

Der Umgang mit Noten ist in der deutschen Bildungslandschaft geprägt von großem Traditionsbewusstsein - Schule ohne Noten ist uns schlicht fremd. Zudem existiert eine stabile - nicht hinterfragte - Überzeugung, dass Noten aussagekräftig, überwiegend leistungsindizierend und -motivierend sowie orientierungsgebend sind. Folglich macht man sie zur Basis für sehr zentrale Entscheidungen: Versetzungen am Schuljahrsende, Übergänge auf anderen Schularten, Vergabe von Studienplätzen und Lehrstellen. Damit einher gehen gravierende Chancenungleichheiten für berufliche Karrieren, Sozialprestige und Einkünfte. Andere Formen der Leistungserhebung werden häufig pauschal als „Kuschelpädagogik“ abgetan - Bildung kann nicht entstehen, wenn Lernen nicht weh tut, so die offensichtliche Überzeugung in unserer Gesellschaft.

Dass solch ein Umgang mit schulischen Noten in hohem Maße naiv ist, zeigen seit den 70er Jahren zahlreiche Forschungen zur Entstehung und Wirkung von Noten. Allerdings werden sie bei uns überwiegend nicht zur Kenntnis genommen.

Dass sowohl ein gesundes Mißtrauen gegenüber unserer Notenpraxis angebracht ist, als auch alternative Verfahren ins Auge gefasst werden sollten - wenn es wirklich um die Erfassung und Förderung von Leistung und nicht um soziale Sortierprozesse geht - soll in den folgenden Abschnitten verdeutlicht werden.

A) **Die Unschärfe der Noten bei der Erfassung von Leistung**

Es scheint ja so einfach und nachvollziehbar zu sein: Es gibt ein im Unterricht behandeltes Unterrichtsthema, über dessen Inhalte nach Abschluss eine von der Lehrperson zusammengestellte „Klassenarbeit“ geschrieben wird und von ihr „korrigiert“ wird. Die dabei erzielten Punkte werden über eine lineare Skala in Notenwerte umgerechnet. Aus dem erreichten Notenwert wird dann pro Schüler / Schülerin abgeleitet, wie erfolgreich / leistungsstark / begabt sie / er ist. Dies wiederum öffnet oder verschließt weitere Wege.

Bei genauerer Betrachtung wird sehr schnell deutlich, dass es so einfach nicht ist:

- *Leistungsanspruch:*

Bei der Zusammenstellung der Klassenarbeit hat die Lehrperson die freie Entscheidung, wieviele Fragestellungen / Aufgaben auf welchen Schwierigkeitsstufen in der Klassenarbeit bewältigt werden sollen. Dabei sind auf das Thema bezogen ca. fünf Schwierigkeitsstufen (z. B. nach der Bloom'schen Taxonomie) möglich. Je nach Lehrperson werden hier deutlich unterschiedliche Anforderungen im Rahmen einer Klassenarbeit gestellt, so dass schon vom Schwierigkeitsgrad der Arbeiten her ein Vergleich praktisch nicht möglich ist. Nun könnte man sich zumindest innerhalb der Parallelklassen einer Schule auf gemeinsame Arbeiten einigen - was aber im Regelbetrieb eher selten geschieht - zwischen verschiedenen Schulen ist es kaum durchführbar.

- *Personeneffekte:*

Schon beim durchschnittlichen Schwierigkeitsgrad einer Klassenarbeit spielt die Person des Lehrers / der Lehrerin eine zentrale Rolle - beim Korrekturverfahren gilt dies ebenso uneingeschränkt. Es gibt hier die Ausprägungen der sog. „Hochbewerter“ und „Tiefbewerter“, d. h. Lehrpersonen, die bei der Punktevergabe „von unten“ anfangen und Punkte addieren oder eben „von oben“ anfangen und Punkte abziehen bei fehlenden oder falschen Angaben, was zu durchaus unterschiedlichen

Gesamtpunkten führen kann. Schließlich gibt es diejenigen, die eine „Tendenz-zur-Mitte“ pflegen und in ihrem Korrekturstil Extreme nach unten oder nach oben grundsätzlich vermeiden.

