# Leistungsbeurteilung im Schulalltag: Wozu vergleicht man was womit?[1]

Falko Rheinberg

## Messen und Beurteilen

Voraussetzung für jede Leistungsbeurteilung ist zunächst, dass man irgendein Ergebnis ermitteln kann, das sich nach Menge (z. B. Zahl heute richtig geschriebener Vokabeln) und/oder Güte (z. B. Qualität der Aussprache dieser Vokabeln) näher bestimmen lässt. Dieser Punkt betrifft die Leistungs*messung*. Nun lässt sich leicht zeigen, dass die bloße Feststellung eines noch so exakt ermittelten Leistungspunktwertes für sich allein genommen noch wenig besagt. Die Mitteilung: "Sie haben bei einem Lerntest zum Inhalt des jetzigen Kapitels 28,5 Punkte erreicht" würde den Leser nicht sonderlich klüger machen. Er wüsste nämlich nicht, ob das viel oder wenig ist. Für solche Einschätzung würde er Vergleichsstandards benötigen. Diese Standards könnten nun verschieden hoch oder niedrig sein. Abhängig davon können 28,5 Punkte sehr viel oder sehr wenig, eine bessere oder schlechtere Leistung sein. Dies ist eine Frage der Leistungs*beurteilung*. Beurteilung bedeutet hier: Vergleich eines ermittelten Ergebnisses mit einem Standard.

## Eine kleine Beurteilungsaufgabe

Interessanter als die bloße Höhe dieser Vergleichsstandards ist die Frage, woher ein jeweiliger Standard stammt. Überraschenderweise gibt es nämlich qualitativ scharf unterscheidbare Quellen, aus denen man solche Standards herleiten kann. Statt dies abstrakt abzuhandeln, ist es anschaulicher, vorweg einmal die nachfolgende Übung zu machen. Sie ist als "Kleine Beurteilungsaufgabe" in vielen Untersuchungen bei Lehrern eingesetzt worden.

---

[1] erscheint in F.E. Weinert (Hrsg.). *Leistungsmessung in Schulen*. Weinheim: Belz.

## Kleine Beurteilungsaufgabe

Eine durchschnittliche Schulklasse macht in monatlichen Abständen Schulleistungstests, in denen jeweils der Unterrichtsstoff des letzten Monats abgefragt wird. In jedem Test kann man maximal 100 Punkte erreichen. Die Tests sind so aufgebaut, dass der Klassendurchschnitt bei ca. 50 Punkten liegt. Neun Schüler erreichten bei den letzten drei Tests die unten angeführten Punkte.

Ihre Aufgabe besteht darin, bei jedem der neun Schüler das letzte Testergebnis zu beurteilen. Wenn Sie das Ergebnis eines Schülers für eine gute Leistung halten, so können Sie einen bis fünf Pluspunkte (++...) geben. Halten Sie dieses Ergebnis für eine schlechte Leistung, so können Sie einen bis fünf Minuspunkte (--...) geben. Bitte geben Sie pro Ergebnis entweder nur Plus- oder nur Minuspunkte, also nicht beides gleichzeitig! Wenn sie in eine Zeile weder Plus- noch Minuszeichen schreiben, so bedeutet das, dass Sie das Ergebnis weder für eine gute noch für eine schlechte Leistung halten. Beziehen Sie sich bei Ihrer Beurteilung bitte auf eines Ihrer Unterrichtsfächer.

| | Erreichte Punkte | | | Beurteilung des letzten Testergebnisse (bitte Plus- bzw. Minuszeichen in die Kästchen schreiben) |
|---|---|---|---|---|
| | 1. Test | 2. Test | 3. (letzter) Test | |
| ① | 60 | 55 | 50 | |
| ② | 25 | 25 | 25 | |
| ③ | 85 | 80 | 75 | |
| ④ | 50 | 50 | 50 | |
| ⑤ | 65 | 70 | 75 | |
| ⑥ | 15 | 20 | 25 | |
| ⑦ | 40 | 45 | 50 | |
| ⑧ | 75 | 75 | 75 | |
| ⑨ | 35 | 30 | 25 | |

Es kann sein, dass Sie bei einigen Schülern sich unsicher über die „richtige“ Beurteilungsweise sind. Entscheiden Sie sich dann bitte so, wie Sie persönlich das für angemessen halten.

Üblicherweise entsteht bei dieser Übung eine gewisse Unsicherheit. Womit soll man das letzte Resultat zwecks Beurteilung vergleichen? Mit den Resultaten der anderen Schüler? Das wird häufig (wenn auch nicht ganz "richtliniengetreu") bei der Benotung von Klassenarbeiten gemacht. Oder sollte man nicht auch mitberücksichtigen, ob sich der Schüler im Vergleich zu früher verbessert bzw. verschlechtert hat? Das wäre dann eine "für ihn" gute oder schlechte Leistung. Im ersten Fall würde man in der Tabelle senkrecht, im zweiten Fall waagerecht vergleichen.

