

*Josef Thonhauser, Universität Salzburg:*

**Die Leistungsbeurteilung verbessern heißt den Unterricht verbessern.**

Vortrag,

gehalten anlässlich der Tagung *Qualitätsinitiative Q-Hum* des BMBWK  
am 22. November 2006 in Wien

Meine sehr geehrten Damen und Herren! Liebe Kolleginnen, liebe Kollegen!

Ich bin eingeladen worden, anlässlich Ihrer Qualitätsinitiative ein Referat zu halten, bei dem die bekannten und weniger bekannten, aber dennoch gravierenden Probleme der Leistungsbeurteilung in der Schule möglichst umfassend angesprochen werden. Ich möchte das, sofern das in der verfügbaren Zeit überhaupt zu leisten ist, unter einer akzentuierenden Perspektive tun, die ich im Titel meines Vortrags zum Ausdruck brachte. Leistungsbeurteilung ist diesem zufolge nicht unabhängig von Unterricht insgesamt, gleichsam wie ein im Nachhinein anzuhängender Appendix zu sehen, sondern als konstituierende Komponente des Unterrichts. Die Qualität der Leistungsbeurteilung und die Qualität des Unterrichts insgesamt hängen daher notwendiger Weise zusammen.

Ich möchte zum Einstieg eine Lehrerstimme zitieren. Ein engagierter Lehrer leitete seinen Beitrag in einer Pädagogischen Zeitschrift sinngemäß wie folgt ein:

*Das Angenehmste ist das Unterrichten. Schon weniger angenehm ist das Üben. Am unangenehmsten ist das Prüfen. Denn dabei gilt die Schweigsamkeit des Lehrers als höchste Tugend.* (LUDERER, 1972, 119)

Es fällt auf, dass nur die Phasen der Information durch den Lehrer vollgültig dem Unterricht zugerechnet werden. Üben und Prüfen werden davon als pädagogisch weniger interessant deutlich abgesetzt.

Auf einem anderen Blatt stehen Befunde in der neueren einschlägigen Literatur (z.B. EDER, 2002, HELMKE, 2003, OSER & OELKERS, 2001), die sich so zusammenfassen lassen:

- Lehrer/innen mit beeindruckenden fachlichen Kompetenzen lassen sich, zumal (aber nicht nur) an höheren Schulen, relativ leicht finden.
- Es gibt, insbesondere (aber nicht nur) an Pflichtschulen, viele Beispiele für interessante Unterrichtsarrangements, z. T. mit aufwändigen Differenzierungsmaßnahmen.
- Im Vergleich dazu herrschen im Bereich der pädagogischen Diagnostik weithin erhebliche Mängel.

Die folgenden Ausführungen sind in fünf Abschnitte unterschiedlichen Umfangs gegliedert:

1. Für ein schwieriges Feld bräuchte es eine besondere Qualität der Praxis
2. Gewöhnung an begriffliche Unklarheiten und ihre Folgen
3. Neuere Verfahren der Leistungsbeurteilung (LB) zwischen ideologisch motivierter Abwehr und überzogenen Erwartungen
4. Ein kaum verzichtbarer, wenn auch nicht alles entscheidender Ansatz: die Qualität von Aufgaben
5. Die Qualität der Aufgaben(stellungen) im Rahmen neuerer Verfahren der LB

### **1. Für ein schwieriges Feld bräuchte es eine besondere Qualität der Praxis**

Leistungsbeurteilung ist ohne jeden Zweifel ein schwieriger Bereich der schulischen Arbeit. Das hat mehrere Gründe. Auf einige besonders gravierende möchte ich hier eingehen.

Traditionelle Orientierungen, die sowohl bei Lehrerinnen und Lehrern als auch bei Schülerinnen und Schülern sowie Eltern vorherrschen, führen dazu, dass die erste Assoziation zum Begriff *Leistungsbeurteilung* die (schlechte) Note ist. Damit gehen jedoch zwei entscheidende Verengungen des Blickwinkels einher: die auf die Note und die auf die selektive Funktion. Beides ist nicht naturgegeben, wird aber vielfach so empfunden. „Warum gehen wir in die Schule?“, lässt Frank Wedekind in seinem Stück *Frühlingserwachen* einen pubertierenden Schüler fragen. „Damit wir examiniert werden,“ lautet die Antwort. „Wozu werden wir examiniert?“ „Damit wir durchfallen!“ Damit schien der Weg geradlinig vorgezeichnet zu sein. Mit dem bei uns vor einigen Jahren eingeführten und von BildungspolitikerInnen häufig als Qualitätsmerkmal unseres Systems gepriesenen *Frühwarnsystem* wird zwar ein Rettungsanker ausgeworfen oder zumindest dafür gesorgt, dass sich Betroffene an die drohende Katastrophe gewöhnen können, die angeführte primäre Assoziation wird damit jedoch höchstens verstärkt.

Die Leistungsbeurteilung in unserem Schulsystem mutet den Lehrerinnen und Lehrern eine meines Erachtens schwer vereinbare Doppelfunktion zu: denselben Schülerinnen und Schülern gegenüber sowohl die Funktion des Förderns als auch des Auslesens zu erfüllen. (Ich bringe dazu einen sarkastisch anmutenden Vergleich: Diese Doppelfunktion der Lehrer/innen gleicht der Situation eines Chirurgen, der zugleich auch ein Bestattungsunternehmen betreibt. Fehlt der Erfolg in dem einen Beruf, ist er im anderen im anderen gegeben.) Wäre in dieser Situation eine verstärkte *Externalisierung der Leistungsbeurteilung* ein Ausweg? Mehr dazu unter 3.4.

Verschärfend kommt hinzu, dass die Beurteilung der Schülerinnen und Schüler häufig eine Zuteilung von Lebenschancen bedeutet, insofern mit ihr Aufstiegschancen im Schulsystem und darüber hinaus verbunden sind. Das macht diesen Bereich prinzipiell konfliktrichtig. Das ist mit ein Grund dafür, warum sich die Juristen des Bildungsministeriums ausführlich mit dieser Materie auseinandergesetzt haben und das Schulunterrichtsgesetz mit einer *Verordnung über die*

*Leistungsbeurteilung (LBVO)* ergänzt haben.<sup>1</sup> In dieser wird gleich einleitend auf ihre besondere Bedeutung hingewiesen:

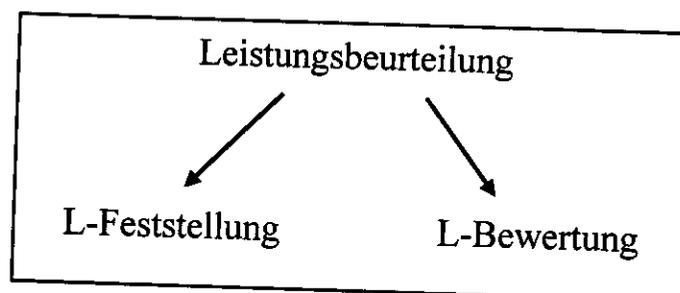
„Die Leistungsbeurteilung ist rechtlich gesehen ein Sachverständigengutachten. Wie bei jedem Gutachten muss der Beurteilung des Sachverhaltes die Erhebung (Feststellung) des Sachverhaltes vorangehen.“

Die Verordnung enthält in der Folge vor allem Bestimmungen darüber, was bei Beurteilungen, die unmittelbare oder mittelbare Konsequenzen für die Berechtigungen der Schülerinnen enthalten, zu beachten ist. Die zum Zwecke der Steuerung des Unterrichtsprozesses und die (Selbst-)Evaluation des Unterrichts vorgenommenen Überprüfungen des Unterrichtsertrags (in der LBVO, § 1, *Informationsfeststellung* genannt) werden in der Verordnung nicht behandelt. Damit trägt die LBVO faktisch zu einer gedanklichen Engführung von Leistungsbeurteilung, Notengebung und Berechtigungen bei. Das kommt auch in der in ihr verwendeten Begrifflichkeit zum Ausdruck:

Bei der Leistungsbeurteilung wird unterschieden zwischen der *Leistungsfeststellung* und der (eigentlichen) *Leistungsbeurteilung*. Damit behindert die LBVO die Vorstellung von einer Leistungsbeurteilung als einem integrativen Bestandteil von Unterricht, wiewohl sie an manchen Stellen Hinweise auf Zusammenhänge mit verschiedenen Qualitätsmerkmalen von Unterricht enthält (z. B. Vertrauensverhältnis zwischen Lehrerinnen/Lehrern und Schülerinnen /Schülern; Bedeutung einer zunehmenden Selbstbeurteilungskompetenz der Schüler/innen; Bedeutung von Prüfungssituationen für nicht betroffene Schüler/innen etc.).

Ich habe in meinem Titel bewusst eine von der LBVO abweichende Akzentuierung vorgenommen, die auch eine etwas andere Verwendung der Begriffe nach sich zieht:

*Leistungsbeurteilung* ist für mich der übergeordnete Begriff. Er umfasst die *Leistungsfeststellung* und die *Leistungsbewertung*. Je nach Situation wird dabei eine *formative*, den laufenden Lehr-/Lern-Prozess steuernde Funktion oder die *summative* Funktion in den Vordergrund gerückt, das heißt eine anhand der von Schülerinnen und Schülern nachzuweisenden Ergebnisse ihrer Lernprozesse (vorläufig) abschließenden Beurteilung.



Nach diesem Begriffsverständnis werden im Unterricht beinahe laufend *Leistungsfeststellungen* vorgenommen. Deren Basis sind *Beurteilungskategorien*, die das Augenmerk auf bestimmte Aspekte der Beurteilung (z.B. sprachlicher Aus-

<sup>1</sup> Ich folge hier der aktuellen Fassung.

druck oder inhaltliche Substanz einer Darstellung) lenken, und bei der *Maßstäbe* (richtig / falsch bzw. gut ↔ schlecht) angewandt werden. Die festgestellten Leistungen brauchen längst nicht in jedem Falle „bewertet“ zu werden.