- *Wahrnehmungseffekte*

Jede Lehrperson unterliegt aber auch den - ganz menschlichen - psychologisch bedingten Wahrnehmungsbesonderheiten, die sich individuell in unterschiedlichen Ausprägungen zeigen können.

So lässt der sog. „*Halo-Effekt*“ andere Persönlichkeitsmerkmale der Schülerinnen und Schüler mit in die Korrektur der Lehrperson einfließen, was - trotz identischer - Arbeitsergebnisse zu deutlich unterschiedlicher Bewertung führt.

Der „*Reinigungseffekt*“ wiederum steuert die Wahrnehmung des Korrigierenden hin auf eine in der „Schulkultur“ verankerte ideale Streuung der Ergebnisse. Bei durchweg positiven Ergebnissen gilt die Arbeit als zu leicht, bei schlechten Ergebnissen als zu schwer. Entsprechend entdeckt der / die Korrigierende bei vielen „guten“ Ergebnissen zunehmend Fehler oder Ungenauigkeiten, die zu negativeren Bewertungen führen, bei „schlechten“ Ergebnissen nimmt die Neigung zu, Fehler oder Ungenauigkeiten zu übersehen, was zu besseren Ergebnissen führt. All dies geschieht unbewusst, ist also in der Regel keine willentliche Manipulation der Lehrperson, sie unterliegt aber eben den Mechanismen ihrer Wahrnehmung - mit Auswirkungen auf die Bewertung.

Beim „*Vorerwartungseffekt*“ kann es passieren, dass eine negative Abweichung von sehr hohen Erwartungen im Hinblick auf die erhobenen Leistungen eines bestimmten Schülers / einer bestimmten Schülerin zu einer deutlich negativeren Bewertung führt (Enttäuschung) oder eine implizite Entschuldigung (eigentlich kann er / sie das doch) die Bewertung abmildert. Auch bei abweichenden deutlich besseren Leistungen als erwartet sind vergleichbare Effekte denkbar.

Schließlich wurde auch festgestellt, dass Bewertungen in hohem Maße an den Zeitpunkt ihrer Entstehung gebunden sind: Wird dieselbe Arbeit durch denselben Korrektor zu einem anderen Zeitpunkt wiederholt korrigiert, ändert sich nicht selten die Bewertung.

- *Referenzgruppeneffekt:*

Im „*Referenzgruppeneffekt*“ schließlich wird die Leistung des Einzelnen auf die Gesamtleistung der Gruppe / Klasse bezogen. Dabei gilt generell: Identische Leistungen werden auf dem Hintergrund einer hohen durchschnittlichen Leistungskultur der Gruppe / Klasse schlechter bewertet, auf dem Hintergrund einer niedrigen durchschnittlichen Leistungskultur der Gruppe / Klasse besser bewertet. Die Bewertung der Einzelleistung hängt daher in hohem Maße davon ab, in welchem Umfeld die Leistung erbracht wird.

- *Pädagogische Effekte:*

Schließlich gibt es noch eine Reihe pädagogischer Effekte, die überwiegend um zwei Kristallisationspunkte herum entstehen: Disziplinierung und Prognose. Im „*Disziplinierungseffekt*“ wirkt sich die Verhaltensdimension negativ auf die Bewertung einer Leistung aus, beim „*Gratifizierungseffekt*“ erfolgt das Gegenteil. Und schließlich sollen in einer Note - insbesondere dann, wenn es um Jahresergebnisse geht - nicht nur mathematische Verfahren zur Anwendung kommen, sondern es soll auch um eine - prognostisch verwertbare - Gesamtwürdigung des Schülers / der Schülerin gehen.

Mit anderen Worten: Neben einer zahlenbasierten Scheinobjektivität des Notenfestsetzungsverfahrens spielen - ganz legal - die prognostischen Überzeugungen der Lehrperson bezogen auf den einzelnen Schüler / die einzelne Schülerin eine zentrale Rolle.

Die in diesen fünf aufgeführten Feldern geschilderten Effekte führen notwendigerweise dazu, dass eine Vergleichbarkeit von Noten nur dann ansatzweise gegeben ist, wenn die Noten

- im selben Umfeld (Klasse),
- zur gleichen Zeit,
- anhand derselben Aufgaben entstanden sind und,
- durch dieselbe Person korrigiert wurden.

