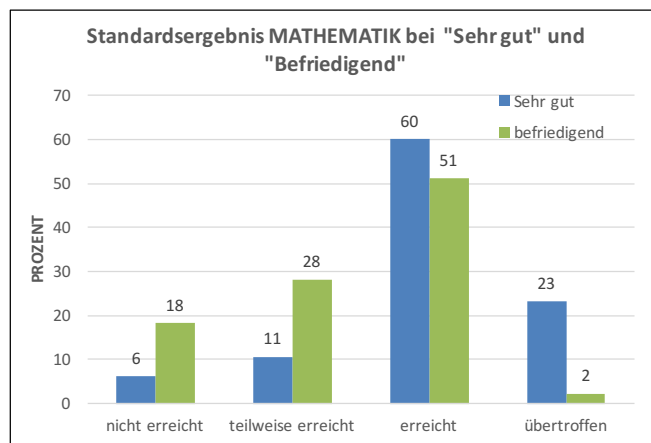
Welche Note würden wir erwarten, wenn eine Schülerin oder ein Schüler die Bildungsstandards für Lesen nur teilweise erreicht hat?
 Grafik 9 gibt uns die Antwort: Alles ist möglich.



Grafik 9: Ergebnis der Bildungsstandards, differenziert nach Noten. Quelle und Berechnung: BIFIE (BIST-Ü M8 2012)

In beiden Fällen wird die Notenskala voll ausgenutzt, wobei für beide Schulformen beide Male die Note 3 die höchste Ausprägung hat.
 Der Zusammenhang zwischen der Kompetenz in Mathematik und der Fachnote in Mathematik erscheint tendenziell eher lose. Eine Schätzung der Fachnote aus den erreichten Bildungsstandards wäre sehr unsicher.

Die Darstellung zur zweiten Frage – welche Kompetenzen haben Schüler/innen, die eine bestimmte Note erhalten haben – bringt natürlich analoge Ergebnisse.



Grafik 9: Standardergebnisse für die Note „befriedigend“. Quelle und Berechnung: BIFIE (BIST-Ü M8 2012)

Sowohl bei „Sehr gut“ als auch bei „Befriedigend“ wird die gesamte Kompetenzskala ausgenutzt. Zwar zeigen sich deutliche Unterschiede zwischen den beiden Noten, aber die Schätzung der erreichten Kompetenz aus einer vorliegenden Note wäre recht unsicher. Bei „Sehr gut“ beträgt die Basisrate, dass die Standards zumindest erreicht wurden, in Mathematik 83%, Bei „befriedigend“, das ja schulrechtlich bedeutet, dass ein Schüler die wesentlichen Inhalte des Unterrichts zur Gänze beherrscht, sind die entsprechenden Raten 53% bzw. 51%.

Die berichteten Analysen zeigen, dass es um die „Notenwahrheit“ tatsächlich sehr schlecht bestellt ist. Noten sind einigermaßen in der Lage, die relative Position von Schüler/innen in ihrer Klasse abzubilden. Wenn jedoch Noten, die in verschiedenen Kontexten entstanden sind, miteinander in Beziehung gesetzt werden, werden auf allen Ebenen Inkonsistenzen sichtbar, die nachweisen, dass gleiche Noten für ungleiche Leistungen oder ungleiche Noten für gleiche Leistungen vergeben werden.

Wie entstehen die großen Unterschiede in der Benotung?

Wie kommt es, dass Schülerinnen und Schüler, die von ihren Leistungen her vielleicht sehr ähnlich sind, so unterschiedliche Noten erhalten? Die wichtigsten Gründe sind