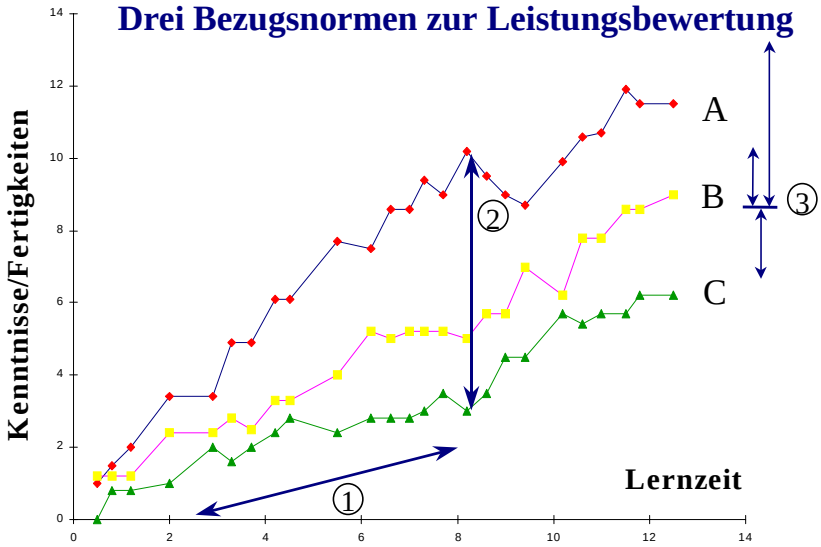
Das hätte bei einigen Schülern erhebliche Konsequenzen. Schüler 3 hätte je nach Vergleichsperspektive einmal Pluspunkte (weil überdurchschnittliches Niveau) und einmal Minuspunkte (weil abfallende Tendenz). Bei Schüler 6 wäre das genau umgekehrt. Solche Vergleichsperspektiven nennt man *Bezugsnormen*. Der Vergleich mit anderen ist ein sozialer Vergleich. Deshalb spricht man von einer *sozialen* Bezugsnorm. "Gut" ist das, was über dem Durchschnitt ist, "schlecht" ist das, was darunter liegt. Der Vergleich des Schülers mit sich selbst (ansteigende oder abfallende Tendenz) ist ein individueller Vergleich. Man spricht deshalb auch von einer *individuellen* Bezugsnorm. (Gleichbedeutend werden mitunter auch die Begriffe „autonome“ oder „temporale“ Bezugsnormen verwandt.)

Diese beiden Bezugsnormen sind übrigens keineswegs eine Erfindung theoretisierender Psychologen. Sie finden sich in vielen alltäglichen Leistungsbeurteilungen wieder, wie sie z. B. in Sportsendungen vorgenommen werden: "Er ist eindeutig der Schnellste von allen. Sein Sieg ist eine hervorragende Leistung" (soziale Bezugsnorm). "Er ist über sich hinausgewachsen und ist Dritter geworden. Diese Steigerung hätte in dieser Saison niemand für möglich gehalten - eine hervorragende Leistung" (individuelle Bezugsnorm). Im übrigen ist die Unterscheidung der beiden Vergleichsperspektiven der Leistungsbeurteilung nicht neu. Sie findet sich mehr oder weniger implizit z. B. in Äußerungen von A. Fischer, J. F. Herbart oder J. H. Pestalozzi.

## Genauere Klärung des Sachverhaltes

Wenn auch als Vergleichsperspektiven also schon länger bekannt und im Alltag intuitiv angewendet, sind die genaueren Bedingungen und Folgen der verschiedenen Bezugsnormen erst in den letzten 25 Jahren genauer untersucht worden. Dass die Bezugsnormen trotz ihrer erheblichen Folgen für die Beurteilung so lange eher unbedeutend erschienen, liegt wohl daran, dass sie typische Hintergrundvariablen sind. Das ist genauso wie bei der Wahrnehmung. Das gleiche Objekt wirkt vor einem dunklen Hintergrund heller als vor einem hellen Hintergrund. Der Hintergrund hat also erheblichen Einfluss auf das, was man sieht. Trotzdem wird er selbst als wirksame Größe nicht erkannt. Er geht statt dessen als nicht weiter beachtete Konstante in den Wahrnehmungsprozess ein.

Was bei der Wahrnehmung (meistens) Sinn macht, kann bei der Leistungsbeurteilung deshalb Unklarheiten schaffen, weil es - wie oben bei der kleinen Beurteilungsaufgabe erlebt - qualitativ verschiedene Bezugsnormen gibt, die als Hintergrund ganz verschiedene Seiten desselben Resultates sichtbar machen. Die folgende Abbildung soll das verdeutlichen.




1 = individuelle Bezugsnorm
2 = soziale Bezugsnorm
3 = sachliche Bezugsnorm

A, B, C = Schüler mit unterschiedlichem Lerntempo

Die Abbildung zeigt die Kenntnisse und Fertigkeiten, die drei fiktive Schüler A, B und C in einer bestimmten Lernzeit (z. B. einem Schulhalbjahr) in einem bestimmten Bereich erworben haben. Dieser Bereich könnte z. B. der aktive Wortschatz in einer Fremdsprache sein oder die Qualität der Aussprache oder die Zahl richtig geschriebener Wörter oder die übersprungenen Zentimeter beim Hochsprung und anderes mehr. Da als Folge von Lernen Kenntnisse und Fertigkeiten in der Regel zunehmen, steigen alle drei Lernkurven an. Weiterhin haben Lernkurven üblicherweise Schwankungen. Das liegt z. T. an der wechselnden Tagesform oder an anderen Zufälligkeiten. Wichtiger ist in diesem Zusammenhang aber der Einfluss, den Anstrengung und Bemühen des Schülers sowie Art und Intensität des Übens auf den Lernzuwachs haben. Da dieser Lerneinsatz erheblich variieren kann, schwankt auch der jeweilige Lernzuwachs über die Lernzeit hinweg.

Schließlich trägt die Abbildung realistischerweise noch der Tatsache Rechnung, dass nicht alle Schüler gleich schnell lernen. Die drei Lernkurven sind verschieden steil. Üblicherweise gibt es nämlich individuelle Unterschiede in der Lernfähigkeit für bestimmte Dinge - worauf auch immer diese Unterschiede wiederum zurückzuführen sind. Beim aktiven Wortschatz wird das sicher ganz andere Ursachen haben als beim Hochsprung. Das soll hier nicht weiter interessieren. Gemeinsam ist Fähigkeitsunterschieden, dass sie sich in der Regel nicht von einem Tag auf den anderen verändern.