Aber auch dann lassen sich die Auswirkungen der psychologischen Effekte und pädagogischen Effekte nicht ausschließen.

Außerhalb dieses Entstehungsrahmens verliert eine Note jegliche Aussagekraft und ist ungeeignet für Vergleiche. Wird dies z. B. im Rahmen von Bewerbungsverfahren für Ausbildungsplätze nicht berücksichtigt, wird es in vielen Fällen dazu führen, dass die Personalchefs diejenigen mit den schlechteren schulischen Leistungen einstellen, nur weil sie die besseren Zahlenwerte in ihren Zeugnissen haben - weil diese eben auf dem Hintergrund völlig anderer Rahmenbedingungen entstanden sind. Denn die Abweichungen sind durchaus erheblich: Für identische Leistungen existiert im Alltag eine durchschnittliche Abweichung von bis zu zwei Notenstufen, in wissenschaftlich begleiteten Experimenten von bis zu fünf Notenstufen.

B) Die psychologische Wirkung von Notengebung

Ging es bisher „lediglich“ um die seit langem bekannte mangelnde Präzision der Leistungsbestimmung per Notenvergabe, so zeigt sich in den psychologischen Wirkungen von Noten eine deutlich fatalere Konsequenz, die insbesondere die Leistungsentwicklung bei Kindern und Jugendlichen nachhaltig blockiert, und die sehr deutlich vor Augen führt, wie sehr unsere Annahme, dass Noten und Leistung zusammengehören, vollkommen falsch ist.

Bereits Ende der 80iger Jahre hat Ruth Butler die Auswirkungen von Notengebung, Kommentaren (Feed-Back) sowie der Kombination von Feed-Back und Note in ihren Auswirkungen auf die Leistungsentwicklung in verschiedenen Gruppen untersucht. Das Ergebnis ist sehr deutlich: „Erhielten (...) die Schüler keine Noten, sondern nur Kommentare, dann verbesserten sie ihre Leistungen. Die Gruppe, die Noten und Kommentare erhielt, zeigte ein signifikantes Absacken der Leistungen, ähnlich wie die Gruppe, die nur Noten erhielt.“¹

Damit bestätigte sie die Forschungen im Rahmen der Sozialpsychologie zu sog. intrinsischen Motivatoren, zu denen die Note zählt. Hier wurde in zahlreichen Versuchen nachgewiesen, dass durch Belohnungen, die außerhalb des eigentlichen

¹ Butler, R.: Enhancing and undermining intrinsic motivation: the effect of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58 (1988), S. 1-14. Übersetzung bei Wellenreuther a.a.O. S. 5

Themenbereichs lagen, ein intrinsisches Interesse an der untersuchten Tätigkeit untergraben wurde, weshalb man für diesen Prozess den Begriff der „Korrumpierung“ wählte. Im Endeffekt zeigte sich stabil, dass die untersuchten Kinder an einer weiteren Beschäftigung mit dem Thema nicht mehr interessiert waren, wenn eine Belohnung von außen nicht mehr erfolgte. Lediglich diejenigen, die keine Belohnung von außen bekamen, widmeten sich weiterhin ihrer themenbezogenen Beschäftigung. Insofern war das Ergebnis von Ruth Butlers Untersuchungen zur Notengebung nicht wirklich unerwartet.

In einer weiteren Veröffentlichung - und hier ist jetzt wieder der Blick in den Binnenraum von Schule interessant - legte Ruth Butler die Mechanismen offen, die zu der gemachten Beobachtung führten.