Unterschiedliche Zielkriterien und Maßstäbe der einzelnen Lehrpersonen	Lehrpersonen interpretieren Lehrplan und vorgegebene Bildungsstandards für ihre jeweilige Klasse, sodass unterschiedliche Anforderungen entstehen, die unterschiedlich gut erfüllt werden können
Fehlende Vergleichsmöglichkeiten	Schulklassen bilden einen abgeschlossenen Raum. Über die Vorgänge in diesem Raum wird in der Regel nicht oder nur wenig mit anderen Lehrpersonen kommuniziert. Austausch von Materialien, Prüfungsinstrumenten findet nur selten statt ????
Unterschiedliche und problematische Messinstrumente und Messverfahren	Die Art der Prüfung oder die Qualität von Schularbeiten beeinflusst das Ergebnis. ZB wird die vom Gesetz geforderte „Eigenständigkeit“ des Arbeitens für die Note SG von manchen Lehrpersonen in Prüfungsverfahren umgesetzt, von anderen nicht. Die ständige Beurteilung der laufenden Mitarbeit ist ein sehr unzuverlässiges Verfahren.
Verzerrungen durch Erwartungseffekte (Halo-Effekt, Priming-Effekte)	Lehrpersonen entwickeln Erwartungen, welche Leistungen die einzelnen Schülerinnen und Schüler voraussichtlich erbringen, und interpretieren die tatsächlichen Leistungen im Sinne dieser Erwartungen. In den klassischen Untersuchungen von Rudolf Weiss reichte der Zusatz „Schüler aus gutem Hause“ bei einem Aufsatz, um seine Bewertung durch eine große Gruppe von Lehrpersonen um etwa 1 Notenstufe zu steigern, gegenüber dem gleichen Aufsatz, der die Bemerkung „Sohn eines Journalisten“ trug. Solche Einflussfaktoren sind insbesondere das äußere Erscheinungsbild und Auftreten von Schülern und Schülerinnen, ihr familiärer Hintergrund, oder auch das Geschlecht. -> Genderergebnisse Grafik
Äußerer Druck	Lehrpersonen werden manchmal direkt oder indirekt bedroht, einem Kind jene bessere Note zugeben, die für den Zugang zu einem bestimmten Schultyp notwendig ist
Absichtliche Verzerrungen durch strategische Überlegungen	Lehrpersonen geben sowohl bessere als auch schlechtere Noten als es den tatsächlichen Leistungen entsprechen würde, in der Erwartung, die Diskrepanz würde Schüler/innen zum Lernen motivieren. Sie verschenken aber auch Noten, „um einem Kind in seiner Schullaufbahn nicht im Wege zu stehen“.

Die Bedeutung der aufgezeigten Ergebnisse bzw. der mit ihnen verbundenen Einflüsse ergibt sich, wenn wir sie mit den wichtigsten Aufgaben des Bildungssystems in Beziehung setzen.

Systemziele

- **Leistungsförderung:** ein möglichst hohes mittleres Niveau an Wissen, Kenntnissen, Kompetenzen und Qualifikationen bei den SchülerInnen (und letztlich in der Gesamtbevölkerung) zu erreichen und dabei gleichzeitig die Leistungsunterschiede zwischen unterschiedlichen Schülergruppen möglichst gering zu halten, d.h. das Entstehen von „Scheren“ der Kompetenzentwicklung zu vermeiden (**Chancenausgleich**);
- **Integration:** Jede Form des sozialen Ausschlusses beim Zugang zu Bildung zu vermeiden, d.h. den Anteil derjenigen Risikogruppen unter den SchülerInnen, deren Partizipationschancen am lebenslangen Lernprozess durch unzureichende Kompetenzentwicklung eingeschränkt werden, zu minimieren.
- **Gerechtigkeit:** Die Vergabe von Berechtigungen an Schnittstellen des Schulwesens an objektiven Leistungsstandards zu orientieren.

Auf individueller Ebene

- Voraussetzungen für die individuelle Entwicklung der Persönlichkeit zu schaffen;
- Bewusstsein und Kompetenzen der (zukünftigen) Bürger/innen für die Partizipation in den verschiedenen Gemeinschaftsstrukturen zu entwickeln.
- Qualifikationsvoraussetzungen für einen aktiven Beitrag am Arbeitsmarkt und damit letztlich für die wirtschaftliche Entwicklung zu gewährleisten, und

Im Nachfolgenden wird die Ziffernnote im Hinblick auf drei dieser Ziele geprüft:

3 Drei Fragen

3.1 Fördern Noten das Lernen?

Fördern Noten das Lernen?

