

LehrerInnen sollten Lernprozesse im Blick auf einzelne SchülerInnen und auf die Klasse insgesamt beobachten und auswerten, den eigenen Unterricht beobachten (lassen) und Erfahrungen im Gespräch mit KollegInnen und Eltern durchdenken, um die Wirkungen der eigenen Arbeit zu überprüfen.

Das *Kollegium* sollte gemeinsam – im Blick auf die Schule als ganze – ein Schulprogramm entwickeln und anhand fokussierter Bestandsaufnahmen (auch mit Hilfe von externen “critical friends”) regelmäßig fortschreiben.

Die *Schulaufsicht* sollte in den schulinternen Diskussionen den Fremdblick stärken, kritische Fragen stellen, Weiterentwicklung fordern, vor allem aber konstruktive Ansätze anerkennen und unterstützen.

Das *Ministerium* sollte mit Blick auf das Gesamtsystem allgemeine Problemstellen ausmachen, politische Prioritäten überprüfen und Ressourcen ziel- und bedarfsorientiert organisieren.

Und noch eins: Vieles, was im Moment als “Evaluation” verkauft wird, ist eigentlich Grundlagenforschung. Zu den Zusammenhängen zwischen verschiedenen Faktoren im System gibt es doch schon eine Fülle von Publikationen und Daten, die man mit vergleichsweise geringerem Aufwand aufbereiten und gezielt auf aktuelle Fragen hin auswerten könnte, statt mit hohen Kosten Berge neuer Daten zu erheben. Wenn von Ökonomie die Rede ist, sollte dieser Anspruch auch für die Evaluation selbst gelten.

Müsste man nicht stattdessen SchülerInnen, Eltern, LehrerInnen und die Öffentlichkeit differenzierter nach ihren Ansprüchen an die Schule, nach ihren unterschiedlichen Erfahrungen und Kritikpunkten befragen? Damit würden die *Wertfragen* und die konkurrierenden Perspektiven auf Unterricht deutlicher, und man könnte den Blick zugleich auf unzulängliche *Bedingungen* schulischer Arbeit erweitern. Es geht doch nicht nur darum, was ist, sondern auch darum, was *sein* soll.

KAPITEL 3

Alltägliche Leistungsbeurteilung durch Lehrer

Friedrich-Wilhelm Schrader und Andreas Helmke

Die Beurteilung von Schülerleistungen ist integraler Bestandteil der beruflichen Tätigkeit von Lehrkräften. Die Zensurenvergabe wird meistens als die prototypische Beurteilungsaufgabe von Lehrkräften angesehen. Leistungsbeurteilungen dienen aber auch dazu, die Unterrichtsgestaltung und das Lernen zu verbessern. Für beide Aufgaben spielen alltägliche Leistungsbeurteilungen im Klassenzimmer eine zentrale Rolle. Deshalb sollen diese im Folgenden genauer beleuchtet werden. Dass es dabei um *Leistungsbeurteilungen* geht, heißt nicht, dass die Einschätzung anderer Schülermerkmale, insbesondere affektiv-motivationaler Merkmale, unwichtig für das erzieherische Handeln wäre. Aber die Lehrziele und Lernergebnisse, die in der schulischen Praxis im Vordergrund stehen, betreffen primär die Leistung.

In diesem Beitrag werden allgemeine Fragen des Diagnostizierens angesprochen, die Lehrkräfte aller Schulformen und Altersstufen betreffen. Auch wenn die grundlegenden Aufgaben und Funktionen auf einer abstrakten Ebene vergleichbar sind, unterscheidet sich die konkrete Ausgestaltung dieser Aufgaben – man vergleiche dazu etwa nur den Unterricht in der Grundschule mit dem in der gymnasialen Oberstufe – ganz erheblich.

Die Lehrkraft als Diagnostiker

Zwei Arten der Leistungsbeurteilung

Manche der Urteilsaufgaben von Lehrkräften ähneln denen professioneller Diagnostiker: Es wird ein *explizites* Urteil (sozusagen eine “Diagnose”) abgegeben; die Lehrkraft stützt sich dabei auf geeignete Informationen (“Daten”), die eigens zum Zwecke der Beurteilung erhoben wurden (Klassenarbeiten; mündliche Prüfungen; informelle Tests usw.). Das diagnostische Urteil kommt dadurch zustande, dass die gewonnenen Informationen mit einer *Norm*, d. h. mit einem Vergleichmaßstab in Beziehung gesetzt werden (Ingenkamp, 1997a). Man unterscheidet gewöhnlich drei Bezugsnormen: Eine Leis-

tung ist gut, wenn sie (a) besser ist als die der meisten anderen Lernenden bzw. des Durchschnitts (soziale Norm); (b) sich der Lernende im Vergleich zu früher verbessert hat (individuelle Norm); oder (c) ein angestrebtes Lehrziel oder Kriterium erreicht wurde (sachbezogene oder kriteriale Norm). Die Beurteilungen erfolgen hierbei in Situationen, in denen die Lehrkraft die Aufmerksamkeit gezielt (bei einer mündlichen Leistungsprüfung) und im Idealfall nahezu ungeteilt (bei der Korrektur schriftlicher Arbeiten) auf die diagnostische Aktivität richten kann. Die vorhandenen Informationen können (außerhalb der Unterrichtssituation) anschließend eingehend gesichtet, bewertet und gewichtet werden, und der gesamte Urteilsvorgang kann gründlich reflektiert werden.

Andere Beurteilungsleistungen sind *impliziter* Natur, d. h. es wird kein ausdrückliches Urteil verlangt, sondern der Urteilsvorgang läuft stark verkürzt ab, ohne am Ende in eine sprachlich mitteilbare Aussage übersetzt zu werden: Schülerleistungen werden lediglich insoweit registriert und intuitiv eingeschätzt, wie es nötig ist, um Entscheidungen treffen zu können (die Behandlung eines Themas abschließen; eine neue Aufgabenstellung einführen; eine bestimmte Frage stellen; einen bestimmten Schüler aufrufen). Solche Entscheidungen müssen während des Unterrichtens getroffen werden und daher meist sehr schnell erfolgen; sie können selten gründlich reflektiert werden, sondern erfolgen auf der Grundlage eingübter Denkroutinen mit einem vergleichsweise geringen Grad an bewusster Aufmerksamkeit und Kontrolle. Solche "Mikrodiagnosen" kommen dadurch zustande, dass allgemeine Erwartungen an die Leistung der Klasse oder einzelner Schüler mit aktuellen Beobachtungen abgeglichen und verknüpft werden; sie erfolgen eher unsystematisch, ungezielt und beiläufig neben der eigentlichen Unterrichtstätigkeit. Die Aufmerksamkeit ist häufig auf andere Aspekte der Situation gerichtet. Es werden intuitive und subjektive Einschätzungen vorgenommen, die beim erfahrenen Lehrer eng mit bestimmten Handlungsentwürfen und -routinen verknüpft sind und bei störungsfreien Abläufen kaum über die Bewusstseinsschwelle gelangen.