Bei einer *Bewertung* kommen zu Kategorien und Maßstäben noch *Standards* hinzu, die in einem vergleichenden Prozess erlauben anzugeben, was eine erbrachte Leistung im gegebenen Kontext „wert“ ist. In unserem System bringen das die Noten zum Ausdruck. Lehrerinnen und Lehrer, die es genau nehmen, legen sich zumindest für die Schulstufe und den Schultyp solche Standards zu recht, die z. B. für ein *sehr gut* oder ein *genügend* erfüllt sein müssen. Die Vergleiche können sich allerdings an verschiedenen Parametern (Vergleichsdaten) orientieren. Die LBVO bietet diesbezüglich wenig Hilfe an.

### **Exkurs 1: Bezugsnormen**

In der einschlägigen Literatur werden in der Regel drei *Bezugsnormen* diskutiert:

- *Sozialnorm oder Kollektivnorm* (Durchschnitt und ggf. Streuung der Leistungen innerhalb einer Lerngruppe, z.B. einer Schulklasse);
- *Sachnorm bzw. curriculare oder kriterielle Norm* (Katalog der jeweils zu erreichenden, ausreichend klar definierten Lernziele [Kriterien]);
- *Individualnorm* (Leistungen der Lernenden, die mit ihren jeweils entsprechenden Ausgangsleistungen verglichen werden).

Die Anwendung dieser Bezugsnormen hat auf die Lernenden unterschiedliche Auswirkungen. Diese hängen zusammen mit der Leistungsfähigkeit und mit Persönlichkeitsmerkmalen. Auf der Basis zahlreicher Untersuchungen lassen einige zuverlässige allgemeine Aussagen treffen (vgl. z. B. ARNOLD, 1999, 74; RHEINBERG, 2001):

- Leistungsschwache Schüler/innen werden durch die *Sozialnorm* (über negative Einflüsse auf die Leistungsmotivation sowie das Selbstkonzept der eigenen Fähigkeiten) in ihrer Leistungsentwicklung in der Regel negativ beeinflusst. Denn sie müssen sich fortwährend mit einer Bezugsgruppe vergleichen (lassen), die leistungsfähiger ist als sie selbst es sind, und haben wenig Aussicht, von der ungünstigen relativen Position (am unteren Rand der Verteilung) weg zu kommen (WAGNER, 1999, 156).
- Die Anwendung der *Sachnorm* ist der Tendenz nach für alle Gruppen positiv, insbesondere dann, wenn differenzierte Aufgabenstellungen („*innere Differenzierung*“) allen Schülerinnen und Schülern die Chance auf richtige Lösungen bieten (grundlegend darüber HERBER, 1983).
- Die Anwendung der *Individualnorm*, für die eine „*innere Differenzierung*“ die logische Voraussetzung ist, wirkt sich ebenfalls auf alle Gruppen positiv aus. Sie setzt jedoch eine genaue Buchführung über die Leistungsentwicklung der einzelnen Schüler/innen voraus und erfordert daher einen erheblichen Aufwand.

Die Anwendung der *Individualnorm* stößt jedoch auch noch aus anderen Gründen auf Schwierigkeiten. Erstens: Von klein auf wird uns in vielen gesellschaftlichen Bereichen der soziale Vergleich nahe gebracht. Ein Grund dafür dürfte auch darin liegen, dass der soziale Vergleich fast immer möglich und vergleichsweise einfach ist. Die Menschen lernen seit der Antike, sich – in der „sportlichen“ Hoffnung auf *citius, altius, celerius* – mit anderen zu messen, und rechnen daher mit Leistungsvergleichen. Zweitens: Die Anwendung der *Individualnorm* setzt streng genommen die Eindimensionalität der beurteilten Leistungen voraus. Das heißt, sie ist nur dann anwendbar, wenn Leistungen in gleichen Leistungsbereichen (z. B. der

Wortschatz bei der Bebearbeitung vergleichbarer Themen; die Lesekompetenz für Sachtexte) aufeinander folgend erhoben (festgestellt) und beurteilt werden.

Die LBVO fordert eine *curriculare Bezugsnorm, also die Orientierung an lehrplankonformen Lernzielen*. In der Praxis dominiert hingegen weithin die *soziale Bezugsnorm*, was zu erheblichen Problemen führt. Man müsste die Folgen nur wirklich zu Ende denken. Ich möchte das an zwei sehr unterschiedlichen Beispielen demonstrieren.

Mit einer Analyse von TIMSS-Daten (EDER, 2002, S. 33) wurde rekonstruiert, wie groß der Punktezuwachs im Laufe eines Schuljahres im Durchschnitt sein müsste (vgl. Abbildung 1). Dann wurde festgestellt, wie weit die beste und die schlechteste Schulklasse auseinander liegen. Das Ergebnis: volle zwei Jahre. Zuletzt wurden die in der besten und der schlechtesten Klasse vergebenen Noten verglichen. Es zeigte sich kein Unterschied. Der Grund dafür ist die

Abb. 1: Erwarteter jährlicher Leistungszuwachs in Mathematik (TIMSS 1997)

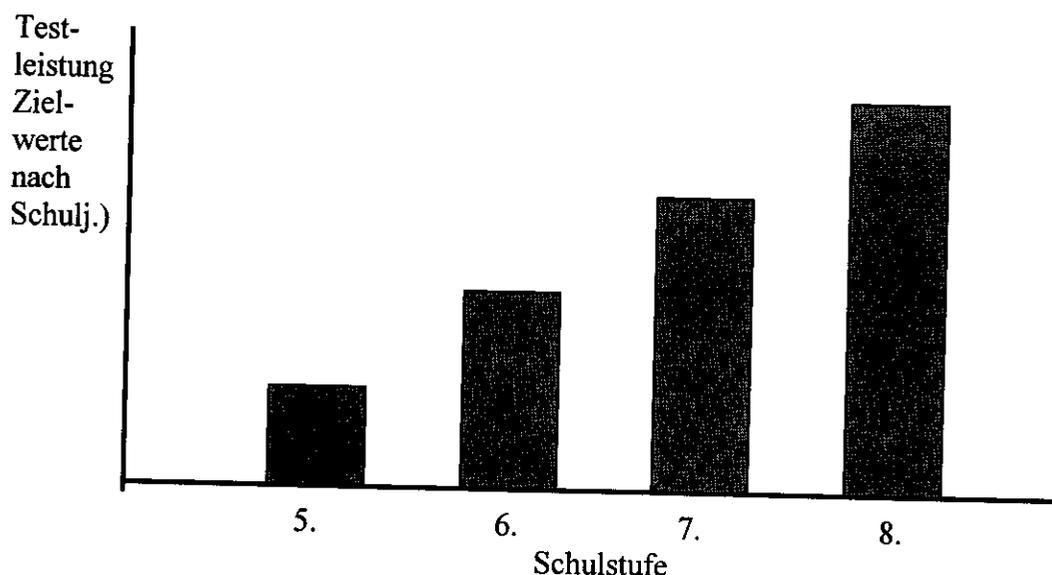
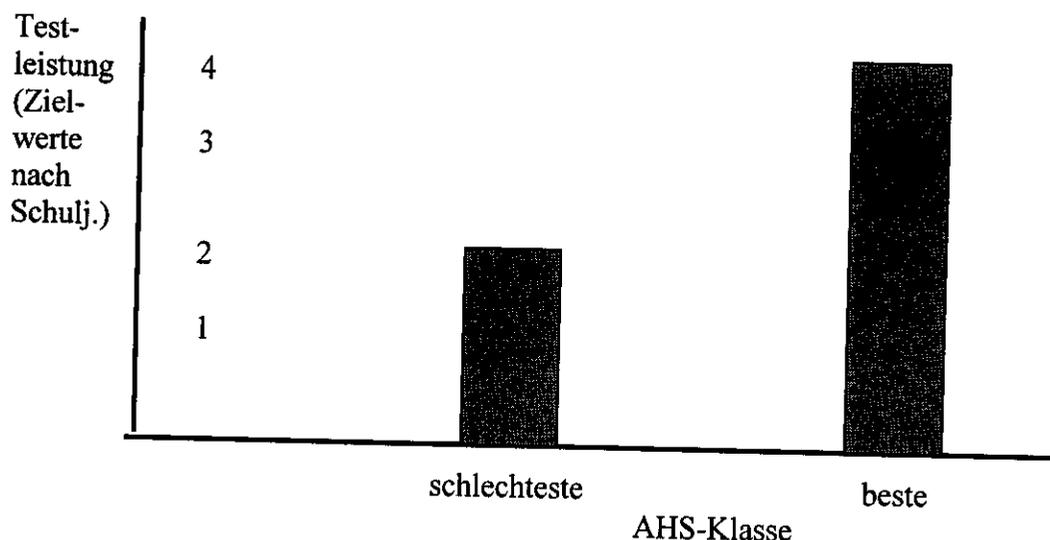


Abb. 2: Erwarteter jährlicher Leistungszuwachs in Mathematik (TIMSS 1997)



Anwendung der sozialen Bezugsnorm, die fast zwangsläufig dazu führt, dass die besten Schüler/innen einer Klasse mit *sehr gut*, die schlechtesten mit *nicht genügend* beurteilt werden.

Die soziale Bezugsnorm spielt aber auch in Situationen, wo man das zunächst nicht vermuten möchte, eine (im wahrsten Sinne des Wortes) entscheidende Rolle. Ein Beispiel: Seit diesem Jahr sind Aufnahmetests für Studierende der Medizin im Gespräch bzw. bereits im Einsatz. Dem begrenzten Platzangebot steht eine erheblich größere Zahl an Bewerberinnen und Bewerbern gegenüber. Von diesen kann nur ein bestimmter Anteil (z. B. die 300 Aspiranten mit den besten Testergebnissen) aufgenommen werden. Es ist aber nicht anzunehmen, dass diese 300 die gleichen Testwerte erreicht haben, wie die 300 Besten im nächsten Studienjahr, wo die Bewerber/innen insgesamt bessere oder schlechtere Voraussetzungen mitbringen könnten. Die Grenzfälle haben Glück, wenn sie sich mit insgesamt weniger guten Konkurrenten um die Aufnahme bewerben, sie haben Pech, wenn die Konkurrenten insgesamt bessere Testleistungen erbringen, obwohl ihre persönlichen Testwerte eine gleiche Prognose für ein erfolgreiches Medizinstudium zulassen müssten, wenn der Test eine gute Qualität hat. Die *berufsbildenden höheren Schulen* haben, solange es österreichweite standardisierte Aufnahmetests gegeben hat, sich (weniger allerdings den betroffenen Aspiranten) mit der Formel „*Aufnahmsprüfung bestanden, aber wegen Platzmangels abgewiesen*“ geholfen.