Noten geben Feedback und sind dadurch eine wichtige Komponente des Lernens

- Allerdings: Ziffernnoten werden in der Regel erst dann gegeben, wenn der Lernprozess bereits abgeschlossen ist

Noten haben eine motivierende Funktion

- Allerdings: Paradoxon der gleichen Wirkung gegensätzlicher Maßnahmen

Noten sind ein Feedback für Schüler/innen und Lehrer/innen; sie geben Rückmeldung darüber, was gekonnt wird oder was im Unterricht erreicht wurde, und bilden damit eine wichtige Komponente jeden Lernprozesses. Diese Funktion wird allerdings dadurch eingeschränkt, dass Noten in der Regel dann gegeben werden, wenn der Lernprozess bereits abgeschlossen ist, sodass sie nicht mehr unmittelbar wirksam werden, und dass die Rückmeldung eher global-bewertend, nicht sachlich-informativ ausgerichtet ist.

Dazu kommt die potentiell motivierende Funktion von Noten. Ausschlaggebend sind hier eher Emotionen: Freude und Stolz über den hinter den Noten stehenden Erfolg, Freude über das darin enthaltene Lob bei guten Noten, Enttäuschung, Angst, Wut wenn der Erfolg nicht so gut ist oder wenn man sich ungerecht behandelt fühlt.

In der Praxis tritt hier das Paradoxon der gleichen Wirkung gegensätzlicher Maßnahmen auf: Wie bereits angesprochen, geben Lehrpersonen sowohl bessere als auch schlechtere Noten als es den tatsächlichen Leistungen entsprechen würde, in der Erwartung, die Diskrepanz würde Schüler/innen zum Lernen motivieren.

Die Antwort auf die Frage der Lernförderung ist daher: Ziffernnoten haben Einfluss auf die Lernmotivation, haben aber wenig Effekt auf den unmittelbaren Prozess des Lernens.



„Korrumpierende“ Effekte von Noten:

Rupert Vierlinger:

„Wenn der Schüler nach der Note schießt, ist das Lernen bereits beeinträchtigt“

Eine Schülerin:

„Das bin ich schon geprüft worden, das brauche ich nicht mehr zu können“

Allerdings: Wenn der Schüler nach der Note schießt, ist das Lernen bereits beeinträchtigt – so hat Rupert Vierlinger das Problem angesprochen. Es geht nicht mehr darum, etwas zu können, sondern eine bestimmte Zertifizierung zu erhalten.

„Das bin ich schon geprüft worden, das brauche ich nicht mehr zu können“ schreibt eine Schülerin empört auf die Frage, was sie an der Schule besonders gestört habe.

3.2 Tragen Noten zur Gerechtigkeit bei?

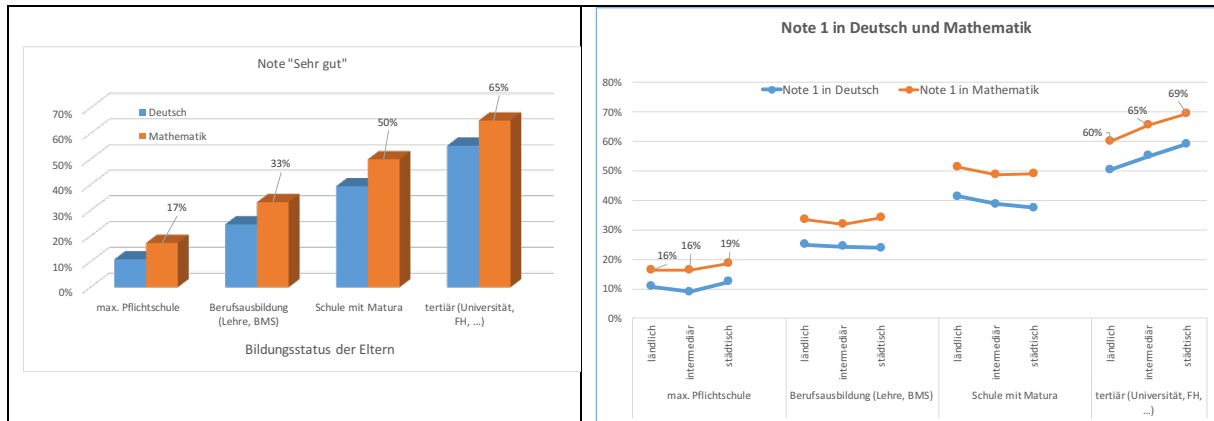
Selektion und Allokation bezeichnet den Sachverhalt, dass auf Basis der Noten Schülerinnen und Schüler für unterschiedliche Laufbahnen selektiert und damit für bestimmte Berufe in Position gebracht werden. In diesem Kontext stellt sich vor allem die Frage, in welchem Ausmaß diese Selektion tatsächlich auf Basis von erbrachten Leistungen erfolgt bzw. in welchem Ausmaß hier andere Aspekte eine Rolle spielen.