Die beiden Vergleichsarten, die in der "Kleinen Beurteilungsaufgabe" oben bereits praktisch demonstriert wurden, sind hier als Pfeile, d. h. als Vergleichsperspektiven 1 und 2 eingezeichnet. (Die dritte Vergleichsperspektive, nämlich die sachliche Bezugsnorm, wird anschließend behandeln). Die soziale Bezugsnorm, also der Vergleich mit anderen (Pfeil 2), macht sehr gut deutlich, wer auf einem jeweiligen Gebiet zu den besseren und wer zu den schlechteren Schülern gehört. Wenn die Schüler hier hinreichend verschieden sind, entsteht ein relativ konstantes Leistungsbild. Schüler A ist gleichblei-

bend besser als Schüler B und der wiederum besser als Schüler C. Da zeitstabile Leistungsunterschiede meist zeitkonstanten Ursachen, insbesondere Fähigkeiten zugeschrieben werden, hebt diese Vergleichsperspektive überdauernde Kompetenzunterschiede zwischen Schülern besonders deutlich hervor.

Die Vergleichsperspektive der sozialen Bezugsnorm ist überall dort sinnvoll, wo es darum geht, die dauerhaft Besten herauszufinden. Wer beispielsweise aus einer Zahl von Bewerbern die oder den Besten auswählen will, weil etwa die zu besetzende Position so anspruchsvoll und wichtig ist, dass niemand überqualifiziert sein kann (z. B. Besetzung eines Lehrstuhls an der Universität), der tut gut daran, möglichst zuverlässig und genau die bislang erbrachte Leistung zwischen den Bewerbern zu vergleichen. Das erhöht die Chance, diejenige Person mit der Position zu betrauen, die nach erwiesener Leistungsfähigkeit auch für die Zukunft die besten Ergebnisse erwarten lässt. Ähnliches gilt, wenn man bei begrenzten Mitteln Talentförderung betreiben will. Auch wenn im Prinzip viele förderungswürdig wären, wird man bei begrenzten Mitteln diejenigen auswählen, die im Vergleich zu den anderen die besten Effekte erwarten lassen. Damit werden die knappen Mittel am wirksamsten eingesetzt. Im übrigen wird die Zuteilung nach bislang erwiesener Leistung als gerecht erlebt, gerechter jedenfalls als eine Zuteilung nach persönlichen Beziehungen, Parteizugehörigkeit, Herkunft, Geschlecht, Religion u. a. (Heckhausen, 1974).

Die Anwendung der sozialen Bezugsnorm muss sich übrigens nicht auf die Ermittlung der (relativ) Besten beschränken. Sind beispielsweise Mittel zur besonderen Förderung der schwachen und langsamen Schüler vorhanden, so ist es natürlich sinnvoll, die (relativ) schwächsten auszuwählen, um sie zu fördern. Auch dieses würde einen sozialen Vergleich erfordern. Rationaler wäre es allerdings, sich hierbei an sachliche Bezugsnormen (s. unten) zu orientieren. Damit könnte man nämlich verhindern, dass unnötigerweise bei den relativ leistungsschwächsten einer Schülergruppe eine Förderungsnotwendigkeit festgestellt wird, obwohl vielleicht alle Schüler dieser Gruppe die hier erforderlichen Kompetenzen bereits in hinreichendem Maß erworben haben und deshalb die Fördermittel in andere Bereiche fließen könnten. Das setzt voraus, dass solche inhaltlich definierten Standards für hinreichende Kompetenzgrade vorliegen. So etwas ist aber eher die Ausnahme, weswegen hilfsweise soziale Bezugsnormen häufiger verwandt werden, als es rational begründbar wäre. Im Schulalltag werden soziale Bezugsnormen wohl am häufigsten bei der Zensurengebung verwandt, obwohl das bei genauer Beachtung von Zensurendefinitionen eigentlich nicht zulässig ist.

Jede Bezugsnorm hat ihre "blinden Flecken". Die sozialen Bezugsnorm hat mindestens drei. Der erste ist, dass man ohne weitere Hilfsmittel immer nur innerhalb einer gegebenen Schülergruppe vergleichen kann. Der Lehrer vergleicht innerhalb einer Schulklasse, bestenfalls innerhalb seiner Schule, ohne zu wissen, wo beim Vergleich mit Schülern aus anderen Schulen seine Schüler liegen würden. Er verwendet ein sog. *klasseninternes Bezugssystem*. Das kann zu bizarren Fehlbeurteilungen führen, die in Deutschland schon von Ingenkamp (1977) in alarmierender Weise empirisch nachgewiesen wurden: Die gleiche Leistung wird mit "gut" oder "mangelhaft" beurteilt, je nachdem, ob der Schüler in einer leistungsstarken oder leistungsschwachen Schulklasse ist. Wie aktuelle Daten (TIMSS) zeigen, hat sich an diesem Missstand nicht viel geändert. Das ist besonders dann kritisch, wenn von solchen (Zensuren-)Beurteilungen Berechtigungen wie Studienplätze, begehrte Ausbildungsplätze, Stipendien usw. abhängen.

Dieser "blinde Fleck" der sozialen Bezugsnorm ist wegen seiner gravierenden Konsequenzen schon frühzeitig bemerkt worden (zusammenfassend Ingenkamp, 1977). Zwei weitere wurden erst später beschrieben (Rheinberg, 1980; zusammenfassend Rheinberg & Krug, 1999). Der zweite "blinde Fleck" liegt darin, dass die soziale Bezugsnorm den gemeinsamen Lernzuwachs aller unsichtbar macht. Dass in Abbildung 1 alle drei Schüler im Verlauf der Lernzeit immer mehr können und dazulernen, wird ausgeblendet. Es zählen ja nur die Unterschiede zwischen den Schülern. So kommt es, dass der Schüler C gleichbleibend "schlechte Leistungen" hat obwohl auch er über die Zeit merklich besser wird. Von daher überrascht nicht, wenn mehr als die Hälfte der Schüler von Lehrern, die sich ausschließlich an sozialen Bezugsnormen orientierten, am Schuljahresende sagten, sie könnten jetzt nur gleichviel oder sogar weniger (!) als zu Schuljahresbeginn (Rheinberg, 1980).