### ***Der beachtenswerte Gehalt der LBVO***

- Unter Berücksichtigung der zuvor getroffenen Einschränkungen soll hier ausdrücklich auf den beachtenswerten Gehalt der LBVO hingewiesen werden (vgl. dazu BUSCHMANN & THONHAUSER, 2000).
- Die Unterscheidung zwischen *Leistungsfeststellung* (anhand von Kategorien und Maßstäben) und Leistungsbeurteilung (in meiner Diktion *Leistungsbeurteilung*) (mit zusätzlichen Standards) begünstigt an sich – sowohl bei Lehrerinnen und Lehrern als auch bei Schülerinnen und Schülern sowie ggf. bei den Eltern – eine Einstellung, zunächst auf die erbrachte Leistung zu schauen und sich mit ihr auseinander zu setzen, und nicht das einzige Augenmerk gleich auf die Note zu richten. Die – unter einer juristischen Perspektive verständliche – Ausklammerung der *Informationsfeststellung*, wie das in der LBVO heißt, bringt die Verordnung – unter einem pädagogischen Aspekt betrachtet – jedoch wieder um einen Teil ihrer möglichen Wirkung.
- Aus der LBVO kann man eine Unterscheidung von unmittelbar veranlassten und nicht (unmittelbar) veranlassten Leistungsnachweisen herauslesen. Das wurde seinerzeit unter dem Reizwort *ständige Beobachtung* der Schüler/innen-Mitarbeit sehr kritisch diskutiert. Die LBVO will mit dieser Unterscheidung jedoch offensichtlich das Augenmerk der Lehrer/innen auf positive Beiträge der Schüler/innen im Unterricht lenken, mit denen sie bereits vor einer allfälligen Prüfung nachweisen, bestimmte Lernziele erreicht zu haben. Das heißt: positive Leistungen zählen immer, negative hingegen nur im Rahmen rechtmäßiger Prüfungen.
- Die LBVO verpflichtet zu einer *curricularen* (anstelle der sozialen) *Bezugsnorm*, deren Vorteile bereits oben beschrieben wurden.
- Nach der LBVO ist die *Leistungsbeurteilung* insofern ein *integrativer Bestandteil und nicht Anhang des Unterrichts*, als ausdrücklich festgehalten wird, sie sei „in den Unterricht so einzubauen, dass auch die übrigen Schüler

der Klasse aus der Leistungsfeststellung Nutzen ziehen können.“ (LBVO § 2 (6))

## 2. Gewöhnung an begriffliche Unklarheiten

Wir haben uns daran gewöhnt, von *Leistungsbeurteilung* zu sprechen. Trifft dieser Begriff den gegenständlichen Sachverhalt überhaupt?

In der *Physik* wird *Leistung als Arbeit in der Zeit* definiert. Die Formel für Arbeit lautet Kraft mal Weg. Für die *Pädagogik* lautet die Frage, was beurteilt werden soll: die *Anstrengung/der Fleiß* (analog zur Physik) *oder* die jeweils aktuelle *Fähigkeit (Kompetenz)* oder die unter günstigen subjektiven und objektiven Voraussetzungen in Zukunft erreichbare Kompetenz?

In der Praxis ist häufig wenn nicht regelmäßig zu beobachten, dass im Zweifelsfalle bei gleichen nachgewiesenen Fähigkeiten die Fleißigen eher die bessere Note erhalten als die Faulen. Das ist nicht selbstverständlich, wenn man fragt, wer bei gleichen aktuellen Fähigkeiten über die größeren Entwicklungsreserven verfügt. Die LBVO fordert hingegen in manchen Fächern ausdrücklich die Miteinbeziehung des *Leistungswillens* zugunsten der Schüler/innen (§ 11 (9)), die schwache Fähigkeiten zeigen. An diesen Beispielen wird evident, dass sich Leistungen bzw. Fähigkeiten mit Noten allein nicht gültig beschreiben lassen. Die LBVO unterstützt ferner eine heimliche Beschränkung der Leistungsbeurteilung auf kognitive und motorische Bereiche. Das steht im Widerspruch zu vielen allgemeinen Zielen von Schule, Schultypen und Fächern, in denen die affektive Komponente (Einstellungen, Werthaltungen) eine große Rolle spielen (vgl. POSCH & THONHAUSER, 1982). Die Schwierigkeiten ihrer Beurteilung sind unbestritten. Aber dürfen sie deswegen von Überlegungen einer Beurteilung gänzlich ausgeklammert bleiben? Spricht das nicht eher für eine verbale Ergänzung des Notenzeugnisses, mit denen die Entwicklung von Stärken und Schwächen so beschrieben werden können, dass für die Beurteilten Bezüge zu affektiven Zielen (z. B. umweltverträgliches Handeln; Solidarität mit (Lern-)Schwachen in der Klasse; Bereitschaft, eigene Wertvorstellungen zu begründen und sich für sie einzusetzen) deutlich werden?

## 3. Neuere Verfahren der Leistungsbeurteilung zwischen ideologisch motivierter Abwehr und überzogenen Erwartungen

Das weite Feld der Erziehung ist übersät mit Problemen. Für viele von ihnen ist eine Lösung dringlich. Oft – so scheint es zumindest – fehlt es an der Zeit für eine seriöse Entwicklung und Erprobung geeigneter Strategien. Das begünstigt die Hoffnung, ja mehr noch: eine fixe Zuversicht zu glauben, das probate Mittel zur Lösung eines Problems bereits gefunden zu haben, wenn mit einem neuen Begriff (für ein Prinzip, ein Verfahren oder ein Instrument) eine geeignete Maß-

nahme versprochen wird. Von dieser Tendenz sind auch Verfahren der Leistungsbeurteilung betroffen. Auf einige möchte ich im Folgenden eingehen.

### 3.1 Die Erlaubnis der Verwendung selbst angefertigter Arbeitsunterlagen

Ich habe für den gegebenen Zusammenhang drei Verfahren ausgewählt. Das erste betrifft die Erlaubnis, bei Prüfungen selbst angefertigte Arbeitsunterlagen zu verwenden. Meinen Klassenkollegen und mir wollte einst ein Lehrer weismachen, im fernen Japan würden Schularbeiten auf isolierten Tischen und in der Badehose geschrieben, um Schwindeln zu verhindern. Inwiefern ist diese Vision überhaupt sinnvoll? Wie gehen wir denn vor, wenn von uns in einem für uns selbst, aber auch für den Auftraggeber wichtigen Zusammenhang eine Leistung gefordert wird? Verzichten wir denn auf Hilfsmittel, die uns Ideen liefern oder bei der Ausführung sicherer machen?

In der Schule hat eine strenge Askese Tradition. Das geht zu Lasten der ökologischen Validität der Prüfungen.<sup>2</sup> Mit diesem Problem haben wir uns an der Universität anlässlich der Neugestaltung der Ersten Diplomprüfung beschäftigt. Über diese handelt der folgende Exkurs.

#### **Exkurs 2: Eine Prüfung unter dem Anspruch von Transparenz, Fairness und ökologischer Validität<sup>3</sup>**

Die Lehrveranstaltungen des ersten Studienabschnitts Pädagogik beanspruchen, den Studierenden Grundlagen wissenschaftlichen Denkens und Handelns zu vermitteln. Wie aber erfahren Lehrende und Studierende, in wieweit dies tatsächlich gelungen ist?

Die früher geltende Prüfungsordnung hat eine Lehr-/Lernkultur begünstigt, die mit einem Hürdenlauf verglichen werden kann. Die Hürden waren allerdings nur von Bedeutung, solange sie vor einem standen. Übersprungen, hatten sie – zumindest die meisten von ihnen – ihren Dienst einer geistigen Disziplinierung erfüllt und konnten gleichsam weggeräumt werden. Am Ziel angelangt, hielten die Studierenden die für ihr Fortkommen nötigen Scheine in der Hand. Ein Reflektieren über die Nachhaltigkeit der erworbenen Qualifikationen blieb der individuellen Initiative oder dem Zufall überlassen. Offiziell wird diese Form einer komplexen Prüfung (euphemistisch) als *kumulativ* bezeichnet.

Unsere neue Prüfungsordnung sieht am Ende des Ersten Studienabschnitts eine Überprüfung von Fähigkeiten vor, die jedenfalls für den Zweiten Studienabschnitt sowie für Diplomarbeit und Diplomprüfung, darüber hinaus aber auch für spezifische Berufe bedeutsam sind. Die Anforderungen dieser Prüfung und die Argumente dafür werden den Studierenden bereits zu Beginn des Studiums bekannt gemacht. Sie erfahren damit einerseits, was von ihnen erwartet wird, andererseits, wo sie einen Anspruch auf Unterstützung durch das Lehrangebot haben. Mit der Ersten Diplomprüfung ist daher die Hoffnung verbunden, dass von ihr sowohl auf das

---

<sup>2</sup> Unter Ökologischer Validität versteht man – einfach ausgedrückt – das Ausmaß, in dem eine Prüfungsleistung auch außerhalb bzw. nach der Schule von Bedeutung ist. Wenn die Bedingungen, unter denen eine Leistung zu erbringen ist, sich zwischen Schule und außerschulischem System stark unterscheiden, sinkt die ökologische Validität.

<sup>3</sup> Wen dieses konkrete Beispiel nicht interessiert, möge bei 3.2 auf Seite 13 weiterlesen.

Studium als auch auf die Lehre steuernde Impulse ausgehen, die den curricularen Zielen dienen und zu einer Steigerung der Qualität führen.

### **(a) *Transparenz der Leistungsanforderungen***

Im Zusammenhang mit Prüfungen herrscht an Schulen und Universitäten seit jeher eine rege Kommunikation. Aufgaben werden weitergereicht, hektische Bemühungen, sich mit Fragen wie „Was kommt denn zur Prüfung?“ oder „Was fragt er/sie denn?“ oder „Was kann man (ohne Risiko) auslassen?“ Gewissheit über die Anforderungen zu verschaffen, sind an der Tagesordnung. Wir haben daher Vorsorge getroffen, dass für die Studierenden von Beginn des Studiums an durch offizielle Informationen die größtmögliche Transparenz über die Anforderungen herrscht (siehe den Kasten auf der nächsten Seite). Damit hoffen wir, „außertourlicher“ Informationen überflüssig zu machen. Am schwarzen Brett finden alle Studienanfänger/innen die im untenstehenden Kasten enthaltenen Mitteilungen. Zu jedem Punkt gibt es weitere Erläuterungen.