Dabei müssen zwei Rückmeldeprozesse unterschieden werden. Kommentare (Feedback) beziehen sich in erster Linie auf die zu bewältigenden Aufgaben und geben Hinweise, wie bessere Erfolge erzielt werden können, sie sind also in hohem Maß an der erfolgreichen Bewältigung von konkreten Aufgaben interessiert (task-involving). Noten hingegen geben in erster Linie Auskunft darüber, wie erfolgreich oder weniger erfolgreich eine Person bei ihrer Aufgabenbewältigung war (ego-involving). Beides führt zu einem völlig anderen Verhalten bei weiteren Lernprozessen. „Aufgaben, Lernprozesse oder, wie ich es nennen möchte, aufgabenorientierte Zielsetzungen, bewegen Schüler dazu, sich lohnendes Wissen, Verständnis und Kompetenzen aneignen zu wollen. Im Gegensatz dazu werden Denken, Emotionen und Handeln von Schülern durch Zielvorstellungen, die sich auf Leistungen, Fähigkeiten und die eigene Person konzentrieren, in einer Weise gesteuert, dass sie überlegene Leistungen demonstrieren oder die Offenlegung von Leistungsschwäche vermeiden wollen.“² In der konkreten Beobachtung ließen sich klare Strategien unterscheiden, die zu unterschiedlichen Grundhaltungen wurden. „Grundschüler, die wiederholt eine aufgabenbezogene Rückmeldung erhielten - und zwar in Form von Kommentaren, die sowohl Hinweise zur Aufgabenbearbeitung enthielten als auch einen konkreten Hinweis darauf, was die Schüler verbessern konnten -, nahmen aufgabenorientierte Zielvorstellungen an, schenkten den Kommentaren Beachtung und nutzten sie, um zu lernen und mit der Zeit Fortschritte zu machen. Im Gegensatz dazu stimmten Schüler, die als Rückmeldung wiederholt Zensuren erhielten, leistungsorientierten Zielvorstellungen zu, leistungsfähige Schüler erhielten ihre hohe Leistung aufrecht, ohne sie jedoch zu verbessern, und die Leistung der lernschwächeren Schüler ging mit der Zeit zurück. Weiterhin war zu beobachten, dass sich Schüler, die sowohl eine Zensur als auch ein konkretes Feedback erhielten, verhielten, als hätten sie ausschließlich eine Zensur erhalten“³.

Butlers abschließendes Urteil ist eindeutig. „Je mehr Bedeutung daher dem jeweiligen Leistungsniveau durch die Bildungssysteme beigemessen wird, desto mehr bewirken sie als unvermeidliche - (...) - Folge die Erzeugung schulischer Disparitäten zwischen anfänglich leistungsstarken und leistungsschwachen Schülern. Darüber hinaus liefern die neuesten Erkenntnisse der PISA-Studie (...) übereinstimmende Hinweise darauf,

² Butler, R.: Ein zielorientiertes Modell für Feedbackverfahren im Unterricht. Unterrichtswissenschaft, 33. Jg. 2005, H2, S. 124

³ ebd. S. 132

dass hochdifferenzierte Bildungssysteme wie hier in der Bundesrepublik Deutschland Disparitäten verschärfen, indem sie eher einen Leistungsrückgang bei lernschwachen Schülern bewirken, als dass sie einen Zuwachs bei den leistungsstärkeren fördern.“⁴ Eine umfassende Vergabe schulischer Noten führt also letztlich zu Stagnation bei den Leistungsstarken, einer Leistungsminderung bei den Lernschwachen und zu größeren Unterschieden zwischen den Leistungsausprägungen als notwendig.