Der dritte "blinde Fleck" betrifft die Schwankungen im Lernzuwachs. Wenn in Abbildung 1 der schwächere Schüler C eine Veränderung in den beurteilten Lernergebnissen sehen soll, so müsste er Schüler B leistungsmäßig überholen. Bei hinreichend leistungsverschiedenen Schülern ist das aber eher unwahrscheinlich. So kann es sein, dass dieser Schüler gleichbleibend "schlechte Leistungen" rückgemeldet bekommt, gleichgültig, ob er sich angestrengt hat und einen für ihn ungewöhnlichen Zuwachs erzielt hat oder ob er gar nichts tut und noch weiter zurückfällt. Dasselbe gilt analog auch für andere Leistungsniveaus. Die soziale Bezugsnorm zeigt nur in Ausnahmefällen, d. h. bei leistungsmäßigen "Überholmanövern", wie das eigene Lernbemühen sowie die Art des Übens Einfluss auf das Lernresultat haben. Der zweite und der dritte "blinde Fleck" der sozialen Bezugsnorm haben ungünstige Auswirkungen auf die Lern- und Leistungsmotivation (Rheinberg, 1980, 1982).

*Individuelle Bezugsnorm*. Gerade unter diesem Aspekt erscheinen individuelle Bezugsnormen günstiger. (Das ist der Pfeil 1 in Abbildung 1). Hier wird im zeitlichen Längsschnitt ein jetzt erzieltes Ergebnis daran gemessen, was der Schüler auf diesem Gebiet zuvor erreicht hat. Damit geht der individuelle Lernzuwachs direkt in die Leistungsbeurteilung ein und wird besonders deutlich gemacht. Bei Lehrern, die sich nicht nur an sozialen, sondern zugleich auch an individuellen Bezugsnormen orientierten, gaben immerhin etwa zwei Drittel der Schüler an, sie könnten jetzt am Schuljahresende mehr als zu Schuljahresbeginn (Rheinberg, 1980).

Auch die Schwankungen im Lernverlauf werden unter individueller Bezugsnorm wie unter einem Vergrößerungsglas sichtbar gemacht. Schließlich sind es ja gerade die Veränderungen der Kenntnisse und Fähigkeiten, die hier in der Leistungsbeurteilung direkt zum Ausdruck kommen. Auf jedem Leistungsniveau sind gute (=besser als zuvor) wie auch schlechte Leistungen (=schlechter als zuvor) möglich. Von daher bestehen für alle Schüler etwa gleich gute Voraussetzungen, den Zusammenhang zwischen eigenen Lernbemühungen und Lernerfolg wahrzunehmen. Es zeigte sich, dass leistungsschwächere Schüler von der individuellen Bezugsnorm besonders profitieren, ohne dass leistungsstärkere benachteiligt wären. Allerdings ist schon hier einschränkend zu beachten, dass in (fast) allen Untersuchungen die individuelle Bezugsnorm als *zusätzliche* Beurteilungsperspektive eingeführt war, d. h. in Kombination mit anderen Bezugsnormen auftrat (Rheinberg, 1998).

Letzteres überrascht nicht. Bei allen motivationalen Vorzügen hat die individuelle Bezugsnorm natürlich auch ihre "blinden Flecke". Der wichtigste ist hier, dass überdau-

ernde Leistungsunterschiede zwischen Schülern ausgeblendet werden. Das hat für schwache, aber auch durchschnittliche Schüler einerseits den Vorteil, nicht durch ständig leistungsstärkere Mitschüler entmutigt zu werden. Andererseits ergibt sich aber der Nachteil, dass der Schüler eine ausgesprochen wichtige Informationsquelle zu sich selbst verliert. Aus der Sozialpsychologie ist lange schon bekannt, dass Menschen soziale Vergleiche suchen, um sich ihrer Einschätzungen sicherer zu werden. Das gilt insbesondere auch für Einschätzungen der eigenen Fähigkeit (Meyer, 1984). Würde man ausschließlich die individuelle Bezugsnorm verwenden, würden vielleicht alle zurecht die Überzeugung gewinnen, dass sie dazulernen können, wenn sie sich anstrengen. Sie wären sich allerdings höchst unsicher, auf welchem Gebiet sie besondere, vielleicht außergewöhnliche Kompetenzen haben und auf welchem weniger. Das verletzt das Bedürfnis nach sicherer Selbsteinschätzung und kann zudem zu unsinnigen Entscheidungen und zu Enttäuschungen führen - etwa bei der Berufs- oder Studienwahl. So überrascht nicht, wenn sich die Schüler im Rahmen eines etwas artifiziellen Unterrichtsexperiments wieder mehr Informationen zur sozialen Bezugsnorm wünschten, nachdem ihre Lehrer zur (fast) ausschließlichen Anwendung individueller Bezugsnormen gebracht worden waren (Rheinberg, 1998).

Abgesehen davon würde die individuelle Bezugsnorm als alleinige Beurteilungsperspektive zu irrationalen Konsequenzen führen, sofern an Leistungsbeurteilungen irgendwelche Berechtigungen geknüpft wären. So würde der Schüler, der sich im letzten Jahr vom "mangelhaft" auf "ausreichend" hochgearbeitet hat, einen Studienplatz in einem anspruchsvollen Numerus Clausus Fach erhalten, der Schüler, der konstant bei "sehr gut" läge jedoch nicht. Von daher sind Beurteilungen unter individueller Bezugsnorm bestens geeignet, wenn es darum geht, möglichst veränderungssensible, detaillierte Rückmeldungen *innerhalb* eines Ausbildungsabschnittes zu geben, die über günstige Motivationsauswirkungen den Lernerfolg fördern. Bei Leistungsbeurteilungen dagegen, die dauerhafte Berechtigungen *außerhalb* dieses Ausbildungsabschnittes vermitteln (z. B. Studienplätze oder andere Zugangsberechtigungen), können individuelle Bezugsnormen keine nennenswerte Rolle spielen. Man könnte sie allenfalls als "letzten Trend" mit berücksichtigen, der zu einer leichten Auf- oder Abwertung einer anders vorgenommenen Beurteilung führt.