### **(b) *Diagnose des aktuellen Leistungspotenzials***

In der einschlägigen Literatur wird zwischen *Leistungskapazität*, (aktuellem) *Leistungspotenzial*, der (konkreten) *effektiven Leistung* und der *beurteilten Leistung* (dem mit der Beurteilung erfassten Ausschnitt aus der effektiven Leistung) unterschieden.

- *Leistungskapazität*, als theoretisches Konstrukt, ist jenes Bündel von Leistungsmerkmalen, zu dem individuelle Entwicklungen bei gegebener genetischer Ausstattung unter günstigsten denkbaren Entwicklungsbedingungen führen. (Daraus folgt, dass in der Realität kein Individuum seine Leistungskapazität vollständig erreicht.)
- Mit *Leistungspotenzial*, ebenfalls ein theoretisches Konstrukt, wird jenes Bündel von Leistungsmerkmalen bezeichnet, das ein Individuum in einer individuell optimalen Prüfungssituation erreicht. (Eine solche Situation dürfte bestenfalls näherungsweise gegeben sein.)
- Unter *effektiver Leistung* versteht man die in einer Prüfungssituation vorgelegten Aufgabenlösungen, (Antworten auf Fragen, Kurzreferate, ausgefüllt Testhefte, Konstruktionen, Zeichnungen etc.).
- *Beurteilte Leistung* bezeichnet jenen Teil der effektiven Leistung, der von Beurteilenden wahrgenommen und für die Urteilsbildung herangezogen wird.

Bei der Ersten Diplomprüfung geht es vor allem um die Diagnose des jeweils aktuellen Leistungspotenzials. Dabei sind folgende Fragen wichtig:

- (a) Wie kann es gelingen, die effektive Leistung möglichst nahe an das aktuelle Leistungspotential heranzuführen?
- (b) Wie kann verhindert werden, dass von den Studierenden für relevant erachtete Aspekte der Leistung, die in der effektiven Leistung zum Ausdruck kommen, zu ihrem Nachteil in der beurteilten Leistung nicht berücksichtigt werden?

Die Antwort auf die erste Frage ist:

- eine möglichst langfristig gegebene Transparenz der Prüfungsanforderungen,
- die freie Wahl des Prüfungstermins,
- Prüfungsbedingungen, die ein leistungsförderliches, zumindest aber leistungsverträgliches Maß an Aufgeregtheit nicht überschreiten, und

## **Klausurarbeit zum Abschluss der Ersten Diplomprüfung\*)**

### **Allgemeine Information**

#### **Liebe Studierende !**

Die schriftliche Klausurarbeit ist die letzte Prüfung im ersten Studienabschnitt. Sie wird im EDV-Raum stattfinden, damit Sie die Möglichkeit haben, Ihre Arbeit mit jenem Medium zu verfassen, an das Sie sich im Studium bereits gewöhnt haben. Dabei sollen Sie vor allem das Wissen anwenden, das Sie im ersten Studienabschnitt erworben haben.

Grundlage der Aufgabenstellungen ist ein aktueller, ca. 15 Seiten umfassender Artikel aus einer wissenschaftlichen Zeitschrift. Ihre Aufgabe besteht aus Folgendem:

- Sorgfältige Lektüre (mit Vorteil unter Anwendung einer Lesetechnik, z.B. SQRRR)
- Verfassen eines verständlichen, gut lesbaren abstracts in deutscher Sprache (25 – 35 Zeilen, à ca. 70 Anschläge) und Ergänzung desselben mit 4 – 6 Schlüsselbegriffen
- Beantwortung von allgemeinen und spezifischen offenen und geschlossenen Fragen zum Text, wobei Sie jeweils drei aus fünf Fragen auswählen können
- Ergänzung des Literaturverzeichnisses um zwei Titel aus der einschlägigen aktuellen Literatur (mit Hilfe des Internets)
- Ergänzung der Angaben zur/zum (Erst-)Autor/in (mit Hilfe des Internets)
- Ergänzung des Artikels durch eine eigene, jedoch durch den Artikel begründbare technologische Empfehlung („Praktische Theorie“) für die pädagogische Praxis (Bildungspolitik, Erziehung in Familie, Schule oder anderen pädagogischen Institutionen etc.)
- Beurteilung des Artikels (im Stile eines Gutachtens) unter Berücksichtigung von Verständlichkeit der Darstellung; Aktualität der verwendeten Literatur (Einschätzung anhand des Literaturverzeichnisses); Angemessenheit der verwendeten Methoden; Bezug auf Theorien; praktische Bedeutung der Ergebnisse; selbstkritische Reflexion der Autorinnen und Autoren.

Sie dürfen bei der Bearbeitung der Aufgaben alle Materialien (Bücher, Skripten, etc.) verwenden, die sie in den Lehrveranstaltungen des ersten Abschnitts gesammelt haben. Sie haben außerdem Zugang zur Bibliothek. Personalressourcen (z.B. Kolleg/inn/en) stehen Ihnen allerdings nicht zur Verfügung. Insgesamt haben Sie zwei Stunden für die Lektüre und dann vier volle Stunden für die Bearbeitung der Aufgaben Zeit.

Wir betrachten die Ergebnisse der Klausur auch als wichtige Rückmeldung für die Leiter/innen der Lehrveranstaltungen des Ersten Studienabschnitts und hoffen daher mit Ihnen auf einen guten Erfolg.

\*) Zu dieser Information erhalten die Studierenden zwei konkrete Prüfungsbeispiele (Texte mit bibliographischen Angaben sowie Prüfungsaufgaben).

- Vermeidung von leistungshemmendem Stress, zum Beispiel dadurch, dass weder im Mikrobereich der einzelnen Aufgaben noch im Makrobereich der Prüfung insgesamt das Prinzip „alles oder nichts“ bzw. „jetzt oder nie“ herrscht, sondern einerseits zwischen zwei verschiedenen Texten gewählt werden kann und einzelne Aufgaben abge- wählt werden können, andererseits für eine negativ abgeschlossene Prüfung ausrei- chend Wiederholungsmöglichkeiten angeboten werden, wodurch das Schreckgespenst gravierender Nachteile (Zeitverlust, Versäumen von Anschlussprogrammen, Gefähr- dung der Studienbeihilfe etc.) einigermaßen gebannt werden dürfte.

Die Antwort auf die zweite Frage ist:

- die möglichst vollständige Entsprechung von curricularen Zielen (Anforderungen) und Beurteilungskategorien,
- ein klar festgelegtes und stabiles Beurteilungsverfahren und
- eine möglichst langfristige Transparenz dieses Verfahrens für die Studierenden.

### ***Fairness***

Das Merkmal der *Fairness* in dem hier gebrauchten, terminologisch engeren Sinn bezieht sich auf die Qualität des Studienangebots. Das heißt: Inwiefern eröffnen die laut Studienplan vor- gesehenen Lehrveranstaltungen den Studierenden seriöse Chancen, die an sie gestellten An- forderungen zu bestehen. Die Beachtung dieses Merkmals hat Konsequenzen in mehrere Richtungen:

- (a) Wenn sich die Studierenden den Anforderungen gewachsen zeigen, ist das zugleich ein positiver evaluativer Hinweis auf eine ausreichende Qualität des Lehrangebots.
- (b) Je mehr Studierende bei bestimmten Aufgaben Schwierigkeiten haben, desto dringen- der ist die (selbst-)kritische Auseinandersetzung der Lehrenden mit diesen Aufgaben, weil die Lehre offenbar nicht ausreicht, auf die für valide erachteten Prüfungsaufgaben ausreichend vorzubereiten.
- (c) Inwieweit wird mit der Ersten Diplomprüfung (alles) evaluiert, was im Ersten Stu- dienabschnitt gelehrt wurde?

An diesen Aussagen bzw. an der Frage wird deutlich, welche eminente Bedeutung den Leis- tungen der Studierenden bei der Evaluation der Lehre zukommt.

### ***Systeminterne prognostische sowie ökologische Validität***

Jede anspruchsvolle Prüfung muss sich am Merkmal der Validität beurteilen lassen. Wir ha- ben uns vorgenommen, nur solche Aufgaben zuzulassen, für die das Argument der internen Validität (Bedeutsamkeit für das weitere Studium, z.B. im Zusammenhang mit Seminar- oder Diplomarbeiten) und/oder der ökologischen Validität (Nähe der Prüfungsaufgaben zu As- sessment-Centers bei Bewerbungen und/oder erwartbaren Aufgaben in späteren professionel- len Situationen). Wir haben uns bisher allerdings auf logische Argumente formaler oder in- haltlicher Analogien beschränkt. Dafür zwei Beispiele:

- (a) Wir gehen von der Annahme aus, dass sowohl während des zweiten Studienabschnit- tes als auch in einem einschlägigen Beruf Studierende bzw. Absolventen häufig in die Situation kommen werden, aktuelle Fachinformationen für sich als abstract zu spei- chern und für andere zusammenfassend darzustellen.
- (b) Wir gehen ferner von der Annahme aus, dass bereits Seminararbeiten, jedenfalls aber Diplomarbeiten Beurteilungen von aktueller, wichtiger bzw. häufig zitierter Fachlite-

ratur enthalten, die einem Fachgutachten ähnlich sind. In vielen einschlägigen Berufen werden gleiche Aufgaben anstehen.

Der Umstand, dass wir uns große Mühe um eine möglichst hohe systeminterne prognostische und ökologische Validität der Prüfung geben, zeigt sich u. a. in dem durchgehend hohen taxonomischen Niveau unserer Aufgaben. Auf der anderen Seite ermöglichen bzw. fördern wir eine Entlastung des Arbeitsgedächtnisses von singulären Fakten, die man in der jeweiligen Situation rasch und zuverlässiger allgemein (z. B. aus dem Internet) oder fachspezifisch verfügbaren Informationssystemen (z. B. der Fachbereichsbibliothek) entnehmen kann. Die erfolgreichen Studierenden zeichnen sich durch geordnete Unterlagen aus, die zu allen erwartbaren Aufgaben Informationen enthalten. Während der Prüfung selbst wird von den Unterlagen unserer Erfahrung nach nur teilweise Gebrauch gemacht, weil das wohlüberlegte Zusammenstellen der Unterlagen positive Gedächtniseffekte zeitigt und ein Nachschauen vielfach überflüssig macht.