Wolfgang Klafki unterschied in seiner Analyse des Leistungsprinzips mit Bezug auf Klaus Offe zwischen einer aristokratischen (wir würden eher sagen: ständischen) und einer meritokratischen Gesellschaft. Ständische Gesellschaften vergeben soziale Positionen und Privilegien nach Merkmalen von Geburt, Herkunft und Familie. Meritokratische Gesellschaften verteilen sozialen Aufstieg und privilegierte Positionen nach der individuellen Leistung.

Aufgrund vieler Analysen lässt sich die These formulieren:

Noten, die von den einzelnen Lehrpersonen vergeben werden, bilden das Werkzeug, durch das ständische Aspekte wirksam werden.

Beispiel Übergang AHS:



Grafik 10: VS-Noten „Sehr gut“ in D und M, in Abhängigkeit von Bildungsvoraussetzungen der Eltern und regionalen Voraussetzungen. Quelle und Berechnung: BIFIE (BIST-B4 2010)

Selektion und Allokation erfolgen, das sollen die Darstellungen belegen, nicht nur nach Leistung, sondern es dominieren familiäre Aspekte, die zum Teil noch durch regionale Einflüsse verstärkt werden.

3.3 Fördern Noten die Persönlichkeitsentwicklung?

Fördern Noten die Persönlichkeitsentwicklung?



Analyse zu zwei Merkmalen der Persönlichkeit:

Leistungselbstkonzept

das Ausmaß, in dem sich Menschen selbst für leistungsfähig und tüchtig halten und Zutrauen in ihre Fähigkeiten haben

Allgemeines Selbstwertgefühl

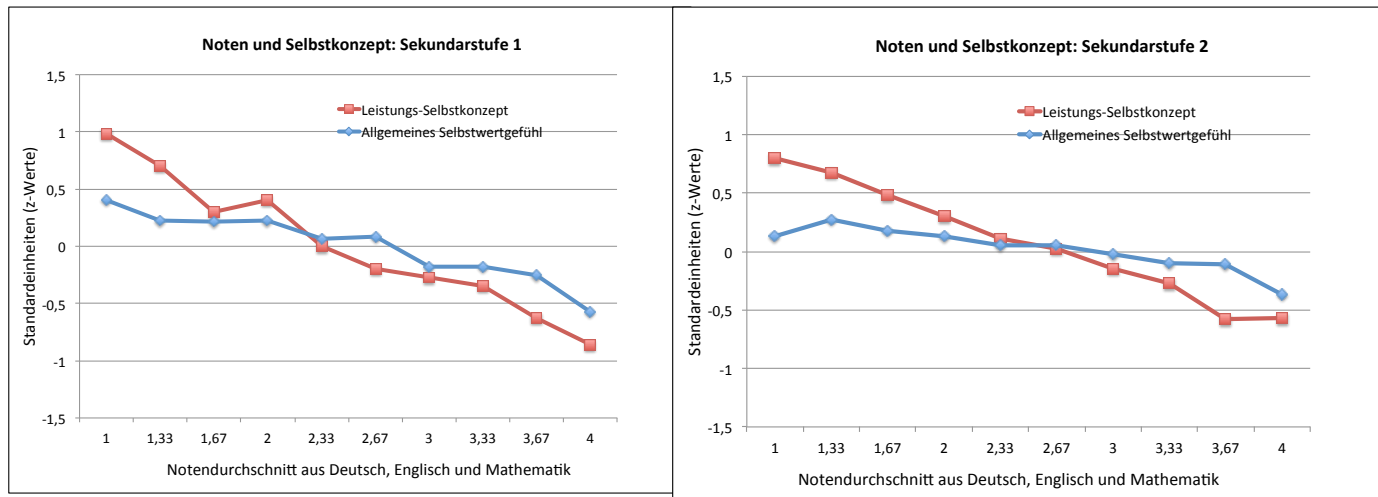
Ausmaß, in dem sich Menschen ganz allgemein für wertvoll und wichtig halten und eine gute Meinung von sich selbst haben

Langfristig wirken Noten auf das Selbstbild / Selbstkonzept der Schülerinnen und Schüler.