*Sachliche Bezugsnorm*. Muss dann die gerade angesprochene "andere" Beurteilung unter sozialer Bezugsnorm erfolgen? Nicht notwendigerweise. Gerade bei Beurteilungen, die bestimmte Kompetenzen ausweisen sollen, die auch Personen oder Instanzen außerhalb eines Ausbildungsabschnittes informieren, sind inhaltlich beschriebene Standards hilfreich. Solche inhaltlich verankerten Standards nennt man sachliche Bezugsnormen. Der Vergleichsstandard liegt hier nicht in bereits erbrachten eigenen oder fremden Leistungen, sondern in Anforderungen, die in der Sache selber liegen: Man schafft es, über einen bestimmten Wassergraben zu springen oder man fällt hinein.

Sachliche Bezugsnormen werden überall dort verwandt, wo bestimmte Mindestkompetenzen, insbesondere wegen gravierender Folgen erreicht sein müssen und wo sich solche Mindestkompetenzen messen lassen. In solchen Fällen sind sachliche Bezugsnormen meist mit Alternativentscheidungen verbunden (z. B. Lernziel erreicht oder nicht?) Sie könnten im Prinzip aber auch bei abgestuften Urteilen (z. B. Zensuren, s. unten) herangezogen werden. So wurden im Rahmen der TIMSS- und der PISA-Studie verschiedene Kompetenzstufen inhaltlich definiert und voneinander abgegrenzt (s. Kap. 16 und 19). Typische Beispiele für Alternativentscheidungen auf der Basis von Min-

deststandards wären dagegen die Führerschein- oder die Pilotenscheinprüfung. Hier sind die Mindeststandards inhaltlich festgeschrieben. Die Beurteilung ist unabhängig davon, ob ein Kandidat besser oder schlechter als die anderen Prüflinge abschneidet (soziale Bezugsnorm). Im Extremfall könnten sogar alle Kandidaten einer Prüfungsgruppe durchfallen. Gänzlich unerheblich ist auch, ob sich der Kandidat gesteigert hat oder nicht (individuelle Bezugsnorm), solange die Steigerung nicht dazu geführt hat, das inhaltlich definierte Kompetenzniveau zu erreichen.

Im schulischen Bereich ist meist der Lehrplan Anker für sachliche Bezugsnormen. Von daher werden sachliche Bezugsnormen auch als "curriculare", mitunter auch als "lehrzielorientierte" oder "kriteriale" Bezugsnormen bezeichnet (Klauer, 1987). Was solche lehrplangeforderten Kompetenzgrade betrifft, so haben sowohl soziale als auch individuelle Bezugsnormen einen weiteren "blinden Fleck". Ob alle Schüler einer Klasse viel mehr oder viel weniger können, als das vom Lehrplan gewünscht ist, bleibt sowohl beim sozialen Vergleich zwischen verschiedenen Schülern, als auch beim individuellen Vergleich mit vorherigen Resultaten desselben Schülers unsichtbar. Das kann man erst sehen, wenn man die vorliegenden Resultate mit klaren inhaltlich bestimmten Maßstäben, also mit sachlichen Bezugsnormen vergleicht.

Liest man amtliche Zensurendefinitionen, so erkennt man häufig den allerdings etwas halbherzigen Versuch, Zensuren über solche sachliche Bezugsnormen zu bestimmen. Halbherzig ist der Versuch insofern, als meist die "durchschnittlichen Anforderungen" als inhaltlicher Anker genannt werden, ohne diese genauer zu bestimmen. Wollte man die Zensurengebung tatsächlich an sachliche Bezugsnormen knüpfen, so müsste man den beurteilenden Lehrern pro Fach, Jahrgangsstufe und Schulform sehr genau sagen, was jemand können muss, um ein "ausreichend" oder ein "gut" zu bekommen. Dabei würde es sicher nicht nur theoretisch, sondern auch tatsächlich geschehen können, dass ganze Schulklassen ein "gut" oder "sehr gut", aber auch ganze Schulklassen (vielleicht sogar ganze Schulen) nur "mangelhaft" oder "ungenügend" erhielten. Die erwähnten Untersuchungen zum klasseninternen Bezugssystem (Ingenkamp, 1977) lassen so etwas sogar für die Schulen innerhalb eines einzigen Schulbezirkes vermuten!

Solche, über Lehrplaninhalte genau definierten Kompetenzgrade wären für die beurteilenden Lehrer sicher hilfreiche Ankerpunkte. Je mehr allerdings Schulen ihr eigenes Profil entwickeln sollen, um so aufwendiger wird eine solche inhaltliche Definition. Das sollte aber kein prinzipielles Hindernis sein. Die Kompetenzstufenanalysen der TIMSS- und PISA-Studie zeigen, daß schulübergreifende Kriteriendefinitionen durchaus möglich und sinnvoll sind. Man muß sich allerdings darüber klar sein, daß hinter solchen Kriterien bestimmte Vorstellungen darüber stehen, welche Kompetenzen Unterricht vermitteln soll. Insbesondere, wenn diese Kriterien nicht gänzlich mit den Lehrplänen übereinstimmen, ist man gut beraten, genau zu prüfen, inwieweit man diese Vorstellungen inhaltlich akzeptiert. Der Hinweis, daß bestimmte Inhaltskonzepte (z.B. Mathematics Literacy) international anerkannt seien, ist zweifellos wichtig, kann aber für sich allein kein hinreichendes Argument sein.(In den deutschen TIMSS- bzw. PISA-Erhebungen beispielsweise wird über zusätzliche Aufgaben versucht, nationale Curriculumsbesonderheiten zu berücksichtigen.)