Das Unterrichten selbst ist eine hochkomplexe Tätigkeit; die Anforderungen, die dabei an den Lehrer gestellt werden, werden von Außenstehenden häufig unterschätzt: Es gibt eine Vielzahl von Informationen, die auf den Lehrer einströmen; vieles passiert gleichzeitig; es ist oft nicht genau vorhersagbar, was als nächstes passiert; es muss häufig sofort reagiert werden; und einmal gezeigte Lehrerreaktionen stellen für die Schüler/innen unter Umständen Präzedenzfälle dar, die die künftigen Verhaltensmöglichkeiten der Lehrkraft einengen können. Diese Anforderungen sind nur dadurch zu bewältigen, dass Unterrichtssituationen und -abläufe effizient organisiert und strukturiert werden (Shuell, 1996). Lehrkräfte unterscheiden sich erheblich darin, inwieweit ihnen das gelingt. Gleichzeitig stellt die so beschriebene Unterrichtssituation auch hohe Anforderungen an die Informationsaufnahme und -verarbeitung des Lehrers.

Die Rolle von Leistungserwartungen und diagnostischen Kompetenzen

Leistungsbeurteilungen von Lehrkräften werden von verschiedenen Faktoren beeinflusst. Selbst außerhalb der Unterrichtssituation erfolgt die Beurteilung oft nicht völlig unvoreingenommen und neutral, sondern wird durch Erwartungen und Voreinstellungen

verzerrt. Untersuchungen zeigen, dass dies selbst bei der Beurteilung schriftlicher Leistungen (und zwar nicht etwa nur bei Aufsatzleistungen, sondern sogar bei Mathematikarbeiten; siehe etwa Ingenkamp, 1997a) der Fall sein kann. Vor allem aber das Handeln von Lehrkräften im Unterricht wird durch *Erwartungen* gesteuert und beeinflusst. Das in Abbildung 1 dargestellte Modell soll die dabei ablaufenden Prozesse und Einflussfaktoren verdeutlichen helfen (zum besseren Nachvollzug sind die einzelnen Kästchen mit Nummern versehen, auf die im nachfolgenden Text Bezug genommen wird). Dabei sind die wichtigsten Einflussrichtungen (natürlich sind noch andere denkbar) als Pfeile eingezeichnet.

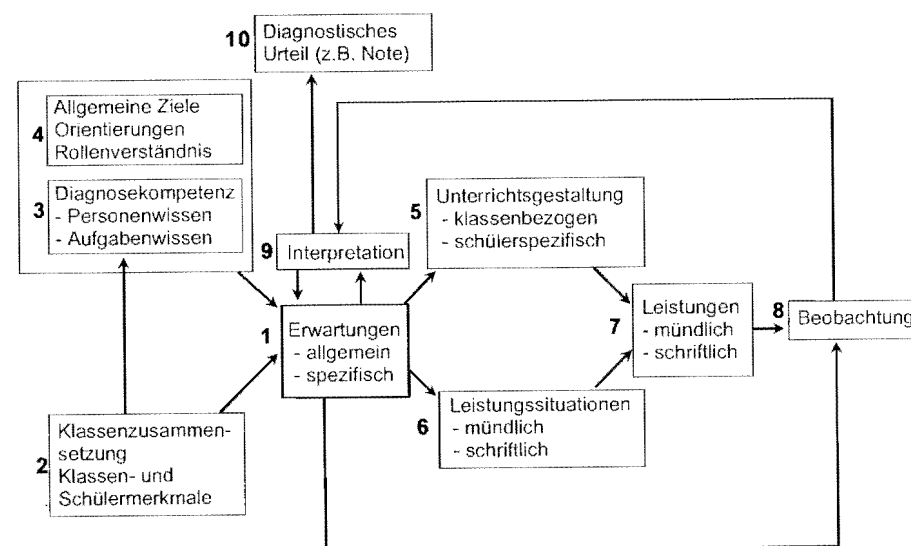


Abbildung 1: Leistungserwartungen, Unterricht und diagnostisches Urteil

Erwartungen [1] sind Vorstellungen darüber, was sein wird (wie sich eine Situation weiterentwickeln oder welche Folgen eine Handlung nach sich ziehen wird). Sie sind oft mit einem hohen Grad subjektiver Gewissheit verbunden, was in der Regel die Handlungssicherheit erhöht. Von Hypothesen würde man sprechen, wenn die Erwartungen als vorläufig und überprüfungsbedürftig betrachtet werden. Erwartungen des Lehrers ("wie sich bestimmte Schüler oder Klassen im Unterricht verhalten, was sie zu leisten imstande sind, wieviel Lernangebote ihnen in einer bestimmten Zeitspanne zuzumuten sind, durch welche Mittel man sie am besten beeinflussen kann, welche Probleme sie bereiten können usw."; Wahl, Weinert & Huber, 1997, S. 54) bestimmen ganz maßgeblich, wie in einer bestimmten Situation gehandelt wird, und tragen somit dazu bei, allgemeine pädagogische Ziele in konkrete Unterrichtshandlungen umzusetzen. Welche

Leistungen von der Klasse insgesamt oder von individuellen Schülern erwartet werden, hängt sicher immer auch vom tatsächlichen Leistungsvermögen der Klasse und der jeweiligen Schüler – allgemein also von der *Klassenzusammensetzung* [2] – ab. Wie genau, flexibel und realitätsangemessen diese Leistungserwartungen sind und wie zutreffend sie also die realen Gegebenheiten widerspiegeln, hängt von der *Diagnosekompetenz* [3] der Lehrkraft ab, also ihrer Fähigkeit, Schülermerkmale und Aufgabenschwierigkeiten zutreffend einzuschätzen. Dazu sind diagnostisches Wissen (über Fähigkeiten und Leistungen von Schüler und die Schwierigkeit von Aufgaben) und diagnostische Fertigkeiten (Beobachtungsfähigkeiten, Beherrschung von Diagnoseinstrumenten) erforderlich. Die diagnostische Kompetenz stellt neben dem Fachwissen, den didaktisch-methodischen Fähigkeiten und der Fähigkeit zur Klassenführung einen von vier Kompetenzbereichen dar, die den erfolgreichen Lehrer auszeichnen (Weinert, Schrader & Helmke, 1990). Das diagnostische Wissen hängt seinerseits von der Klassenzusammensetzung [2] ab. So ist es in einer sehr heterogenen Klasse leichter, Leistungsunterschiede zwischen Schülern zu erkennen.