Unsere Erste Diplomprüfung steht damit in einer konzeptuellen Nähe zu einer „*Schummelzettel-Schularbeit*“. Wir haben diese irreführende Bezeichnung nicht erfunden (vgl. BMBWK, 2001). Im Gegenteil: Ich habe bereits vor geraumer Zeit einen zuständigen Beamten im BMBWK prophezeit, dass die Protagonisten dieser Schularbeitenform für den Gag dieser Bezeichnung büßen werden. Das ist prompt eingetreten, wie die in den Printmedien ausgetragene Diskussion gezeigt hat. Aber nun steht dieses Reizwort gemeinsam mit allen aus diffamierender Absicht oder Unerkenntnis erzeugten Missverständnissen einmal im Raum, und es erweist sich als schwierig, zu einer sachlichen Auseinandersetzung mit dem Prinzip dieses Verfahrens zurück zu finden.

Die traditionelle Prüfungskultur sah (sieht?) die Zulassung von dergleichen Hilfsmitteln oft nicht vor, sofern sie nicht überhaupt strikt untersagt sind. Die Begründung dürfte nicht zuletzt darin zu suchen sein, dass Prüfungen als Maßnahmen zur Disziplinierung (regelmäßiges Mitlernen, zum Teil bis zum auswendig Können) benutzt werden. Mit der Erlaubnis der Verwendung selbst angefertigter Arbeitsunterlagen würde eine Prüfung ungebührlich erleichtert. Dieses Argument zählt jedoch nur, wenn zwischen Qualität der Aufgaben und Nützlichkeit organisierbarer Hilfsmittel folgende regelhafte Zusammenhänge beachtet werden:

- (a) Je leichter und genauer eine konkrete Aufgabenstellung vorher erraten oder in Erfahrung gebracht werden kann (z. B. weil die Vorlieben der Prüfer/innen bekannt sind oder weil es eine undichte Stelle im System gibt), desto leichter und genauer sind Aufgabenlösungen oder Teile von ihnen bereits vor der Prüfung herstellbar.
- (b) Je höher das taxonomische Niveau einer Aufgabe, desto geringer ist der unmittelbare relative Nutzen, wenn man die formalen Anforderungen der Aufgabe kennt. (Z. B.: Das Wissen, dass bei der Erste Diplomprüfung die Beurteilung irgend eines aktuellen Fachartikels nach wissenschaftlichen Gesichtspunkten gefordert wird, hat zwar langfristig eine steuernde Wirkung auf die Vorbereitung und damit die Lernprozesse im Studium insgesamt, eröffnet jedoch nicht kurzfristig die Möglichkeit, sich die Aufgabe mit einem Schummelzettel entscheidend zu erleichtern. Fragestellungen vom Typ „Schreib alles auf, was du über den Pädagogischen Optimismus [oder die Hauptvertreter der konstruktivistischen Didaktik oder Poppers Wissenschaftstheorie] weißt, werden hingegen mit einer entsprechenden Unterlage entscheidend erleichtert.) Gegen Aufgabenstellungen der zuletzt genannten Art spricht auch noch Folgendes: Fähigkeiten auf einem höheren taxonomischen Niveau machen die Lernenden selbständiger in der Anwendung. Fähigkeiten auf einem niedrigen taxonomischen Niveau bleiben hin-

gegen häufig an den Kontext der Schule bzw. des Studiums gebunden („Schulwissen“).

### 3.2 Portfolios als Grundlage der Leistungsbeurteilung

#### ***Seit wann gibt es das Konzept?***

Wer sich bei uns in Österreich mit dem Thema Portfolio auseinandersetzt, hat allen Grund, bei Rupert VIERLINGERs „Direkter Leistungsvorlage“ (DLV) zu beginnen. Von Beginn seiner Tätigkeit als Lehrerbildner an hat der gleichermaßen profunde Kenner der praktischen Nöte schulischer Leistungsbeurteilung und der von der Erziehungswissenschaft erhobenen Ansprüche nach Alternativen zur Ziffernbeurteilung gesucht. Seit den 70er Jahren glaubt er, sie mit der DLV gefunden zu haben. Unter dem Motto „*Leistung spricht für sich selbst*“ hat er seine Ideen und Erfahrungen zusammenfassend dargelegt (VIERLINGER, 1999).

Mit einer Leistungsbeurteilung auf der Grundlage von Portfolios soll Folgendes erreicht werden:

- die Informationsarmut der Ziffernzensur und das – vor allem leistungsschwächere Schüler/innen betreffende – Elend der sozialen Bezugsnorm soll überwunden werden;
- die Angst vor Misserfolg bzw. die Orientierung an den Schwächen soll auch bei leistungsschwächeren Schülerinnen und Schülern der Hoffnung auf den individuell möglichen Erfolg und ggf. berechtigtem Stolz auf eigene Leistungen Platz machen;
- die misserfolgsbedingten Belastungen der Beziehungen zwischen Schülerinnen und Schülern, Lehrerinnen und Lehrern bzw. Eltern kann zumindest gemindert werden.

Zudem komme man mit einer DLV realen Situationen, wie sie bei Aufnahmeprüfungen, Abitur-Prüfungen oder Bewerbungen im Berufsleben gegeben seien, wesentlich näher. Denn dort interessieren auch in erster Linie Leistungen und nicht die Beurteilung von Leistungen (Noten).

VIERLINGERs Vorschlag war zunächst eher vorsichtig aufgenommen worden. Seit Beginn der 90er Jahre aber kam Unterstützung aus den USA. (Für einen Überblick vgl. LISSMANN, 2000).

#### ***Mögliche Funktionen eines Portfolios***

Ihren Funktionen nach können zumindest drei Gruppen von Portfolios unterschieden werden:

- Dokumentationen über Entwicklungen unter bestimmten (mehr oder minder kontrollierten) Bedingungen (Zusammenhang: *Prozessdiagnostik*);
- Dokumentationen über einmal erbrachte bzw. aktuelle Leistungen (Zusammenhang: *Produkt diagnostik*);

- Berichte über die (aktuelle) Leistungsfähigkeit anhand der „besten Leistungen“ (Zusammenhang: Beleg für aktuelle Kompetenzen und für die [Selbst-]Einschätzung des Leistungspotentials).

### ***Erwartungen an eine Portfolio-Beurteilung***

- Die Leistungen werden nicht, wie bei der herkömmlichen Beurteilung in der Schule (als Schularbeit, schriftliche Tests oder Antworten bei einer mündlichen Prüfung) an die Beurteiler/innen abgeliefert und gleichsam gegen das Urteil eingetauscht, sondern verbleiben als geistiges Eigentum, das in einem anderen Zusammenhang legitimer Weise wieder verwendet werden darf, beim Urheber.
- Das Urteil interessiert nicht nur unter summativen, sondern auch unter formativen Aspekten, d.h., insofern es (auch) Hinweise auf positive oder negative, aber veränderbare aktuelle Entwicklungsbedingungen bzw. künftige Entwicklungsmöglichkeiten enthält.
- Die Schüler/innen erfahren ein größeres Zutrauen in ihre Selbständigkeit, indem sie bei der Terminwahl und – in Grenzen – auch bei den Leistungsinhalten mitbestimmen dürfen.
- Beurteilte werden angehalten, sich selbst ein Urteil über die Qualität ihrer Leistungen zu bilden, bevor sie diese einer Fremdbeurteilung zugänglich machen (Förderung der Selbstbeurteilungskompetenz).
- Die Möglichkeit der Orientierung an den Stärken erhöht die Chancen auf die Entwicklung eines realistischen positiven Selbstkonzepts der eigenen Leistungsfähigkeit.

Die bisherigen Erfahrungen rechtfertigen einen gewissen Optimismus hinsichtlich positiver Effekte der Arbeit mit Portfolios. Dennoch kann diese allein nicht mit einem Schlag die vielschichtigen Probleme der Leistungsbeurteilung lösen. Ausreichende Objektivität und Validität der Beurteilung, der Chancengleich unter den Schülerinnen und Schülern bei der Wahl der Themen, die Begründung verpflichtender Leistungsansprüche an alle Schüler/innen, die nicht abgewählt werden dürfen, und last but not least umfassende Beurteilungskompetenzen und ihnen entsprechende Beurteilungskonzepte der Lehrer/innen stellen sich mit der Verwendung von Portfolios nicht automatisch ein.

Das Argument, das ich nach einem Vortrag an einer Pädagogischen Akademie zu hören bekam, mit Portfolios würden die Leistungsanforderungen generell auf ein unvertretbar niedriges Niveau gedrückt, beruht hingegen entweder auf einem Mangel an Erfahrungen mit diesem Verfahren oder verbirgt eine ideologisch motivierte Abwehr konkret dieser innovativen Form oder von Innovationen ganz allgemein.

### 3.3 Zweiphasen- oder Zweistufen-Schularbeit

Bei *Zweiphasen- oder Zweistufenarbeiten* erhalten die Schüler/innen ihre Arbeiten, nachdem sie von der Lehrerin bzw. dem Lehrer durchgesehen und allenfalls mit Hinweisen versehen, nicht unbedingt aber korrigiert und beurteilt worden sind, nochmals zurück, um daran noch kurze Zeit an ihr weiter zu arbeiten. Die Schüler/innen werden damit angeregt, sich mit ihren Texten (als ihren fachspezifischen Leistungen) nochmals intensiv zu beschäftigen und – ggf. durchaus unter Nutzung personeller Ressourcen, z.B. der Eltern – Verbesserungen anzubringen. Dabei ist auch ein gewisses Ausmaß an *Selbstbeurteilungskompetenz* erforderlich (BLÜML, 1994). Prinzipiell sind mehrere Varianten denkbar, die in der folgenden Übersicht zusammengestellt sind.