Die folgenden Grafiken zeigen den Zusammenhang zwischen zwei zentralen Merkmalen des **Selbstkonzepts** – Leistungselbstkonzept und Allgemeines Selbstwertgefühl – und dem Durchschnittswert aus den Noten aus Deutsch, Englisch und Mathematik.

Das Leistungselbstkonzept bezieht sich auf das Ausmaß, in dem sich Menschen selbst für leistungsfähig und tüchtig halten und Zutrauen in ihre Fähigkeiten haben. Das Selbstwertgefühl beschreibt das Ausmaß, in dem sich Menschen ganz allgemein für wertvoll und wichtig halten und eine gute Meinung von sich selbst haben.

Die Daten stammen aus der 2007 durchgeführten Untersuchung zum Befinden von Kindern und Jugendlichen in der Schule.



Grafik 11: Noten und Selbstkonzeptmerkmale. Quelle und Berechnung: Eder, Befindensuntersuchung 2007

Die Verläufe machen sichtbar, wie sensibel vor allem das Leistungsselbstkonzept schon auf kleinste Veränderungen im Notendurchschnitt reagiert. Fast durchgehend ist schon eine Verschlechterung um eine Drittelnote bereits mit einer Veränderung in der Zusammenhangslinie verbunden. Die Differenz zwischen Schülern mit lauter „Sehr gut“ und solchen mit lauter „Genügend“ beträgt in der Sekundarstufe 1 fast zwei Standardeinheiten, in der Sekundarstufe 2 mehr als 1 Standardabweichung.

Ähnliche Verläufe, allerdings nicht so ausgeprägt, zeigen sich für das Allgemeine Selbstwertgefühl. Auch hier zeigt sich, dass der Notendurchschnitt offenbar ein guter Prädiktor dafür ist, in welchem Ausmaß sich junge Menschen selber Wert zuschreiben.

Die hier beispielhaft gezeigten Zusammenhänge finden sich auch für Einstellungen zur Schule oder für die Lernmotivation.



Die folgenreichste Auswirkung liegt jedoch in der selektierenden Wirkung von Noten: **Schüler halten sich dort für fähig und gut, wo ihre Noten gut sind, und dort für schlecht, wo ihren Noten schlecht sind.** Vor allem bei der Entwicklung von Interessen werden in einer bestimmten Phase jene Bereiche als Interessengegenstände ausgeklammert, wo die Noten das Fehlen von Fähigkeiten suggerieren. Bei der Entwicklung von beruflichen Interessen, die ja im Schulalter stattfindet, werden zunehmend jene Bereiche ausgeklammert, in denen man – auf Basis der Notenrückmeldungen – glaubt, keine Fähigkeiten zu besitzen, und es bleiben jene Bereiche über, die nicht durch schlechte Noten belastet sind. Dies ist einer der Gründe, warum Mädchen wenig Interesse für die Naturwissenschaften und Burschen weniger Interessen für Sprache und Kunst entwickeln.

4 Zusammenschau und Implikationen

1. Notenwahrheit ist nicht erreichbar. -> Ende der teacher-made assessments
Noten, die tatsächlich in einem hohen Ausmaß tatsächliche Leistungen vergleichbar wiedergeben sollen, sind nur in einem standardisierten Verfahren möglich. Das bedeutet: zentrale und für alle die gleichen Aufgabenstellungen, klare Durchführungsregelungen, eindeutige Vorgaben für die Auswertung. (bei kriterienorientierter Vorgangsweise, wenn Noten etwas über die tatsächliche Leistung aussagen sollen)
Die Herstellung von Notenwahrheit, wenn sie ernstgenommen werden sollte, impliziert das Ende der Beurteilung seitens der Lehrerinnen und Lehrer.