In die Erwartungen der Lehrkraft fließen außerdem *allgemeine Ziele* (Erziehungsziele wie Selbständigkeit, Fairness, Leistungsbereitschaft, Tüchtigkeit) und *Orientierungen* (z. B. die Tendenz, sich bei der Beurteilung eher an sozialen oder individuellen Bezugsnormen zu orientieren) [4] ein. Genaue, d. h. präzise und realitätsangepasste Erwartungen finden sich bevorzugt bei Lehrkräften mit einem professionellen *Rollenverständnis* [4] und einer damit verbundenen diagnostischen Haltung: der Bereitschaft, sich um Objektivität und Unvoreingenommenheit zu bemühen, Erwartungen und Interpretationen als vorläufige Hypothesen zu betrachten, zu hinterfragen und ggf. zu korrigieren und sich für die Lernerfolge der Schüler verantwortlich zu fühlen. Zu verzerrten Wahrnehmungen kommt es häufig dann, wenn starke Affekte und Beeinträchtigungen des Selbstwerts erlebt werden, was meist zur Folge hat, dass Informationen nur oberflächlich und unvollständig aufgenommen und verarbeitet werden. Natürlich stehen die genannten Merkmale untereinander in einem sehr viel komplexeren Beziehungsverhältnis als hier angedeutet ist. Darauf soll in diesem Beitrag aber nicht weiter eingegangen werden.

Wie die *Unterrichtsgestaltung* [5] für die Klasse insgesamt und für einzelne Schüler aussieht, hängt sowohl von *allgemeinen*, auf das Leistungsniveau von Schülern einer bestimmten Jahrgangsstufe sowie der eigenen Klasse bezogenen, als auch von *schüler-spezifischen* Erwartungen ab. Erstere bestimmen etwa, wie schwer der Unterricht, einzelne Lehrinhalte oder Aufgaben sind; letztere sind dafür maßgeblich, wie *einzelne* Schüler im Unterricht behandelt werden, etwa welche Fragen an sie gerichtet werden. Die Erwartungen der Lehrkraft (sowohl hinsichtlich der Leistungsfähigkeit der Klasse als auch einzelner Schüler) haben ferner Einfluss darauf, wie *Leistungssituationen* [6] – und zwar sowohl mündliche als auch schriftliche Prüfungssituationen – gestaltet werden (wie schwer etwa eine Klassenarbeit oder die mündliche Prüfung für einen bestimmten Schüler ist). Grundlage für die Beurteilung sind dann die schriftlichen und mündlichen *Leistungen* [7] der Schüler, die diese bei gezielten Überprüfungen, aber auch ganz allgemein im Unterricht zeigen.

Die Wahrnehmung und *Beobachtung* [8] der von den Schülerinnen und Schülern gezeigten Leistung wird ihrerseits von den Erwartungen, Voreinstellungen und Hypothesen der Lehrkraft beeinflusst. Man spricht auch von der *Hypothesentheorie der Wahrnehmung*. Laien sind häufig der Meinung, die Wahrnehmung liefere eine vollständige und getreue Abbildung (sozusagen eine Kopie) der Realität. Menschen können aber immer nur einen kleinen Ausschnitt des gesamten Wahrnehmungsfeldes verarbeiten. Abhängig von den jeweiligen Bedürfnissen und Interessen richtet sich die Wahrnehmung bevorzugt auf das, was die Person erwartet. Im günstigen Fall führt dies zu einer gezielteren Beobachtung und damit zu einer genaueren Wahrnehmung des betrachteten Realitätsausschnitts; im ungünstigen Fall kommt es zu einer einseitigen, oberflächlichen und verzerrten Wahrnehmung. Welche dieser Konsequenzen eintritt, hängt davon ab, wie genau und zutreffend die Erwartungen sind und wie reflektiert und flexibel der Wahrnehmende damit umgeht. In der Literatur findet sich eine ganze Reihe von *Urteilsfehlern*, etwa die Tendenz zur Milde oder Strenge im Urteil, zur Abschwächung oder Akzentuierung von Unterschieden; die Tendenz, von bestimmten Merkmalen auf andere zu schließen oder die Tendenz, sich bei der Beurteilung von einem globalen Gesamteindruck leiten zu lassen (vgl. etwa Ingenkamp, 1997a).

Die wahrgenommenen Leistungen wirken schließlich auf die vorhandenen Erwartungen und Hypothesen zurück. Ob diese bestätigt oder korrigiert werden, hängt unter anderem von der *Interpretation* [9] der Beobachtungen durch die Lehrkraft ab. Hier spielt eine Rolle, wie die Leistungen erklärt, d. h. auf welche Ursachen sie zurückgeführt werden (Attribution). So können z. B. unerwartet gute Leistungen eines leistungsschwachen Schülers durch Glück oder besondere Anstrengung erklärt werden, die unerwartet schlechten Leistungen eines leistungsstarken Schülers dagegen durch Pech usw. Derartige Interpretationen können nicht nur die vorhandenen Erwartungen beeinflussen, sondern werden ihrerseits auch von diesen beeinflusst. Darüber hinaus werden die (modifizierten oder nicht-modifizierten) Einschätzungen und Erwartungen den Schülern im Zuge der Unterrichtsgestaltung oder in Leistungssituationen mehr oder weniger explizit zurückgemeldet.

