

# Leistungsfeststellung und Leistungsbeurteilung

Ferdinand Eder, Georg Hans Neuweg und Josef Thonhauser

Die „Leistungsbeurteilungsverordnung“ (LBVO)<sup>1</sup> unterscheidet zwischen *Leistungsfeststellung* als einem Vorgang des Messens und *Leistungsbeurteilung* als einem Vorgang des Bewertens der gemessenen Leistungen. Zugleich scheint *Leistungsbeurteilung* in der Verordnung auch als Oberbegriff für den gesamten Prozess des Prüfens und Beurteilens in der Schule verwendet zu werden. Dieser Sprachgebrauch wird für den vorliegenden Text übernommen.<sup>2</sup>

Terminologische  
Vorbemerkung

B

## 1 Die pädagogische und bildungspolitische Relevanz des Themas

Leistungen werden im Schulsystem in vielfältigen Formen und mit vielfältigen Intentionen und Funktionen gemessen. Tabelle 1 gibt einen groben Überblick.

Systemebene	Formen	primäre (gesetzliche) Intention	Funktionen		
			pädagogische Funktionen	gesellschaftliche Funktionen	personbezogene Auswirkungen
Unterricht	Informationsfeststellungen <sup>a)</sup>	Lernstandsdiagnose (unbenotet)	Setzen von Bedeutsamkeitssignalen, Feedback für SchülerInnen und LehrerInnen, Motivierung	Allokation, Selektion, Bericht an Abnehmer	Selbstkonzept, Einstellung zur Schule, Bildungsaspirationen
	Punktuelle Leistungsfeststellungen	notenförmige Beurteilung der Schülerleistungen; Bericht an Erziehungsberechtigte			
	Unterrichtsbegleitende Leistungsfeststellungen („Mitarbeitsfeststellung“)				
Gesamtsystem	Internationale Leistungsvergleiche	international vergleichende Feststellung der Systemleistung		Grundlage für bildungspolitische Konsequenzen, z. B. Schulsystemreform, Adaptierung der Anforderungen, Ressourcenentscheidungen, Lehrerbildung	
	Standards-Testungen	kriterienorientierte Feststellung der Systemleistung; Rückmeldung an Schulen	Setzen von Bedeutsamkeitssignalen, Feedback für LehrerInnen und SchülerInnen		

Tab. 1: Übersicht über Formen und Funktionen der Leistungsfeststellung

Anmerkungen. <sup>a)</sup> Informationsfeststellungen sind „Feststellungen der Leistungen der Schüler, die dem Lehrer nur zur Information darüber dienen, auf welchen Teilgebieten die Schüler die Lehrziele erreicht haben und auf welchen Teilgebieten noch ein ergänzender Unterricht notwendig ist“ (§ 1 Abs. 2 LBVO).

1 Verordnung des Bundesministers für Unterricht und Kunst vom 24. Juni 1974 über die Leistungsbeurteilung in Pflichtschulen sowie mittleren und höheren Schulen (Leistungsbeurteilungsverordnung), BGBl. Nr. 371/1974, zuletzt geändert durch BGBl. II Nr. 35/1997.

2 Eine begriffliche Bereinigung könnte darin bestehen, „Leistungsfeststellung“ für den Prozess des Messens, „Leistungsbewertung“ für die Umsetzung eines Messergebnisses in eine Note oder sonstige Bewertung und „Leistungsbeurteilung“ als Oberbegriff für beides zu verwenden.