Mit anderen Worten: die momentan so sehr betonten Leistungsunterschiede (Heterogenität) innerhalb der Schülerschaft ist nur zu einem Teil im de facto unterschiedlichen individuellen Leistungsvermögen begründet, in einem anderen nicht unwesentlichen Teil sind es die Effekte einer zu frühen und flächendeckend durchgeführten Notengebung, die diese Unterschiede herbeiführt. Noten sind nun einmal kein neutrales Messinstrument von Leistung, sondern ein Verfahren, dessen Ergebnisse in vielfältiger Weise das weitere Handeln der Bewerteten beeinflusst und zu Folgen führt, die so nicht erwartet wurden, und letztlich unerwünscht sind. Was für die unreflektierte Vergabe von Noten gilt, lässt sich in ähnlicher Weise auch für die auf der Notengebung basierende Nicht-Versetzung feststellen. Auch hier gilt, dass unsere Annahmen, dass Nicht-Versetzung leistungssteigernd wirkt, falsch ist - im Gegenteil, in der Hattie-Studie wird darauf hingewiesen, dass keine einzige Studie existiert, die für eine Nicht-Versetzung positive Effekte nachweist. Dagegen lässt sich eindeutig belegen: „die Nicht-Versetzung hat den Ergebnissen zufolge einen negativen Effekt auf die akademische Lernleistung in den Bereichen Sprachen, Kunst, Lesen, Mathematik, Arbeits- und Studierfertigkeiten, Sozialkunde und Notendurchschnitt.“⁵ Die (langfristig) negativen Auswirkungen betreffen selbst solche Bereiche wie den Schulbesuch, denn dort fällt auf, dass „die Anwesenheit im Unterricht [...] bei Lernenden, die eine Klasse wiederholen, geringer ist.“⁶ Das Fazit der Bildungsforschung ist eindeutig: „In Bezug auf die Nicht-Versetzung ist die einzig interessante Frage, warum eine solche Praxis angesichts derart vernichtender empirischer Belege immer noch angewandt wird.“⁷

C) Die effizientere Form: An Kriterien orientiertes Feedback

In den Forschungsergebnissen von Ruth Butler wurde schon deutlich, dass ein Umgang mit Leistung, der sich stark auf die Sachebene konzentriert, indem Hinweise für eine erfolgreiche Bewältigung der Anforderungen gegeben werden, letztlich ein deutlich höheres Maß an Leistung hervorzubringen vermag, als ein permanentes, die Personen in den Vordergrund stellendes Messverfahren, das keinerlei entwicklungsorientiertes Potential hat, sondern überwiegend die Ergebnisse der gemessenen Personen hierarchisiert und zu Aussagen über „bessere“ oder „schlechtere“ Schüler führt. Dass die Ergebnisse von notenbezogenen Messverfahren

⁴ ebd. S. 140

⁵ Hattie, J.: Lernen sichtbar machen. Baltmannsweiler 2013, S. 116

⁶ ebd. S. 116

⁷ ebd S. 117

danach systematisch für den Ausgleich von Defiziten oder für die Entwicklung von Exzellenzleistungen herangezogen würden, ist nicht Teil des Bewertungsansatzes. Im Unterschied zu den in unserer Gesellschaft verbreiteten Überzeugungen führen Noten zu schlechteren, der Verzicht auf Noten hingegen zu letztendlich besseren Leistungen - wenn die Lehrkräfte die vorhandenen alternativen Verfahren konsequent und kompetent anwenden.

D) Die Position der Gemeinschaftsschulen

Der Verzicht auf Notengebung in den unteren Klassen an den Gemeinschaftsschulen hat daher nichts mit mangelnder Leistungsforderung oder „Kuschelpädagogik“ zu tun. Im Gegenteil: Bewertungen von Arbeitsergebnissen und Hinweise zur Weiterarbeit erfolgen in der Regel kontinuierlicher und mit dem Ziel einer positiven Leistungsentwicklung. Die Einbindung von Selbstbewertungsverfahren führen zudem zu einer größeren Fähigkeit, eigene Leistungen adäquat einzuschätzen. Es ist die Absicht, frühe Festlegungen und mögliche Blockaden im Hinblick auf die eigene Leistungsfähigkeit zu vermeiden. Die Gemeinschaftsschule geht daher den Weg, durch ein höheres Maß an sog. „formativer Bewertung“, d. h. durch eine Bewertung, die unmittelbar in den Lernprozess einmündet, die Leistungsentwicklung kontinuierlicher zu steuern und zu begleiten. Dazu eignen sich Notensymbole nun einmal nicht, sondern es Bedarf einer genaueren Beschreibung dessen, was geleistet wurde, inwieweit es vom eigentlichen hohen Leistungsziel noch abweicht und welcher Weg dorthin am erfolgversprechendsten ist.

Auf diesem Weg sieht sich die Gemeinschaftsschule in guter Gesellschaft mit anderen in der Bildung sehr erfolgreichen Ländern, in denen ebenfalls eine Leistungsbeschreibung durch Zahlensymbolik frühestens ab der achten Jahrgangsstufe erfolgt.