Die beobachteten und interpretierten Leistungen sind schließlich Grundlage für *diagnostische Urteile* [10], die von Lehrkräften zu bestimmten Anlässen (Versetzung, Schulübertritt) gefordert werden. Hier ist vor allem an Urteile gedacht, die nach Abschluss des Unterrichts abgegeben werden (z. B. die Note in einem Versetzungs- oder Übertrittszeugnis), so dass dieses Modell keine Rückwirkung des resultierenden diagnostischen Urteils auf die Erwartungen und den Unterricht vorsieht. Um die Darstellung nicht unnötig zu komplizieren, ist hier ebenfalls nicht berücksichtigt worden, dass Beurteilungen und Erwartungen längerfristig auch zu einer Veränderung des diagnostischen Wissens führen. Und schließlich ist die strikte Hintereinanderanordnung von Erwartungen, der Gestaltung des Unterrichts und von Leistungssituationen, daraus resultierenden Leistungen und davon abhängigen Beobachtungen auch eine sehr stark vereinfachte Darstellung von Abläufen, die in der Realität vielfältig miteinander verflochten sind.

Genauigkeit von Schülerbeurteilungen

In einer Reihe von Untersuchungen zur Notengebung ist belegt worden, dass Lehrerurteile häufig nicht den Anforderungen genügen, die man an professionelle Diagnosen stellen muss (zusammenfassend Tent, 1998). Aussagen zur Güte von Lehrerurteilen werden häufig dadurch gewonnen, dass man diese mit den Testleistungen der Schüler vergleicht. Neuere Untersuchungen haben gezeigt, dass Lehrkräfte im Durchschnitt betrachtet die Rangreihe der Leistungen innerhalb ihrer Klasse recht gut einschätzen können, auch wenn man dabei mit nicht unerheblichen Unterschieden zwischen Lehrern rechnen muss (Schrader, 1998).

Wenn man Lehrer bittet, die Leistungen ihrer Schüler in einem von ihnen bearbeiteten Test vorherzusagen, dann stellt man jedoch häufig fest, dass das allgemeine Niveau der Leistungsurteile (also die im Durchschnitt für die Klasse abgegebene Einschätzung) deutlich von dem der tatsächlichen Schülerleistungen abweicht. Dass Lehrkräfte Schwierigkeiten haben, das absolute Leistungsniveau ihrer Klasse zutreffend einzuschätzen, zeigt sich auch im Bereich der Notengebung: Während die Noten die innerhalb einer Klasse bestehenden Leistungsunterschiede meist recht zutreffend widerspiegeln, gibt es für vergleichbare Leistungen in verschiedenen Klassen oft ganz unterschiedliche Noten. Im Extremfall kann das bedeuten, dass Leistungen, die von der einen Lehrkraft als gut bewertet werden, von einer anderen als nicht ausreichend angesehen werden (vgl. Ingenkamp, 1997a). Auch wenn Lehrkräfte gut in der Lage sind, die Schüler ihrer eigenen Klasse nach ihren Leistungen in eine Rangordnung zu bringen, heißt das also nicht unbedingt, dass gleichen Noten in unterschiedlichen Klassen auch vergleichbare Leistungen zu Grunde liegen.

Lehrer orientieren sich vornehmlich an einem *klasseninternen* Bezugssystem, vergleichen also die Leistungen der Schüler mit denen anderer Schüler in der Klasse (soziale Bezugsnorm) und weniger mit einem absoluten Bezugsmaßstab. Eine sachgerechte und faire Beurteilung erfordert aber den Vergleich mit einem objektiven Kriterium, nämlich den vom Lehrplan geforderten Lehrzielen – unabhängig von den tatsächlich in der Klasse vorhandenen Leistungsunterschieden (kriteriale oder sachbezogene Bezugsnorm). Eine zutreffende Einschätzung des absoluten Leistungsstandes ist allerdings eine außerordentlich schwierige Aufgabe, die ohne den Einsatz von professionell entwickelten, am Lehrplan orientierten diagnostischen Instrumenten kaum zu erfüllen ist. Hinzu kommt, dass Lehrkräfte in inhaltlicher wie zeitlicher Hinsicht gewisse Spielräume bei der Realisierung der vom Lehrplan vorgegebenen Lehrziele haben und diese auch unterschiedlich ausnutzen, so dass verschiedene Klassen am Ende allein schon deshalb unterschiedliche Kenntnisse und Fertigkeiten besitzen können. Dies erschwert natürlich die Vergleichbarkeit. Schließlich ergibt sich für Lehrkräfte das psychologische Dilemma, dass sie mit der Beurteilung der Schülerleistungen immer auch ein Stück weit den Erfolg ihres eigenen Unterrichts bewerten.

Dass Lehrkräfte Schwierigkeiten haben, den Leistungsstand klassenübergreifend zutreffend einzuschätzen, ist vor allem deshalb problematisch, weil Lehrerurteile die Grundlage von Schulabschlüssen darstellen, die für das weitere Leben von erheblicher Bedeu-

tung sind. In manchen Ländern versucht man dieses Problem durch klassenübergreifende Leistungsüberprüfungen mit Hilfe von Tests zu lösen, die den vom Lehrplan geforderten Lehrstoff abprüfen. Allerdings sind auch die Ergebnisse einer einmaligen Testuntersuchung, die anderen Einschränkungen (erhöhter Erfolgsdruck, Tagesform, ungewohnte Situation usw.) unterliegen, nicht unbedingt das beste Kriterium für die Bewertung von Schülerleistungen. Im Vergleich dazu basieren Lehrerurteile in der Regel auf vollständigeren und repräsentativeren Informationen (die Leistungen des Schülers in einem längeren Zeitraum in unterschiedlichen schriftlichen und mündlichen Leistungssituationen). Dies spricht dafür, beide Informationsquellen zu kombinieren und die herkömmliche Urteilspraxis durch diagnostische Hilfsmittel abzusichern und zu ergänzen. Hieraus ergibt sich für die Wissenschaft die Forderung nach einer verstärkten Bereitstellung geeigneter diagnostischer Verfahren und für die Lehreraus- und -fortbildung die Forderung nach einer verbesserten Ausbildung in pädagogischer Diagnostik.