Abb. 3: Varianten der Zweiphasen oder Zweistufen-Schularbeit

Variante	Phase 1	Zwischenphase	Phase 2	Erläuterungen
a	S/S schreibt (1 Stunde oder mehr)	L/L korrigiert, kommentiert und benotet	S/S stellt eine verbesserte Fassung her	S/S lernt (auch) aus Fehlern; L/L benotet die verbesserte Fassung (zusätzlich)
b	S/S schreibt (1 Stunde oder mehr)	L/L korrigiert bzw. kommentiert (ohne Note)	S/S setzt Arbeit fort	Schreibberatung durch L/L; Anregung zur Selbstbeurteilung
c	S/S schreibt (1 Stunde oder mehr)	L/L benotet (ohne Korrekturzeichen oder Kommentare)	S/S setzt die Arbeit fort	Anregung zum Nachvollzug der L/L-Beurteilung zur Selbstbeurteilung und Verbesserung
d	S/S schreibt (1 Stunde oder mehr)	L/L verwahrt die Erstfassung lediglich; S/S beschäftigt sich ggf. gedanklich mit der Arbeit	S/S setzt die Arbeit fort	Selbstbeurteilung durch S/S und Verbesserung nach einem zeitlichen Abstand; L/L korrigiert und benotet die Endfassung

Gegen die Zweistufen-Schularbeit sind anfangs Bedenken vorgebracht worden, z. B. wegen Überschreitung der höchstzulässigen Zeit für Schularbeiten. Später wurden sie als „Schulversuch“ zugelassen, um Erfahrungen zu sammeln. Das Prinzip, eine Arbeit in mehreren Phasen zu betreuen und in jeder Phase formativ (das heißt, zum Zwecke ihrer Verbesserung, bevor sie endgültig abgegeben wird) zu beurteilen, ist in anderen Zusammenhängen, z. B. bei Fachbereichsarbeiten, nichts Außergewöhnliches, wiewohl sich auch dagegen Bedenken konstruieren lassen, weil es schwer werden könne, die Eigenanteile der Schülerleistung zu beurteilen.

### 3.4 Externalisierung der Leistungsbeurteilung zur Entflechtung von Förder- und Auslesefunktion

Die schulische Leistungsbeurteilung ist in dem bei uns geltendem System, wie bereits erwähnt, durch ein Problem belastet: Die Lehrer/innen haben ein und demselben Schüler gegenüber zwei miteinander schwer vereinbare Funktionen zu erfüllen: *Qualifikation* (im Sinne einer umfassenden Entwicklungsförderung) und *Selektion* (im Sinne des zeitweiligen oder dauerhaften Ausschlusses von bestimmten Unterstützungsangeboten, z.B. einer weiterführenden Schule oder einem qualifizierten Beruf). Nimmt man noch hinzu, dass Schülerleistungen ein wichtiger Indikator für die Qualität des Unterrichts sind, befindet man sich in einer klassischen Trilemma-Situation:

(a) Der Gesetzgeber verpflichtet die Lehrer/innen einerseits gleichsam zu Optimismus bis zur letzten Sekunde. Denn keinem Schüler darf vor dem Ende einer Beurteilungsperiode die Aussicht auf eine positive Beurteilung abgesprochen werden.

(b) Andererseits erwartet jedoch die Gesellschaft, dass Schülerinnen und Schüler, die für bestimmte schulische Karrieren als ungeeignet erscheinen, beizeiten erkannt und ausgeschieden werden, damit ihnen die Enttäuschung und der Gesellschaft unnütze Kosten erspart bleiben mögen.

(c) Lehrer/innen, die auch sich selbst als eine Ursache schlechter Schülerleistungen sehen, werden zunächst überlegen, wie sie ihren Unterricht (die Qualifikationsfunktion) für diese Schüler/innen verbessern könnten. Lehrer/innen, die sich von schlechten Schülerleistungen bestraft fühlen oder in ihnen eine Gefahr für die Beurteilung ihres Unterrichts sehen, werden entweder rasch zur Selektionsfunktion neigen, weil sie damit die (scheinbar Allein-) Schuldigen markieren, oder eine günstige Benotung vornehmen, um damit die (scheinbar hohe) Qualität ihres Unterricht zu belegen.

Gibt es Möglichkeiten, Lehrerinnen und Lehrer aus diesem Trilemma zu befreien? Eine bestünde in der Trennung der beiden Funktionen: Lehrer/innen, die unterrichten, wären nicht mehr für Selektionsentscheidungen zuständig. Schüler/innen müssten an entscheidenden Punkten ihrer Karriere ihren (mit Hilfe ihrer Lehrer/innen erreichten) Leistungsstand im Sinne einer *externalen Leistungsbeurteilung* vor fremden Instanzen nachweisen.

In der Diskussion wird Externalisierung der Leistungsbeurteilung für gewöhnlich mit dem Einsatz zentral entwickelter und zentral durchgeführter Tests gleichgesetzt, denen Vor- und Nachteile nachgesagt werden. Die Vorteile scheinen auf der Hand zu liegen: Die Beurteilung erfolgt anhand von Aufgaben, die für gleiche Ansprüche stehen; es gelten gleiche Maßstäbe und gleiche Standards, die Beurteilung ist daher per definitionem *objektiv* und *zuverlässig* und erleichtert Leistungsvergleiche auf allen Ebenen. Als Nachteile werden die zu befürchtende verengende Ausrichtung des Unterrichts auf die Testaufgaben, der tendenzielle Verzicht auf individuell bzw. lokal bedeutsame Lernangebote oder aber

die Gefahr des unlauteren Umgangs mit den (kaum geheim zu haltenden) Tests genannt.

Das Für und Wider betreffend *Standardisierte Tests* verdient angesichts der Bedeutung von TIMSS und PISA und der zu erwartenden *Bildungsstandards* eigens behandelt zu werden. Das kann hier nicht geleistet werden. Hier sei jedoch darauf hingewiesen, dass eine Externalisierung von Leistungsbeurteilung nicht auf Tests beschränkt ist. Wenn Lehrer/innen in Teams eine Konkretisierung des Rahmenlehrplans vornehmen, wie es z.B. das Konzept des Lehrplans '99 innerhalb einer Schule vorgesehen hatte, was aber auch in mit einander vernetzten Schulen denkbar wäre, ist es nur ein kleiner und durchaus konsequenter Schritt zur wechselseitigen Unterstützung in der Funktion der Leistungsbeurteilung. Die Entflechtung der widersprüchlichen Funktionen innerhalb eines Teams, in dem alle sowohl qualifizierende als auch selektierende Funktionen behalten würden, nur nicht gegenüber ein und denselben Schülerinnen und Schülern ausüben müssten, sollte als Preis attraktiv genug sein, um sich an die Erprobung solcher Formen einer *kollegialen Externalisierung* zu wagen.

### 3.5 Schüler/innen-Selbstbeurteilung: Warnung vor einer Zerrform

Schon mehrfach wurde in dieser Arbeit die Wichtigkeit der *Selbstbeurteilungskompetenz* der Schüler/innen angesprochen (vgl. auch LBVO, Anmerkung 1), dem Hinweise auf geeignete Maßnahmen folgten. Hier soll vor einer in der Praxis hin und wieder anzutreffenden Zerrform gewarnt werden: der Selbstbenotung. Diese Praxis übersieht, dass die Benotung (in unserem System) – nimmt man die Mitteilung darüber (Rückmeldung) aus – der letzte Schritt einer Leistungsbeurteilung ist. Es bringt daher wenig, Schüler/innen – vermeintlich zur Förderung ihrer Selbständigkeit – mit diesem Schritt beginnen zu lassen. Die Förderung der Selbstbeurteilungskompetenz muss grundlegender ansetzen.

### **Exkurs 3: Ein Grundmodell des Lehrens und Lernens in einem aufgabenorientierten Unterricht**

Viele, wahrscheinlich sogar die meisten Unterrichtseinheiten, zumindest Teile davon, können als aufgabenorientiert angesehen werden. Die Schüler/innen lernen anhand mehr oder minder explizit gestellter Aufgaben. (Vgl. dazu den langjährigen Bestseller von GRELL & GRELL mit dem provokanten, teilweise irreführenden Titel *Unterrichtsrezepte*, in dessen Zentrum *Lernaufgaben* stehen. Dieses Konzept wurde neuerdings von SCHAFFENRATH & WIESER, 2005, mit Blick auf die Berufsschulen aufgegriffen.)

In einem aufgabenorientierten Unterricht sind drei entscheidende Fragen zu stellen:

- In wieweit ist den Schülerinnen und Schülern die Aufgabe(nstellung) klar?
- In wieweit kennen die Schüler/innen erfolversprechende Lösungswege?

- In wieweit können die S/S ihre Lösungen oder Lösungsversuche selbst beurteilen?

Abb. 4: Grundmodell des Lernens und Lehrens in einem aufgabenorientierten Unterricht

<b>Ergebnisse der Lernprozesse (Produktebene Lernen)</b>	<b>Aktivitäten der Lernenden (Prozessebene Lernen)</b>	<b>Aktivitäten der Lehrenden (Prozessebene Lehren)</b>
<b>Bewusstsein über die Aufgabe (Problembewusstsein)</b>	<b>sich die Aufgabe bewusst machen / den Problemraum definieren</b> <ul style="list-style-type: none"> <li>• Ausgangszustand ↓ transformatorische Operationen</li> <li>• ↓ Zielzustand</li> </ul>	<b>Komponenten der Aufgaben / Problemraumkomponenten</b> vermitteln (insbesondere Ziele) Rahmenbedingungen des Lernens arrangieren (zeitliche, räumliche, materielle, soziale)
<b>Lösungen / Lösungsversuche (Produkte bzw. Verhalten):</b> = Indikatoren für veränderte psychische / physische Dispositionen, die damit beobachtbar gemacht werden	<b>Aufgaben / Probleme bearbeiten</b>	<b>Hilfen für das Lösen der Aufgaben bereitstellen:</b> <ul style="list-style-type: none"> <li>• inhaltliche / lösungsbezogene</li> <li>• formale / lösungswegbezogene</li> </ul>
<b>Urteile der Lernenden über die Qualität ihrer Lösungen / Lösungsversuche</b>	<b>Lösungen / Lösungsversuche selbst beurteilen (evaluieren)</b>	<b>Kategorien und Maßstäbe, (ggf., aber nicht unbedingt Standards) für die (Selbst-) Beurteilung der Lösungen vermitteln; Lösungen beurteilen</b>

Diesem Modell zufolge sind Urteile der Lernenden über die Qualität ihrer Lösungen / Lösungsversuche ein konstitutiver Bestandteil von Lehr-/Lernprozessen. Würde ein Lernender diesbezüglich keinen Kompetenzzuwachs verzeichnen und sich selber keinerlei Rückmeldungen geben, sondern immer vollkommen abhängig bleiben von einer Beurteilung der Lehrer/innen, hätte er im Endeffekt nichts gelernt. Das zu vermeiden helfen ihm Kategorien und Maßstäbe, nicht aber die vordergründige Anwendung von Standards, über die man zu Noten kommt.