2. Die derzeit vergebenen Noten, von der Volksschule angefangen, weisen – neben den bekannten und vielfach nachgewiesenen Einflüssen auf das Lehrerurteil - massive Unterschiede zwischen einzelnen Schulstandorten sowie regionale Verzerrungen auf

3. Noten weisen einen starken Zusammenhang mit dem Bildungsstatus der Eltern auf.
Ein solcher Zusammenhang ist einerseits teilweise dadurch erklärt, dass Eltern mit höherem Bildungsstatus auch besser in der Lage sind, ihre Kinder zu fördern. Es kann aber gezeigt werden, dass es zusätzliche Verzerrungen in der Richtung gibt, dass Kinder von hochgebildeten Eltern vor allem in den Städten noch einen zusätzlichen Bonus erhalten, der dazu führt, dass sie beispielsweise in der Volksschule jene Notenkombinationen erhalten, die zum Eintritt in die AHS berechtigen.

4. (Schlechte) Noten haben längerfristig einen überwiegend negativen Einfluss auf die Entwicklung des Selbstbildes, der Motivationen und der beruflichen Interessen der Schülerinnen und Schüler und bestimmen so die längerfristige Entwicklung

Implikationen

Noten setzen sich im Großen aus zwei Komponenten zusammen: Rückmeldung und Bewertung. Dem entspricht auch die Terminologie in den einschlägigen Gesetzen, die zwischen Leistungsfeststellung und Leistungsbeurteilung differenzieren.

Noten umfassen	
Sachliche Rückmeldung	Bewertung
Essentiell für jeden Lernprozess Gibt die Möglichkeit der Weiterentwicklung	„Gut für die Guten, schlecht für die Schlechten“ („Matthäus-Effekt“)

Diese beiden Komponenten haben für das Lernen eine unterschiedliche Funktion: Die Feststellung und Rückmeldung von Lernergebnissen sind für das Lernen in der Schule essentiell. Ohne Rückmeldungen ist erfolgreiches Lernen nicht möglich. Noten haben allerdings den Nachteil der fehlenden Sachinformation.

Die Bewertungskomponente ist ambivalent.

Implikationen

- Noten in der bisherigen Form sind ein Instrument aus einem ständisch-selektiv orientierten Bildungssystem und in dieser Form historisch überholt.
- Ihre sachliche Aussage über das tatsächliche Können oder die Leistungen der Schülerinnen und Schüler ist gering, „Notenwahrheit“ ein nicht erfüllbarer Anspruch.
- Ihre Verwendung als Basis für schulische Laufbahnentscheidungen ist kontraproduktiv und begünstigt Bildungsarmut bei Kindern aus Elternhäusern mit niedrigem Bildungs- oder Sozialstatus.
- Schlechte Noten führen zu einer frühen Ausgrenzung von Inhaltsbereichen aus dem Spektrum möglicher beruflicher Interessen.

5 Was sind die Alternativen?

Aus dem Streifzug durch die Forschungsergebnisse lässt sich resümieren, dass Ziffernnoten nicht in der Lage sind, das Lernen effizient zu fördern, und ihre sonstigen Funktionen zunehmend obsolet geworden sind.

Beurteilungssysteme, die besser geeignet sind, das Lernen zu unterstützen, müssten auf jeden Fall die folgenden Merkmale aufweisen:

- Es müssen Formen der direkten und unmittelbaren Leistungsrückmeldung entwickelt und ausgebaut werden, die das Lernen der Kinder und Jugendlichen direkt unterstützen.
- Sie sollen den Leistungsstand bzw. das Können der Schülerinnen und Schüler sachlich wiedergeben
- Sie sollen zum weiteren inhaltlichen Lernen animieren
- Sie sollen die Persönlichkeit stärken bzw. die Persönlichkeitsentwicklung unterstützen
- Sie sollen nach Möglichkeit nicht-diskriminierend sein

Die sich derzeit abzeichnenden Möglichkeiten wie

- nach Ergänzung oder Ersatz der Ziffernnoten durch verbale Beurteilung,
- nach Ersatz der lehrervergebenen Noten durch externe Beurteilungen oder
- nach Ersatz der Ziffernbeurteilung durch Rückmeldeformen, die unmittelbar bei den Kompetenzen der Schülerinnen und Schüler ansetzen (Kompetenzlisten, Lernziellisten, Rubrics).
- sind m.E. wichtige Ansätze, aber noch nicht durchgehend so weit entwickelt, dass sie den oben genannten Ansprüchen genügen. Ideen für die Weiterentwicklung der Leistungsbeurteilung finden sich zurzeit eher außerhalb der Schule, z.B. im europäischen Referenzrahmen für Sprachen oder im Leistungssport (z.B. Judo).