Leistungsbeurteilung im Rahmen des Unterrichts

Unterrichtsform und Lehrerrolle

Dass diagnostische Informationen das Handeln von Lehrkräften beeinflussen – wie dies in Abbildung 1 dargestellt wurde –, heißt natürlich nicht, dass sie nur für den ausschließlich lehrergesteuerten Unterricht wichtig wären. Unterrichtsformen, in denen Schüler selbständig und eigenverantwortlich tätig sind (Gruppen- und Projektarbeit oder andere Formen des offenen Unterrichts) stehen heute zu Recht sehr viel stärker im Vordergrund. Obwohl sich bei einem solchen Unterricht die traditionelle Lehrerrolle verändert, bedeutet das aber keinesfalls, dass die erzieherischen, didaktischen und organisatorischen Aufgaben von Lehrkräften an Bedeutung verlieren. Eine effektiver offener Unterricht erfordert vom Lehrer sogar in ganz besonderem Maße intensive vorbereitende, begleitende und nachbereitende Bemühungen: Es müssen geeignete Aufgabenstellungen ausgewählt und auf die Fähigkeiten, Vorkenntnisse und Bedürfnisse der Schüler zugeschnitten werden; die Schüler müssen bei der Bearbeitung dieser Aufgabenstellungen überwacht und unterstützt werden; und schließlich müssen die Arbeitsergebnisse aufgearbeitet, gesichert und vertieft werden. Die Forschung hat gezeigt, dass Lehrkräfte, die im Unterricht eine aktive Rolle übernehmen – und dies gilt auch für die wirkungsvolle Unterstützung von Lernvorgängen in offenen Unterrichtsphasen –, erfolgreicher sind. Darüber hinaus setzt die erfolgreiche Realisierung dieser auch im offenen Unterricht wichtigen didaktischen Maßnahmen bei den Lehrkräften ein hohes Maß an diagnostischer Sensibilität voraus.

Ziele und Erwartungen

Erwartungen, was Schüler einer bestimmten Altersstufe und Schulform im allgemeinen und die Schüler der eigenen Klasse im besonderen leisten können und wo ihre Stärken

und Schwächen liegen, beeinflussen die Unterrichtsgestaltung und das unterrichtliche Handeln der Lehrkraft. Dies betrifft vor allem die Auswahl und Schwierigkeit des Stoffs sowie das Tempo des Unterrichts. Erwartungen, die leicht über dem aktuellen Leistungsniveau der Schüler liegen und somit mit einiger Anstrengung gerade noch erfüllt werden können ("dosierte Diskrepanzen"), sind für den Lernerfolg und die Leistungsentwicklung besonders günstig. Ungünstig ist es dagegen, wenn die Erwartungen die tatsächliche Leistungsfähigkeit entweder stark übersteigen oder zu niedrig sind (Schrader & Helmke, 1987).

Viele Untersuchungen haben sich mit der Rolle von schülerspezifischen Erwartungen beschäftigt. Dass die unterschiedlichen Erwartungen, die Lehrkräfte gegenüber verschiedenen Schülern haben, deren Leistungen beeinflussen, ist seit langem bekannt. Schüler, von denen eine Leistungsverbesserung erwartet wird, verbessern sich häufig auch überzufällig (Erwartungs- oder Pygmalion-Effekt); umgekehrt – und in der Praxis sicher häufiger und pädagogisch problematischer – entwickeln sich die Leistungen von Schülern, von denen (zu) wenig erwartet wird, eher ungünstig. Der Grund für diese "sich-selbst-erfüllenden Prophezeiungen" ist, dass sich Lehrer ihren Erwartungen entsprechend gegenüber verschiedenen Schülern unterschiedlich verhalten und so ihre unterschiedlichen Erwartungen kommunizieren; die Schüler nehmen diese Verhaltensunterschiede wahr und ziehen daraus Rückschlüsse für ihre Selbsteinschätzung. Dies kann dazu führen, dass das Selbstkonzept der eigenen Fähigkeit, also das Vertrauen in die eigene Leistungsfähigkeit, und damit die Lernmotivation, beeinträchtigt wird.

Für erfolgreiche Lehrkräfte kennzeichnend ist die allgemeine Erwartung, dass die Schüler grundsätzlich in der Lage sind, die im Lehrplan vorgesehenen Lehrziele zu erreichen (Shuell, 1996). Für den Unterrichtserfolg ist nicht nur wichtig, dass die Erwartungen angemessen sind, sondern auch, dass sie vom Lehrer als verbindliche, im Unterricht zu realisierende Ziele betrachtet werden. Hierbei gibt es aber deutliche Unterschiede zwischen Lehrern: Manche von ihnen versuchen, ihre Zielvorstellungen konsequent zu realisieren und dabei insbesondere vorhandene Leistungsschwächen zu beseitigen (proaktiv); andere nehmen die vorhandenen Leistungsunterschiede als gegeben hin und stellen sich in ihrem Verhalten darauf ein, ohne sich gezielt um einen Ausgleich der Leistungsunterschiede zu bemühen (reaktiv); wiederum andere verstärken die Leistungsunterschiede sogar noch durch ihr Verhalten (überreaktiv).

Unterrichtsgestaltung und unterrichtliches Handeln

Erfolgreicher Unterricht erfordert eine Abstimmung auf die Lernvoraussetzungen der Schüler/innen. Man bezeichnet dies auch als adaptiven Unterricht oder spricht von Passung. Dazu ist eine enge Verknüpfung von diagnostischen Informationen und didaktisch-methodischen Unterrichtsmaßnahmen erforderlich (Ingenkamp, 1997a). Von adaptivem Unterricht spricht man vor allem dann, wenn es darum geht, *unterschiedliche* Lernvoraussetzungen zwischen Schülern, die es in jeder Klasse gibt, zu berücksichtigen. Darüber hinaus ist es aber auch wichtig, den Unterricht an die sich fortlaufend

verändernde Leistungsfähigkeit der Schüler anzupassen. Dies kann auf unterschiedlichen Ebenen erfolgen.

Wenn größere Unterrichtseinheiten betroffen sind, spricht man von *Makroadaptationen*. Je nachdem, wie die Diagnose der Lernvoraussetzungen ausfällt, werden unterschiedliche Maßnahmen eingesetzt. In solchen Fällen bietet es sich an, die Lernvoraussetzungen mit Hilfe von Tests und anderen diagnostischen Hilfsmitteln genauer abzuklären. Dies betrifft insbesondere die Ermittlung fehlender Vorkenntnisse, über deren Ausmaß und Beschaffenheit sich der Lehrer so ein besseres Bild machen kann. Vor allem lehrzielorientierte Tests, die eine differenzierte Diagnose auf der Feinzielebene ermöglichen, stellen ein geeignetes Hilfsmittel dar, um den nachfolgenden Unterricht genauer zu planen und auf die festgestellten Stärken und Schwächen der Klasse zuzuschneiden.