#### **4. Ein, wenn auch nicht alles entscheidender Ansatz: die Qualität von Aufgaben**

Eine zentrale Frage der Leistungsbeurteilung richtet sich auf die Qualität von Aufgaben, das heißt, auf die Qualitätsmerkmale und ihre Ausprägungen. Die Auseinandersetzung damit ist nicht nur für die Leistungsbeurteilung (mit der LBVO gesprochen: im Sinne einer *Leistungsfeststellung* **und** einer *Informationsfeststellung*) von Bedeutung. Sie betrifft prinzipiell auch Aufgaben, die zum Lernen anregen (*Lernaufgaben*), auch wenn diese weniger dramatische Folgen zu haben scheinen.

Die einschlägige Forschung hat die folgenden Merkmale für wichtig befunden:

- die Verständlichkeit der Aufgabe für die Schüler/innen (Präzision)
- die Sinnhaftigkeit (Bedeutsamkeit) der Aufgabe – auch für die Schüler/innen
- die Fairness
- das Anspruchsniveau (taxonomische Niveau)
- den Schwierigkeitsgrad (die Lösbarkeit)
- die Trennschärfe
- die inhaltliche, prognostische und *ökologische* Validität
- das Merkmal Offenheit versus Geschlossenheit
- den Spielraum bei Annahme und Bearbeitung

Die *Verständlichkeit* der Aufgaben für die Schüler/innen wird – abgesehen von fehlerhaften Formulierungen – oftmals überschätzt. Das zeigte u. a. die Auswertung der Mathematik-Aufgaben bei PISA. Es ergab sich ein erheblicher Zusammenhang zwischen mathematischen Kompetenzen und der Lesekompetenz. Das heißt: ohne eine entsprechende Lesekompetenz kamen die Probanden gar nicht dazu, ihre Fähigkeiten in Mathematik zu beweisen.

Die *Sinnhaftigkeit* einer Aufgabe bezieht sich auf die Dimension der *Bedeutsamkeit* des Unterrichts insgesamt. Beobachtungen ergaben häufig, dass Lehrer/innen ein Thema oder auch Aufgabenstellungen gut begründen konnten, es aber verabsäumten, die Begründung (Bedeutsamkeit) den Schülerinnen und Schülern zu vermitteln. Was objektiv (lehrplanbezogen) bedeutsam ist, ist es nicht automatisch auch subjektiv für die Schüler/innen. Es bedarf der Vermittlung. Eine umfassende Studie über die *Qualität von Rückmeldungen* über Aufgabenlösungen hat bestätigt, dass diese die Schüler/innen nur dann wirklich erreichen, wenn sie die Aufgabenstellung als bedeutsam ansehen (JACOBS, www).

Mit *Fairness* (in diesem Zusammenhang) ist gemeint, Schüler/innen müssen durch den Unterricht eine faire Chance erhalten haben, eine Aufgabe zu lösen, und dürfen dabei nicht auf außerschulische Ressourcen angewiesen sein.

Auf die Bedeutung des Anspruchsniveaus von Aufgaben (des taxonomischen Niveaus, wie es Benjamin S. BLOOM, 1956, konzipiert hat) ist bereits weiter oben, S. 12, hingewiesen worden. In der schulischen Praxis spielen nach wie vor Aufgabenstellungen, die mit auswendig gelerntem, im Extremfall kaum verstandenem Wissen bewältigt werden können, eine große Rolle. Solche Aufgabenstellungen haben eine geringe (ökologische) Validität. Außerhalb des schulischen Prüfungskontextes ist das mit ihnen abgefragte Wissen nur sehr begrenzt verwertbar. Das ändert sich, sobald das Niveau Verstehen und Anwendung erreicht wird, und nochmals, wenn den Schülerinnen die Anwendung ihres Wissens in (für sie) neuen Zusammenhängen zugemutet wird.

#### **Exkurs 4: Output oder Outcome?**

Die bei PISA eingesetzten Aufgaben weisen durchwegs ein hohes taxonomisches Niveau auf. Zudem beanspruchen sie, nicht kurzfristige Lerneffekte, wie das bei „gut vorbereiteten“ Schularbeiten der Fall ist (output), sondern nachhaltige Lerneffekte (outcome) zu erfassen. Das gegen PISA häufig vorgebrachte Bedenken, mit ihm würde ein *Teaching and Learning (only) to the tests* begünstigt, trifft auf PISA jedenfalls weniger zu als auf Praktiken im herkömmlichen Unterricht.

Beim Merkmal *Schwierigkeitsgrad (Lösbarkeit)* ist zu unterscheiden zwischen einer von Lehrerinnen und Lehrern geschätzten, und der empirisch ermittelten Größe. Es ist sinnvoll, wenn Lehrer/innen vorab eine Schätzung vornehmen. Sie sollte jedoch mit dem empirisch zu ermittelnden Maß (= wie viel Prozent der Schüler/innen lösen eine Aufgabe) verglichen werden. Der Vergleich erlaubt Rückschlüsse auf die Qualität von Lehre und Lernen.

Wenn man, um ein Beispiel zu nennen, mit einem Test die Geeigneten von den Ungeeigneten trennen will (Selektion), ist eine hohe Ausprägung der *Trennschärfe* unverzichtbar. Im Unterricht steht hingegen nicht unbedingt die Selektionsfunktion im Vordergrund. Oftmals wird man sich daher als Lehrer/in wünschen, dass eine Aufgabe von möglichst allen Schülerinnen und Schülern gelöst wird (geringe Trennschärfe). Trennscharf müssen die gestellten Aufgaben allenfalls insofern sein, als ihre Lösung entsprechende, im Unterricht geförderte Kompetenzen voraussetzt und nicht, wie man so sagt, mit dem Hausverstand zu lösen sind. Wenn diese Kompetenzen von allen Schülerinnen und Schülern nachgewiesen werden: umso besser. Sollte bei einer Korrektur einmal zutage treten, dass insgesamt leistungsstarke Schüler/innen eine Aufgabe schlechter lösen als insgesamt leistungsschwache, ist das ein Hinweis auf einen Mangel der Aufgabe (z. B. eine mangelnde Verständlichkeit, die dazu führen kann, dass gerade leistungsstarke Schüler/innen in die Aufgabe Ansprüche hinein interpretieren, an die Lehrer/innen gar nicht gedacht haben.)

Ein weiteres, wichtiges Merkmal der Aufgaben ist ihre *Validität*. Dabei sind die inhaltliche, die prognostische und die ökologische Validität zu unterscheiden. Die *inhaltliche Validität* ist dann gegeben, wenn mit der Aufgabe Fähigkeiten geprüft werden, die für die Kompetenz in einem Fach oder einem fächerübergreifenden Bereich tatsächlich bedeutsam sind. So können genaue Kenntnisse über den zeitlichen Ablauf von Verhandlungen und Interventionen vor Ausbruch eines Krieges mit Blick auf die Frage, ob er verhindert werden können hätte, durchaus inhaltlich valide sein; ohne diese weiterführende Frage kann es dagegen belanglos sein, ob man das genaue Datum einer Konferenz oder eines Geheimabkommens weiß.<sup>4</sup>

---

<sup>4</sup> Ein konkretes Beispiel in diesem Zusammenhang wäre der Hergang um den zwischen Österreich und Italien vereinbarten Waffenstillstand im November 1918, bei dem zunächst als geringfügig erscheinende Differenzen des Informationsstandes der Verhandlungspartner langfristige politische Auswirkungen hatten.

Wer die Formulierung von Prüfungsaufgaben ernst nimmt, wird darüber nachdenken, ob er im jeweils konkreten Fall (eher) offene oder (eher) geschlossene Aufgaben stellt. Beide Formen haben Vorteile und Nachteile. Zudem sind sie Einwänden ausgesetzt, die prinzipiell nicht zutreffen.

*Offene Aufgaben* (z. B. die Aufforderung zur Beantwortung einer Frage, die Aufforderung zur Ergänzung eines Textes, die Aufforderung zur freien Äußerung über ein Thema in Form eines Aufsatzes, eines Referats oder einer Gestaltung) haben den Vorteil, dass Schüler/innen von sich aus mehr von ihrer Kompetenz zeigen können; ihr Nachteil liegt in der Schwierigkeit, die konkret oft nicht vorhersehbaren Lösungen objektiv zu beurteilen. In der Praxis werden häufig ad hoc gebildete, instabile Maßstäbe angewandt. Eine seriöse Beurteilung bei offenen Aufgaben erfordert hingegen ein begründetes Konzept und relativ viel Zeit.

*Geschlossene Aufgaben* (z. B. Multiple-Choice-Aufgaben, Um- oder Zuordnungsaufgaben) sind hinsichtlich der Formulierung zeitaufwändig, hinsichtlich der Auswertung hingegen zeitökonomisch. Da die Maßstäbe, die auf verbindlichen Konventionen über richtig und falsch bzw. besser und schlechter beruhen, bereits bei der Formulierung festgelegt werden, begünstigen sie eine hohe Objektivität der Beurteilung. Der Einwand, mit ihnen könne nur einfaches Wissen abgefragt werden, trifft allenfalls für eine defizitäre Praxis, aber nicht prinzipiell zu. TIMSS und PISA haben eine große Menge anspruchsvoller geschlossener Aufgaben in Umlauf gebracht und ihr Potenzial aufgezeigt.

Das Merkmal offen ↔ geschlossen ist kein dichotomes, sondern ein kontinuierliches Merkmal. Oft wird daher die Qualität einer Aufgabe in einer bestimmten Situation verbessert werden können, indem man bei offenen Aufgaben den Spielraum einschränkt, bei geschlossenen hingegen (z.B. mit der Aufforderung, die Antwort zu begründen) erweitert.