Schwieriger ist es, wenn die Schulaufsicht die Konsequenzen ihrer eigenen Anweisungen nicht völlig überschaut. So gibt es in einigen Bundesländern den sog. Dritteler-lass. Danach muss eine Klassenarbeit, bei der ein Drittel der Schüler die Note "mangel-

haft" und "ungenügend" hat, wiederholt werden, es sei denn, der Schulleiter erteilt nach genauer Prüfung eine Ausnahmegenehmigung. Diese Regelung drängt dem Lehrer im unteren Teil der Notenskala faktisch die soziale Bezugsnorm auf. In leistungsschwachen Schulklassen bringt ihn das zwangsläufig in Widerspruch zur Zensurendefinition, die den Anker ja bei den Anforderungen (sachliche Bezugsnorm) und nicht beim Klassendurchschnitt (soziale Bezugsnorm) vorsieht. Solche Inkonsistenzen werden in der Praxis deshalb kaum auffällig, weil die sachliche Bezugsnorm, wie schon erwähnt, nur halbherzig und vage vorgegeben ist. Wären die Zensurendefinitionen statt des Verweises auf "durchschnittliche Anforderungen" tatsächlich inhaltlich exakt beschrieben, würde dieser Widerspruch viel schärfer in Erscheinung treten und zu produktiven Kontroversen führen.

Man darf allerdings nicht verkennen, dass die Erstellung solcher inhaltlich definierten Zensurenstandards für jedes Fach, jede Klassenstufe und jede Schulform nicht nur einen enormen Konstruktionsaufwand erfordern, sondern auch die Flexibilität und die Freiheitsgrade der didaktischen Gestaltung des Lehrers einengen. Schließlich werden über die Zensurenstandards notwendig bestimmte Inhalte festgelegt. Ob man solche stärkere Standardisierung des Curriculums begrüßen oder bedauern sollte, kann hier nicht diskutiert werden. Sie würde jedenfalls eine wahrscheinliche Folge sein.

Abgesehen von solchen praktischen Schwierigkeiten, hat auch die sachliche Bezugsnorm ihre "blinden Flecken". Sie informiert genau genommen nur über die jeweils umschriebenen Fertigkeiten oder Kenntnisse. Ob diese Fertigkeiten auch auf besondere Lernfähigkeiten auf diesem Bereich verweisen oder eher Selbstverständlichkeiten in dem jeweiligen Ausbildungsgang sind, ist dem inhaltlichen Kriterium selbst nicht anzusehen. Die Tatsache, dass sich jemand z. B. im Zahlenraum von 100 ohne Hilfe frei bewegen kann, kann eine nützliche Information für den Lehrer sein, der wissen will, worauf er sich bei seiner Unterrichtsplanung stützen kann. Mit Blick auf das zu erwartende Lerntempo wäre es allerdings schon wichtig zu wissen, ob dieses Kriterium soeben von einem Erstklässler oder einem Zehntklässler erreicht wurde. Ohne den Vergleich mit anderen ist kaum zu beurteilen, wie schwer oder leicht es im allgemeinen ist, das fragliche Kriterium zu erreichen und ob man deshalb auf besondere Fähigkeiten und/oder besonderen Lerneinsatz dieses Schülers rückschließen kann. Diese Einschränkung gilt übrigens auch für den Schüler selbst. Auch ihm fehlen bei alleiniger Verwendung sachlicher Bezugsnormen Hinweise, auf welchem Gebiet er besondere Fähigkeiten besitzt und auf welchen Gebieten er sich besser nicht spezialisieren sollte, weil es viele andere Personen gibt, die die dort erforderlichen Kompetenzen schneller und besser erwerben.

Insbesondere, wenn sachliche Bezugsnormen als Mindeststandards alternativ formuliert sind (bestanden oder durchgefallen), haben sie als zweiten "blinden Fleck" die Unsensibilität gegenüber Lernfortschritten. Die Person, die zum fünften Mal durch die Führerscheinprüfung gefallen ist, weiß ohne Zusatzinformationen nicht, ob sie inzwischen schon etwas besser Auto fährt oder nicht. Sie weiß nur: es reichte auch dieses Mal nicht. Dieses Ausblenden der Lernzuwachsinformation - die ja den Kern der individuellen Bezugsnorm ausmacht - dürfte motivational ungünstig sein.

## Einige praktische Konsequenzen

Wie gezeigt, hat also jede Bezugsnorm ihre "blinden Flecken" und kann nicht für alle Zwecke eingesetzt werden. Es wäre ohnehin ein Irrtum zu glauben, ein Lehrer müsse sich auf eine Bezugsnorm festlegen. Folgt man Heckhausen (1974; 1989), so kommt es darauf an, dass Schüler lernen, sich unter *verschiedenen* Bezugsnormen zu bewerten. Dabei soll für die Zufriedenheit mit der eigenen Leistung (die sog. Selbstbewertung) die individuelle Bezugsnorm die Leitfunktion übernehmen, ohne dass Informationen zu anderen Bezugsnormen ignoriert werden. Um es an einem Grenzfall zu verdeutlichen: Ein Schüler, der sich in Mathematik von "ungenügend" auf "mangelhaft" hochgearbeitet hat, sollte wegen dieser Steigerung ähnliche Freude und Stolzaffekte erleben wie jemand, der sich von "gut" auf "sehr gut" steigert (individuelle Bezugsnorm). Gleichwohl sollte er zur Selbsteinschätzung wissen, dass es z. Z. noch viele andere Schüler gibt, denen Mathematik offenbar leichter fällt (soziale Bezugsnorm) und dass es noch viele Dinge gibt, die er zu lernen hat, um das versetzungssichernde "ausreichend" zu bekommen (sachliche Bezugsnorm). Entscheidend ist aber, dass es die Fortschritte unter individueller Bezugsnorm unübersehbar machen, dass auch er dazu lernt und dass sein bislang unbefriedigender Leistungsstand in Bewegung ist. Diese Wahrnehmung ist die Grundlage für den erwähnten positiven Affekt. Voraussetzung dafür ist allerdings, dass der Schüler solche Informationen über seine Lernzuwächse überhaupt bekommt.