Auch während des Unterrichts muss für die ständige Anpassung einzelner Maßnahmen und Handlungen an das sich fortlaufend verändernde Wissen der Lernenden gesorgt werden. Man spricht bei diesen kurzfristig in der Situation erfolgenden Anpassungen von *Mikroadaptationen*. Während des Unterrichts kann der Lehrer nur in sehr begrenztem Umfang auf diagnostische Hilfsmittel, etwa Tests, zurückgreifen. Er ist stattdessen auf Beobachtungen und subjektive Einschätzungen angewiesen. Der Lehrer muss insbesondere ständig Lernvoraussetzungen abschätzen, Lernerfolge einschätzen und darüber hinaus Ursachen für Lernschwierigkeiten abklären.

Unterrichten wird gelegentlich als ein fortlaufendes Treffen von Entscheidungen beschrieben. Solche Entscheidungen können beim erfahrenen Lehrer mit einem relativ niedrigem Grad bewusster Aufmerksamkeit ablaufen. Dies fängt bereits vor dem Unterricht mit dem Aufstellen eines Unterrichtsplans an, der beim erfahrenen Lehrer meist nicht aus ausführlichen Aufzeichnungen besteht, sondern vielleicht nur einige wenige Stichpunkte umfasst. Entscheidend sind weniger die schriftlichen Aufzeichnungen, sondern der kognitive Plan, also das geistige Bild vom Verlauf einer Unterrichtsstunde. Die dabei zu treffenden Entscheidungen betreffen die Auswahl von Aufgaben, die Einleitung von Unterrichtsaktivitäten, den Übergang von einer Aktivität zur nächsten, das Einschleusen von Überprüfungsphasen, das Stellen von Fragen, das Aufrufen bestimmter Schüler, das Reagieren auf Antworten oder den Einschub zusätzlicher Erklärungen oder Übungen. Erfahrene Lehrer greifen dabei auf *Handlungsroutinen* zurück, die im Laufe der Zeit entwickelt wurden.

Untersuchungen deuten darauf hin, dass die diagnostische Kompetenz – und hier vor allem die Fähigkeit, Leistungsunterschiede einzuschätzen – eine Katalysatorfunktion für den Lern- und Unterrichtserfolg hat: Für sich allein genommen ist sie unwesentlich für den Lernerfolg der Schüler; ein hoher Lernerfolg ergibt sich erst dann, wenn hohe diagnostische Kompetenz mit bestimmten didaktischen Maßnahmen gekoppelt ist (vgl. Schrader & Helmke, 1987; Weinert & Lingelbach, 1995). Dies konnte empirisch für zwei Unterrichtsmerkmale belegt werden: den Einsatz von Strukturierungshilfen und von individueller fachlicher Unterstützung. Der Umstand, dass ein Lehrer häufig strukturierende Hinweise gibt (Zusammenhänge deutlich macht, Wichtiges hervorhebt usw.) ist für sich genommen nicht ausschlaggebend für den Lernerfolg der Schüler, sondern

nur dann, wenn die Lehrkraft gleichzeitig über eine überdurchschnittliche diagnostische Kompetenz verfügt. Dies deutet darauf hin, dass Strukturierungshilfen dann wirksam sind, wenn sie an die unterschiedlichen Lernvoraussetzungen der Schüler angepasst sind, wenn sie also in der richtigen Situation und gegenüber dem richtigen Schüler erfolgen. Ähnliches gilt für individuelle fachliche Unterstützungsmaßnahmen, also Hilfen und Hinweise, die der Lehrer einzelnen Schülern während Stillarbeitsphasen gibt.

Die Ergebnisse dieser Studien unterstreichen die Bedeutung einer engen Verflechtung von Diagnoseaktivitäten und Unterrichtsmaßnahmen. Das fortlaufende Diagnostizieren im Rahmen der Unterrichtstätigkeit könnte sich allerdings sogar ungünstig auswirken, wenn es nicht diskret erfolgt und sensibel gehandhabt wird, sondern wenn sich die Schüler dadurch ständigen Leistungsüberprüfungen und -bewertungen ausgesetzt fühlen. Wie aus Abbildung 1 hervorgeht, stützen sich Lehrkräfte bei ihren Leistungsurteilen in der Regel nicht nur auf die Leistungen in speziell dafür vorgesehenen Prüfungssituationen, sondern auch auf die während des Unterrichts gewonnenen Eindrücke.

In letzter Zeit wurde speziell am Unterricht in Deutschland kritisiert, dass Lern- und Leistungssituationen zu sehr vermischt würden (Weinert, 1999b): Während für den Erwerb eines anspruchsvollen Wissens selbständige, aber häufig auch mit Fehlern verbundene Lernaktivitäten wichtig seien, gehe es in Leistungssituationen um die fehlerfreie Demonstration der erworbenen Fähigkeiten. Wenn aber bereits in der Lernsituation Leistungsbewertungen erfolgten, sei dies der Lernmotivation und dem Lernen abträglich. Um diese Überlegungen noch stärker zur Geltung zu bringen, ließe sich das in Abbildung 1 dargestellte Modell entsprechend erweitern: Die Tendenz von Lehrkräften, in die Unterrichtsgestaltung Leistungsbewertungen einfließen zu lassen und somit unterrichtliche Lernsituationen mit Leistungssituationen zu verknüpfen, könnte als eine allgemeine pädagogische Orientierung [4] angesehen werden, die über einen zusätzlichen Pfad mit der Unterrichtsgestaltung [5] in Verbindung zu bringen wäre.

Verbesserung von Leistungsbeurteilungen

Der Nutzen objektiver Informationen

Das subjektive Lehrerurteil ist also in verschiedener Hinsicht korrektur- und ergänzungsbedürftig. Objektive, zuverlässige und gültige Informationen über Schülerleistungen, wie sie wissenschaftlich erprobte Testverfahren bieten, können dafür eine wertvolle Hilfe sein; sie dienen nicht zuletzt auch der Überprüfung und Korrektur der eigenen Erwartungen und Einschätzungen. *Normorientierte* Testverfahren können Lehrkräfte dabei unterstützen, die Leistungen ihrer Schüler über den Rahmen der eigenen Klasse hinaus zutreffend einzuschätzen. Schulleistungstests, bei denen der Schüler mit einer Alters- oder Jahrgangsnorm verglichen wird, geben dem Lehrer Informationen darüber, wie einzelne Schüler und die gesamte Klasse im Vergleich zum *gesamten Altersjahrgang* bzw. zur jeweiligen Klassenstufe einzustufen sind. Abgesehen davon,

dass Altersnormen irgendwann einmal veralten und der ständigen Aktualisierung bedürfen, haben die herkömmlichen normorientierten Schulleistungstests allerdings einen Nachteil: Sie sind meistens recht allgemeiner Natur, d. h. vornehmlich auf die Grobziele des Curriculums, aber nur wenig auf die Feinziele des Unterrichts abgestimmt (vgl. Ingenkamp, 1997a). Deshalb sind sie oft wenig instruktionssensitiv, d. h. nur begrenzt geeignet, den Erfolg einzelner Unterrichtseinheiten und -maßnahmen zu beurteilen.