Wenn Aufgaben gestellt werden, stellt sich schließlich die Frage nach dem Spielraum bei ihrer Annahme und Bearbeitung. Muss (im Rahmen einer Prüfung) jede Aufgabe bearbeitet werden? Wird die Möglichkeit unterschiedlicher Zugänge (z.B. verbal, graphisch oder szenisch; allein oder in einer Partnergruppe etc.) eingeräumt? Nicht jedes Zugeständnis bedeutet eo ipso eine Erleichterung. Zum Beispiel: Wenn Studierende bei einer Klausurarbeit, die 20 Aufgaben umfasst, 3 abwählen dürfen bzw. müssen, ist das nur insofern eine Erleichterung, als sie (a) aktuelle Teilleistungsschwächen haben und (b) diese auch richtig diagnostizieren (Selbstbeurteilungskompetenz). Wenn eine Lösung in einer Kleingruppe erarbeitet wird, ist die Gefahr, dass sich ein Gruppenmitglied daran nicht beteiligt, dann geringer oder gebannt, wenn alle zu einem späteren Zeitpunkt erklären können müssen, wie sie zur Lösung gekommen sind und welche Alternativen sie ggf. verworfen haben.

## 5. Die Qualität der Aufgaben(stellungen) im Rahmen neuerer Verfahren der Leistungsbeurteilung

Abschließend möchte ich mit einer Übersicht versuchen, auf Vor- und Nachteile ausgewählter Verfahren und zugleich auf den Umstand hinzuweisen, dass mit keinem Verfahren alle Probleme gelöst werden können.

Abb. 5: Merkmalsausprägungen ausgewählter Verfahren der Leistungsbeurteilung

Verfahren	Merkmalsausprägung		
	<i>positiv</i>	<i>ambivalent</i>	<i>negativ</i>
<i>Informelle Tests mit geschlossenen Aufgaben</i>	Objektivität, Ökonomie,	Verständlichkeit, Sinnhaftigkeit, Lösbarkeit, Fairness, Anspruch, Nachhaltigkeit	ökologische Validität
<i>Verwendung selbst angefertigter AU</i>	Verständlichkeit, Fairness, ökologische Validität Nachhaltigkeit	Sinnhaftigkeit, Anspruch(!) Lösbarkeit(!) Objektivität, Ökonomie	
<i>Zweistufen-Schularbeit</i>	Nachhaltigkeit, ökolog. Validität, Lösbarkeit	Sinnhaftigkeit, Fairness, Anspruch, inhaltliche Validität	Objektivität (offene Aufgaben), Ökonomie
<i>Portfolios (direkte Leistungsvorlage)</i>	Verständlichkeit, Fairness, Lösbarkeit, Sinnhaftigkeit, ökologische Validität Nachhaltigkeit	Anspruch, inhaltliche / prognostische Validität	Objektivität, Ökonomie

### Exkurs 5: Ökonomie der Beurteilung

Spätestens jetzt sollte klar geworden sein, dass eine Verbesserung der Leistungsbeurteilung und damit eine Verbesserung des Unterrichts mit einem erhöhten Zeitaufwand einhergeht. Damit wird der Aspekt der Arbeitsökonomie wichtig.

Immer vorausgesetzt, die Beurteiler/innen streben eine hohe Qualität im Sinne der aufgezählten Merkmale an, dürfen – mit aller Vorsicht – die folgenden beiden Faustregeln genannt werden:

- Wer bei der Formulierung von Aufgaben – zu Lasten ihrer Qualität – spart, büßt dafür bei der zeitaufwändigen Auswertung.
- Wenn Aufgaben bei vielen Personen oder ohne Schaden wiederholt eingesetzt werden, lohnt sich der Aufwand für anspruchsvolle geschlossene Aufgaben.

## Literaturbasis (eine Auswahl an neueren Arbeiten)

- ARNOLD, K.-H.: Diagnostische Kompetenz erwerben. Wie das Beurteilen zu lernen und zu lehren ist. In: Pädagogik, 51. Jg. (1999) H. 7-8, 73-77.
- BAMBACH, H.: Ermutigungen. Nicht Zensuren. Lengwil a. B.: Libelle, 1996.
- BLÜML, K.: Die sogenannte Zweiphasen- oder Zweistufenarbeit. In: *die*, 18 (1994), H.2, 73-78.
- BLOOM, B. S. et al.: Taxonomy of Educational Objectives. The Classification of Educational Goals. Handbook I: Cognitive Domain. New York: David McKay Co., 16<sup>th</sup> Ed. 1971 (1956).
- BUSCHMANN, I. & THONHAUSER, J.: Im Namen, aber nicht im Sinne des Gesetzes: zur Praxis der Leistungsbeurteilung. In: *Erziehung und Unterricht*, 150. Jg. (2000), 173-191.
- EDER, F.: Fähigkeits- und Leistungsunterschiede auf der Sekundarstufe. In: Eder, F., GROGGER, G. & MAYR, J. (Hg.): *Sekundarstufe I: Probleme – Praxis – Perspektiven*. Innsbruck u. a.: Studienverlag, 2002, 135-157.
- EDER, F.: Matcomp. Ein Verfahren zur Erfassung mathematischer Kompetenzen am Ende der Sekundarstufe I erstellt auf Basis von TIMSS-Aufgaben. Forschungsbericht, Linz, 2002.
- EDER, F. u. a. (Hg.): *Qualitätsentwicklung und Qualitätssicherung im österreichischen Schulwesen*. Innsbruck u. a.: Studienverlag, 2002. (*mehrere einschlägige Beiträge!*)
- FRIEDRICH JAHRESHEFTE 1996 (Prüfen und Beurteilen. Zwischen Fördern und Zensieren.) und 2003 (Aufgaben. Lernen fördern – Selbständigkeit entwickeln.), Seelze: Friedrich Verlag. (*mehrere einschlägige Beiträge*)
- GRELL, J. & GRELL, M.: *Unterrichtsrezepte*. Weinheim und Berlin: Beltz Taschenbuch, 1999.
- HELMKE, A.: *Unterrichtsqualität – erfassen, bewerten, verbessern*. Seelze: Kallmeyer, 2003.
- HERBER, H.-J.: *Innere Differenzierung*. Stuttgart: Kohlhammer, 1983.
- JACOBS, B.: *Aufgaben stellen und Feedback geben*. <http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/index.htm>
- JÄGER, R. S.: *Von der Beobachtung zur Notengebung. Ein Lehrbuch*. Landau: Verlag Empirische Pädagogik, 2000.
- LISSMANN, U.: *Beurteilung und Beurteilungsprobleme bei Portfolios*. In: JÄGER, R.S: a.a.O., 283-329.
- LUDERER, W.: *Die Lateinschularbeit und ihre Bewertung*. In: *Erziehung und Unterricht* 122 (1972) Heft 2, 119-127.
- OSER, F. & OELKERS, J. (Hg.): *Die Wirksamkeit der Lehrerbildungssysteme: von der Allrounderbildung zur Ausbildung professioneller Standards*. Zürich: Ruediger, 2001.
- PÄDAGOGIK*, 2001, Hefte 1 – 7-8: Serie: *Fachleistungen beobachten, beschreiben, bewerten*.

- PAGE, E.B.: Is the world an orderly place? A review of teacher comments and student achievement. In: Journal of Experimental Education. 60. Jg. (1992) 161-181.
- POSCH, P & THONHAUSER, J.: Die lehrplanbedingte Erosion affektiver Ziele. In: Unterrichtswissenschaft 10 (1982), 212-224.
- RHEINBERG, F.: Bezugsnormen und schulische Leistungsbeurteilung. In: WEINERT, F. E. (Hg.): Leistungsmessungen in Schulen. Weinheim und Basel: Beltz, 2001, 59-371.
- SACHER, W.: Leistungen entwickeln, überprüfen und beurteilen. Grundlagen, Hilfen und Denkanstöße für alle Schularten. Bad Heilbrunn: Klinkhardt, 2001.
- SALDERN, M. von: Schulleistung in Diskussion. Hohengehren: Schneider, 1999.
- SCHAFFENRATH, M. & WIESER, I.: Basiskompetenzen für lebenslanges Lernen aufbauen. In: EXENBERGER, S. & SCHÖBER, P. (Hg.): Baustelle Lehrlingsausbildung. Innsbruck u. a.: Studienverlag, 2005, 108-145 (+ CD).
- SCHEIFLINGER, W. & PETRI, G.: Probleme der Lernerfolgseinstellung. Wie kann Schulstress abgebaut, Lernfreude verstärkt und die Leistungsbeurteilung objektiviert werden? Graz: ZSE, Forschungsbericht 28, 1999.
- TEDDLIE, C. & REYNOLDS, D. (Eds.): The International Handbook of School Effectiveness Research. London und New York: Falmer Press, 2000.
- TERHART, E., LANGKAU, Th. & LÜDERS, M.: Selektionsentscheidungen als Problembereich professionellen Lehrerhandelns. Bochum: Ruhr Universität Eigenverlag, 1999.
- THONHAUSER, J.: Qualitätsentwicklung mit Hilfe einer Prüfung? In: Salzburger Beiträge zur Erziehungswissenschaft, 6 (2002) 2, 63-72.
- THONHAUSER, J.: Diagnostische Kompetenzen von Lehrerinnen und Lehrern und die aktuelle Diskussion über Standards. Ein Beitrag über Österreich. In: DÖBERT, H. & FUCHS, H.-W. (Hg.): Leistungsmessungen und Innovationsstrategien in Schulsystemen. Münster u. a.: Waxmann, 2005, 103-124.
- VALTIN, R.: Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand. Weinheim: Beltz, 2002.
- VIERLINGER, R.: Leistung spricht für sich. „Direkte Leistungsvorlage“ (Portfolios) statt Ziffernzensuren und Notenfetischismus. Heinsberg: Dieck-Verlag, 1999.
- WAGNER, J. W. L.: Soziale Vergleiche und Selbsteinschätzungen. Münster u.a.: Waxmann, 1999.
- WEINERT; F. E. (Hg.): Leistungsmessungen in Schulen. Weinheim und Basel: Beltz, 2001. (*mehrere einschlägige Beiträge!*)
- WINTER, F.: Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen. Hohengehren: Schneider, 2004.
-