Trivialerweise werden es Schüler kaum lernen, sich unter verschiedenen Bezugsnormen zu bewerten, wenn Lehrer nur eine einzige Bezugsnorm verwenden. In der Praxis scheint die Beurteilungssituation, dass ein Lehrer vor vielen etwa gleich alten Schülern steht, die alle das gleiche Curriculum mit den gleichen Tests durchlaufen, die soziale Bezugsnorm aufzudrängen. Es fanden sich jedenfalls Lehrer, die diese Bezugsnorm durchgängig anlegten - gleichgültig, ob sie Übergangsempfehlungen zu geben hatten, ob sie Noten verteilten oder ob sie mit dem Schüler allein über seine Leistungen sprachen (Rheinberg, 1980). Damit sorgen sie zwar für eine unübersehbare Konsistenz ihrer Urteile. Sie machen ihren Schülern allerdings klar, dass es nur die eine gültige Weise gibt, gute Leistungen zu haben, nämlich besser zu sein als andere. Das wirkte sich, wie schon erwähnt, besonders ungünstig bei leistungsschwächeren Schülern aus. Letzteres überrascht um so weniger, als diese Lehrer gute oder schlechte Schulleistungen bevorzugt mit hoher oder niedriger Fähigkeit/Begabung erklären. Das liegt ja - wie eingangs erwähnt - bei sozialer Bezugsnorm recht nahe. Weiterhin bieten sie für ihre Klassen einen möglichst gleichförmigen Unterricht an ohne für individuelle Abstimmungen wie z. B. Zusatzerklärungen oder besondere Übungen zu sorgen. Für leistungsschwächere Schüler ergibt sich damit die Situation, häufig vor Lernanforderungen zu stehen, die sie nicht schaffen, wobei der Leistungsvergleich mit anderen zeigt, dass andere ständig viel besser sind. Zudem geht aus den Reaktionen des Lehrers hervor, dass der desolate Leistungsstand auf einen Fähigkeitsmangel zurückgeht, der mithin kaum änderbar ist (Rheinberg, 1980; 1998).

Lehrer, die sich dagegen (auch) an individuellen Bezugsnormen orientierten, erzielten sehr viel günstigere Motivationseffekte bei ihren Schülern. Allerdings wechseln sie je nach Beurteilungskontext die verwandte Bezugsnorm. Die individuelle Bezugsnorm wird von ihnen typischerweise bei Leistungsrückmeldungen im Unterricht oder im Gespräch mit dem Schüler allein verwandt. Sie selbst wie auch der Schüler sehen durch den Vergleich mit vorangegangenen Leistungen viel deutlicher, wie im Verlauf der

Lernzeit die Kompetenzen wachsen und wie Lerngewinne vom eigenen Lerneinsatz abhängen (Rheinberg, 1980; 1998). Gleichwohl werden soziale (und sachliche) Bezugsnormen nicht ausgeblendet, weil (a) Zensuren zu vergeben oder Übergangsentscheidungen zu treffen sind und (b) die Situation des Klassenverbandes wie auch das Informationsbedürfnis der Schüler so etwas ohnehin kaum zulassen. Die dadurch erzeugte Bezugsnorm-Vielfalt fördert das Ziel, dass Schüler lernen, sich selbst mit Hilfe verschiedener Bezugsnormen zu bewerten. Dieses Ziel wäre allerdings gefährdet, wollte man versuchen, durch Benotungserlasse und besondere Unterrichtsformen die individuelle Bezugsnorm als alleinigen Bewertungsmaßstab durchzusetzen - was vernünftigerweise wohl niemand ernsthaft betreiben würde.

Bei den erwähnten Vorzügen einer zusätzlichen Berücksichtigung individueller Bezugsnormen darf man allerdings nicht übersehen, dass der Beurteilungsaufwand erhöht ist, weil der Lehrer ja viele individuelle Entwicklungen als Beurteilungsanker im Kopf haben muss. Zudem drängt diese Art der Beurteilung erfahrungsgemäß auch zu Versuchen, Unterrichtsanforderungen zumindest zeitweilig verschiedenen Lernständen der Schüler anzupassen. Die genauere Betrachtung individueller Lernverläufe macht nämlich unübersehbar, dass bestimmte Schüler von dem durchschnittsorientierten Unterrichtsangebot überfordert, andere hingegen gelangweilt sein werden. So etwas regt dazu an, bestimmte Zusatzangebote zu planen und anzuwenden. All das und anderes machen dem Lehrer bei der Verwendung individueller Bezugsnormen mehr Arbeit und erfordern zusätzliche Kompetenzen. In einer Reihe von Trainingsstudien wurden verschiedene Möglichkeiten erprobt, die Lehrern dieses Vorgehen unter Schulalltagsbedingungen ermöglichen sollen (Rheinberg & Krug, 1999).

Erst wenig weiß man über die Kombination von sachlichen und individuellen Bezugsnormen. Diese Kombination würde in der Leistungsbewertung ausdrücken, wie sehr sich jemand in der Annäherung an das aktuelle Lehrziel verbessert hat. Gleichwohl würde ganz spezifisch deutlich, was auf dem Weg zur Lehrzielerreichung im Einzelnen noch zu tun bleibt. Sofern die Lehrziele nicht zu hoch sind oder zu weit in der Zukunft liegen, müsste diese Bezugsnorm-Kombination günstige Auswirkungen haben. Das ist aber erst wenig untersucht (Rheinberg, 1998).Erste laborexperimentelle Versuche führten zu ermutigenden Ergebnissen. Damit diese Kombination für Lehrer unter Schulalltagsbedingungen realisierbar wird, müssten sachliche Bezugsnormen in Form exakter lehrzielbeschreibender Kriterien vorliegen. Diese Voraussetzung scheint aber nur ausnahmsweise gegeben.