Lehrzielorientierte Tests geben dagegen direkt Aufschluss darüber, inwieweit ein Schüler bestimmte Lehrziele beherrscht. Ihnen liegt eine sachbezogene Bezugsnorm zugrunde. Um feststellen zu können, was ein Schüler oder eine Schülerin im Hinblick auf ein vorgegebenes Lehrziel bereits kann bzw. noch nicht kann, ist eine differenzierte Aufschlüsselung von Leistungen erforderlich. In vielen Sachgebieten ist dies aber entweder nicht möglich oder zu aufwendig, wohingegen soziale Vergleiche praktisch immer möglich sind, so dass viele Lehrkräfte einen sozialen Vergleichsmaßstab für die Bewertung zu Grunde legen (Ingenkamp, 1997a). Sofern der Test die im Unterricht behandelten Lehrziele abdeckt, ist er *curricular valide*. Lehrer haben oft Schwierigkeiten, spezifische Leistungen differenziert zu beurteilen, insbesondere wenn es sich um weniger offensichtliche oder weniger vertraute Leistungsaspekte handelt, z.B. die Fähigkeit zum Lösen bestimmter Probleme. Testinformationen stellen dafür eine Hilfe dar.

Es ist seit langem bekannt, dass die Leistungsfähigkeit eines Lernenden nicht nur im aktuell erreichten Leistungsstand zum Ausdruck kommt, sondern auch in den bestehenden Entwicklungsmöglichkeiten, also dem durch Unterrichtsmaßnahmen oder bestimmte Hilfestellungen potentiell erreichbaren Leistungsniveau. Mit sogenannten *Lerntests* kann die Fähigkeit, vom Unterricht bzw. von Hilfen verschiedener Art zu profitieren, gemessen werden. Dabei wird im Grunde nichts weiter getan, als den Test zu zwei verschiedenen Zeitpunkten vorzugeben und eine Instruktions- oder Trainingsphase zwischen beiden Messungen einzuschieben. Entscheidend ist dann nicht, wieviel jemand bei der ersten oder zweiten Testung leistet, sondern wie effizient er die angebotenen Hilfen nutzen und sich dadurch verbessern kann. Ein solches Vorgehen, das der Logik der Förderdiagnostik verwandt ist, sensibilisiert die Lehrkraft für das Leistungspotential eines Schülers (vgl. etwa Schrader, 1997).

Testergebnisse können nicht nur die Leistungsbeurteilung der Schüler verbessern, sondern erlauben auch eine vom subjektiven Urteil des Lehrers unabhängige Evaluation des Unterrichtserfolgs. Dieser lässt sich strenggenommen erst dann beurteilen, wenn die vom Lehrer vorgefundenen Eingangsvoraussetzungen der Schüler in Rechnung gestellt werden (ob es sich z. B. um eine besonders leistungsstarke oder leistungsschwache Klasse mit günstigen oder ungünstigen Bildungsvoraussetzungen handelt).

Die Möglichkeit, unterschiedliche Eingangsbedingungen statistisch zu "bereinigen", bieten bislang vor allem die landesweit angelegten Leistungsstudien (wie LAU in Hamburg; QuaSUM in Brandenburg oder MARKUS in Rheinland-Pfalz). Zusätzlich zu den dort eingesetzten, auf den Lehrplan abgestimmten Tests werden z. B. mit Hilfe von Fragebögen verschiedene lernerleichternde oder -erschwerende Kontextmerkmale (Klassengröße, soziale Schicht, muttersprachliche Kompetenz usw.) erfasst. Anschlie-

ßend wird mit statistischen Techniken (Regressionsmethoden) berechnet, welche Leistungen bei bestimmten Ausprägungen der Kontextmerkmale zu erwarten sind. Mit Hilfe dieser *Erwartungswerte* kann man dann abschätzen, wie die Leistungen einzelner Klassen ausgesehen hätten, wenn der Kontext (z. B. die Klassengröße) in allen Klassen gleich gewesen wäre. Bei dieser Bereinigung bekommen Klassen mit ungünstigen Kontextbedingungen gewissermaßen einen Bonus, Klassen mit günstigen Ausgangsbedingungen dagegen einen Malus. Man bekommt auf diese Weise eine anschauliche Vorstellung davon, wie die Leistungsergebnisse vermutlich ausgefallen wären, wenn alle Klassen die gleichen Ausgangsbedingungen gehabt hätten. (Die statistisch bereinigten Ergebnisse haben den Charakter einer gut fundierten Hypothese. Um zu prüfen, ob die Ergebnisse in der realen Situation auch tatsächlich so ausfallen, wie durch die Bereinigung vorhergesagt, müsste man in allen Klassen tatsächlich die selben Ausgangsbedingungen herstellen, was in der Realsituation praktisch nicht möglich ist). Diese bereinigten Leistungsergebnisse werden den Lehrkräften ebenfalls zurückgemeldet. Welchen Wert dies für eine vergleichende Einschätzung des eigenen Unterrichtserfolgs hat, ist zur Zeit noch völlig ungeklärt und muss in entsprechenden Untersuchungen, etwa dem von uns gerade begonnenen Projekt WALZER, abgeklärt werden.

Eine Einschränkung besteht jedoch darin, dass bei Studien, die lediglich im nationalen Rahmen durchgeführt werden, das Curriculum und die ihm zugrundeliegenden Lehrziele meistens nicht problematisiert, geschweige denn in Frage gestellt werden. Auch bestimmte in den Unterrichtstraditionen eines Landes verwurzelte Einseitigkeiten (z. B. der stark auf die Vermittlung von Routineverfahren ausgerichtete deutsche Mathematikunterricht im Unterschied zum stärker auf Verständnis und Problemlösefähigkeit abzielenden Unterricht in Japan) können so nur schwer erkannt werden. International vergleichende Leistungsstudien wie TIMSS können insofern für bestimmte Leistungsaspekte (z. B. Problemlösefähigkeiten und Schlüsselqualifikationen), die im nationalen Rahmen nicht genügend beachtet werden, sensibilisieren. Sie bieten darüber hinaus Anhaltspunkte für eine vergleichende Einschätzung der Angemessenheit der Leistungsanforderungen und können so Einfluss auf die Bildung anspruchsvoller Unterrichtsziele haben.

Rückmeldung von Leistungen und ihre Nutzung

Dass Lehrer als Folge externer Evaluationen Rückmeldungen über die Leistungen ihrer Schüler und den Erfolg ihres Unterrichts erhalten, stellt zur Zeit sicher noch eine Ausnahmesituation dar, obwohl der Trend eindeutig in diese Richtung geht. Es ist daher auch noch wenig darüber bekannt, welchen Wert solche Informationen für die Lehrkräfte tatsächlich haben, wie man die Rückmeldungen zur Erzielung höchstmöglicher Verständlichkeit und Brauchbarkeit gestalten müsste, wie die Lehrkräfte damit umgehen und welche Wirkungen sie zeitigen.

In den seltensten Fällen kommt es allein schon deshalb zu Veränderungen, weil Lehrkräften ein schlechtes, durch ungünstige äußere Umstände nicht mehr erklärbares Leistungsergebnis ihrer Klasse mitgeteilt wird. Zwischen dem Erhalt einer solchen Leis-

tungsrückmeldung, einer dadurch veranlassten Verbesserung des Unterrichts und letztendlich effizienterem Lernen und besseren Leistungen liegt ein weiter Weg, der noch gezielter erforscht werden muss: Erforderlich ist unter anderem, dass die Rückmeldungen überhaupt zur Kenntnis genommen, verstanden und genauer analysiert werden. Dabei müssen genauere Vorstellungen über die Ursachen etwaiger Leistungsdefizite – auch was den eigenen Unterricht anbetrifft – entwickelt und überprüft werden; und es müssen dann tatsächlich auch geeignete Maßnahmen zur Verbesserung der Leistungen in Gang gesetzt werden. Von entscheidender Bedeutung ist vermutlich auch, wie in den Schulen mit derartigen Rückmeldungen umgegangen wird, welche Unterstützung Lehrkräfte bei der Nutzung und Umsetzung der vermittelten Informationen bekommen.

Leistungsstudien könnten nicht nur allgemein für die Verbesserung des Unterrichts, sondern auch direkt für die Verbesserung diagnostischer Kompetenzen genutzt werden. So könnte man z. B. die beteiligten Lehrkräfte einschätzen lassen, wie die Leistungen ihrer Schüler aussehen, und ihnen später Rückmeldung über die Angemessenheit dieser Einschätzungen geben. Auf diese Weise würde man die Lehrkräfte dazu veranlassen, ihre Erwartungen und Einschätzungen explizit zu machen und zu reflektieren, um so Aufschluss über verschiedene Beurteilungstendenzen zu gewinnen (z. B. ob man dazu neigt, Leistungen zu überschätzen oder zu unterschätzen) (vgl. dazu Wahl, Weinert und Huber, 1997). Man könnte Lehrkräfte so auch anregen, dieses Vorgehen dann später selbständig zu praktizieren.

Die Nutzung solcher Rückmeldungen für die Verbesserung des Unterrichts und der eigenen diagnostischen Kompetenzen verspricht vor allem dann Erfolg, wenn von verschiedenen Beteiligten gemeinsame Anstrengungen unternommen werden, die in ein schulisches Qualitätsmanagement eingebettet sind. Maßnahmen wie Fachkonferenzen, wechselseitige Unterrichtsbesuche, Team-Teaching oder Vergleichsarbeiten könnten dabei hilfreich sein. Unerlässliche Rahmenbedingungen dafür sind eine gute innerschulische Kommunikation und ein positives Evaluationsklima.

Fazit

Alltägliche Leistungsbeurteilungen im Klassenzimmer spielen eine wichtige Rolle für die Unterrichtsgestaltung und den Lernerfolg. Sie so vorzunehmen, dass das Lernen verbessert wird, ohne dass sich Schüler einer fortgesetzten Bewertung ausgesetzt fühlen, dürfte ein Schlüssel für den Unterrichtserfolg sein.

Die Beurteilung von Schülern ist eine schwierige und fehleranfällige Aufgabe für den Lehrer. Um sie nachhaltig zu verbessern, ist eine kontinuierliche Absicherung und Unterstützung der informellen Diagnoseleistungen des Lehrers nötig (Schrader, 1997). Dies kann durch objektive Leistungsmessungen mit Hilfe standardisierter Tests erfolgen. Der regelmäßige Einsatz derartiger Verfahren kann dem Lehrer gerade zu solchen Fragen Hinweise geben, die er selbst nur sehr schwer beurteilen kann: Wie die Leis-

tungen der eigenen Schüler im Vergleich zu anderen Klassen beziehungsweise zum gesamten Altersjahrgang aussehen, und wie gut bestimmte Lehrziele von den einzelnen Schülern erreicht wurden. Dadurch bekommt sein Urteil nicht nur eine breitere und sicherere Basis, sondern er kann durch den gezielten Vergleich seiner subjektiven Einschätzungen mit den erzielten Testergebnissen auch sein Urteilsvermögen verbessern. Dazu müsste allerdings die Fähigkeit und Bereitschaft, solche Methoden anzuwenden und zu nutzen, in der Lehreraus- und -fortbildung stärker als bislang gefördert werden. Schließlich müssten die Leistungsbeurteilungen einzelner Lehrkräfte sowohl auf Schulebene als auch zwischen einzelnen Schulen stärker koordiniert werden. Vergleichende Leistungsstudien könnten hierfür entscheidende Impulse geben.

KAPITEL 4

Bezugsnormen und schulische Leistungsbeurteilung

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 *Leistungsmessung*. 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 *Leistungsbeurteilung*. 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"¹ in vielen Untersuchungen bei Lehrern eingesetzt worden.

¹ Diese Aufgabe kann auch im Internet mit einer Ergebnismeldung bearbeitet werden.
Adresse:
[http://www.phil.uni-sb.de/~jakobs/paedpsych/rheinberg/kleine Beurteilungsaufgabe.htm](http://www.phil.uni-sb.de/~jakobs/paedpsych/rheinberg/kleine_Beurteilungsaufgabe.htm)