Die Entwicklung solcher Kriterien wäre eine lohnende Aufgabe. Dies wäre zugleich auch eine notwendige Voraussetzung, um *lehrzielorientierte Prüfverfahren* zu entwickeln, die landes- oder gar bundesweit routinemäßig einsetzbar sind. Solche Verfahren wären für Lehrer eine große Hilfe. Testentwickler könnten dann nämlich relativ einfach ermitteln, welche Punktzahlen im jeweiligen Testverfahren durchschnittlich welchen Zensuren entsprechen. Lehrer hätten dann die Möglichkeit festzustellen, ob ihre eigene Zensurengebung in einem bestimmten Fach - verglichen mit dem Durchschnitt der Bundesrepublik - streng oder mild ist. Ob sie dann ihre Zensurengebung anpassen oder nicht, bliebe ihrer pädagogischen Entscheidung überlassen. Sie wüssten aber, dass es evt. Abweichungen gibt und könnten daraus begründet Schlüsse ziehen. Diese Möglichkeit haben sie z. Z. nicht. Allerdings wurden, wie erwähnt, im Rahmen der TIMSS- und der PISA-Studie zunächst für Forschungszwecke Kompetenzstufen inhaltlich definiert, die eine Orientierung an sachlichen Bezugsnormen erlauben. Wie erste Projekte zeigen,

lassen sich die daraus entwickelten Beurteilungsverfahren in der Praxis sinnvoll einsetzen (s. Kap. 16 und 19).

Wo im Schulalltag sachliche Bezugsnormen und darauf abgestimmte Messverfahren aber noch nicht vorliegen, wird man sich mit Provisorien behelfen - sofern man nicht alles so weiterlaufen lassen will, wie bisher. Um zumindest innerhalb derselben Schule gravierende Unterschiede in der Leistungsbeurteilung verschiedener Lehrer sichtbar zu machen, kann man regelmäßig Parallelarbeiten in Klassen derselben Stufe schreiben, die dann von allen hier beteiligten Fachlehrern korrigiert werden. Welche Konsequenzen diese Lehrer dann aus den möglicherweise auffälligen Leistungs- und Beurteilungsunterschieden ziehen, sollte ihnen überlassen bleiben. Wichtig ist zunächst, dass solche Unterschiede überhaupt erkannt werden und dann vielleicht zu Diskussionen und vom Kollegium verantworteten Änderungen der Beurteilungspraxis führen. Dieses Verfahren zum Vergleich von Leistungsniveaus und Beurteilungsstandards ließe sich auch auf verschiedene Schulen ausdehnen.

Man muss sich allerdings klar darüber sein, dass man damit lediglich Urteilsabweichungen unter sozialer Bezugsnorm sichtbar macht - was durchaus schon ein wichtiger Informationsgewinn sein kann. Ob jedoch ganze Schulen leistungsmäßig weit über dem Niveau liegen, das der Lehrplan vorsieht oder ob ganze Klassen oder Schulen diese Standards klar verfehlen, lässt sich damit nicht feststellen. Dazu bräuchte man, wie oben schon gesagt, sachliche Bezugsnormen. Aber auch wenn diese samt Messverfahren geliefert werden, bleibt es nach wie vor Sache des Lehrers, den Lernfortschritt des einzelnen Schülers unter individuellen Bezugsnormen sichtbar zu machen und bewertend hervorzuheben.

## Verwendete Literatur


Furck, C.L. (1975). *Das pädagogische Problem der Leistung in der Schule*. (5. Aufl.). Weinheim: Beltz.

Heckhausen, H. (1974). *Leistung und Chancengleichheit*. Göttingen: Hogrefe.

Heckhausen, H. (1989). *Motivation und Handeln*. (2. Aufl.). Berlin: Springer.

Ingenkamp, K.H. (1977). *Die Fragwürdigkeit der Zensurengebung*. (7. Aufl.). Weinheim: Beltz.

Klauer, K.J. (1987). Fördernde Notengebung durch Benotung unter drei Bezugsnormen. In R. Olechowski & E. Persy (Eds.), *Fördernde Leistungsbeurteilung* (S. 180-206). Wien: Jugend & Volk.

Meyer, W.-U. (1984). *Das Konzept von der eigenen Begabung*. Stuttgart: Huber.

Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation*. Göttingen: Hogrefe.

Rheinberg, F. (Hrsg.). (1982). *Bezugsnormen zur Leistungsbewertung: Analyse und Intervention*. (Jahrbuch für Empirische Erziehungswissenschaft 1982). Düsseldorf: Schwann.

Rheinberg, F. (1998). Bezugsnorm-Orientierung. In D.H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (S. 39-43). Weinheim: Beltz, PVU.

Rheinberg, F. & Krug, S. (1999). *Motivationsförderung im Schulalltag*. (2. Aufl.). Göttingen: Hogrefe.

## Weiterführende Literatur

Heckhausen, H. (1989). *Motivation und Handeln*. (2. Aufl.). Berlin: Springer.

Klauer, K.J. (1987). Fördernde Notengebung durch Benotung unter drei Bezugsnormen. In R. Olechowski & E. Persy (Eds.), *Fördernde Leistungsbeurteilung* (S. 180-206). Wien: Jugend & Volk.

Rheinberg, F. & Krug, S. (1999). *Motivationsförderung im Schulalltag*. (2. Aufl.). Göttingen: Hogrefe.