**Pädagogische Funktionen**

Die *pädagogischen Funktionen* beziehen sich auf die Steuerung des Lehr-Lern-Geschehens. Dass etwas Gegenstand von Leistungsfeststellungen (Prüfungen) wird, verdeutlicht den Schüler/inne/n, welche Inhalte und Leistungen den Lehrpersonen als besonders bedeutsam erscheinen (*Signalfunktion*). Sofern Transparenz hinsichtlich der Anforderungen besteht, können sie ihre Lernaktivitäten darauf ausrichten.<sup>3</sup> Wenn beispielsweise die Messung der Leistungen auf einem höheren taxonomischen Niveau erfolgt, lernen Schüler/innen, dass es letztlich auf das Verstehen oder Anwenden, nicht bloß auf abrufbare Wissensbestände ankommt. Solche Hinweise können Schüler/innen auch aus der Form der Leistungsfeststellung gewinnen. Wenn sie Wahlmöglichkeiten zur Demonstration ihres Könnens vorfinden oder zur Reflexion und Selbstbewertung ihrer Leistungen veranlasst werden, signalisiert dies andere Bildungsziele als eine traditionelle Schularbeit. Neben dieser inhaltlichen Steuerung ist für Lehrende wie Lernende die Funktion der Rückmeldung wichtig (*Feedbackfunktion*). Diese kann sich auf den Prozess des Lernens beziehen oder auf dessen Ergebnis (z. B. Rückmeldung zu gewählten Lernwegen; Analyse von prototypischen Fehlern). Ebenfalls pädagogisch wirksam, aber umstritten sind die *Motivierungs-* und die *Disziplinierungsfunktion* von Noten. Möglicherweise positiven kurzfristigen Effekten steht hier die Gefahr des Entstehens von Schulunlust, der Unterminierung von Selbststeuerungskräften und des Verlustes der intrinsischen Lernmotivation gegenüber.

**Gesellschaftliche Funktionen**

Während Leistungsfeststellung und Leistungsbeurteilung (LFLB) aus pädagogischer Sicht den Lernprozess aller Schüler/innen bestmöglich fördern soll, betonen die *gesellschaftlichen Funktionen* vor allem die Unterschiede zwischen den Schüler/inne/n. Die Zuordnung zu unterschiedlichen Bewertungsklassen (*Klassifizierungsfunktion*) ist Voraussetzung für die Zuweisung zu verschiedenen Laufbahnen innerhalb und außerhalb der Schule (*Allokationsfunktion*) und zur Vergabe von Berechtigungen (*Selektionsfunktion*). Schulnoten berechtigen zum Aufsteigen in die nächste Schulstufe, zum Wechsel der Leistungsgruppe, zum Übertritt in bestimmte anschließende Schulen oder – durch die Matura – zum Besuch postsekundärer Bildungseinrichtungen. Nicht selten knüpfen auch Abnehmer außerhalb der Schule den Zugang zu bestimmten Berufen, insbesondere zu den Lehrberufen, an das Vorliegen guter Noten. Während für die Wahrnehmung der pädagogischen Funktionen vor allem qualitative Rückmeldungen wichtig sind, wird im Selektionsgeschehen die *Note* zum zentralen Instrument; für Schüler/innen, Eltern und Abnehmer außerhalb der Schule bietet sie eine rasch überschaubare Information über den Erfolg bei der Bewältigung der schulischen Anforderungen (*Berichtsfunktion*).

**Personbezogene Funktionen**

Die *personbezogenen Funktionen* der Leistungsbeurteilung betreffen zunächst die Beeinflussung der *Einstellung zur Schule*. Schüler/innen mit guten Noten sind mit der Schule insgesamt zufriedener und erleben bessere Beziehungen zu ihren Lehrpersonen. Zum anderen wird das *Selbstkonzept* der Schüler/innen wesentlich durch die Rückmeldungen aus der Schule beeinflusst. Junge Menschen halten sich in einem hohen Ausmaß dort für leistungsfähig und tüchtig, wo sie in der Schule gute Noten haben und trauen sich in jenen Fächern wenig zu, in denen ihre Noten schlecht sind. Die Vorstellungen, die sie von ihren eigenen Leistungsmöglichkeiten haben, prägen dann auch ihre späteren *Bildungsaspirationen* und ihre Motivation zum (lebenslangen) Lernen. Und schließlich: Unabhängig vom Niveau der Noten ist Leistungsbeurteilung für viele Schüler/innen mit *Schul- und Prüfungsgangst* verbunden (vgl. Abschnitt 2.2).

**Weitere Begleiterscheinungen**

Weitere Begleiterscheinungen von LFLB unterstreichen die Relevanz der Thematik. Die Androhung schlechter Noten, insbesondere die vom System ausgehende Drohung, ein Jahr wiederholen zu müssen, setzt beispielsweise ein umfangreiches Unterstützungssystem in Gang, das sich zu einem geringen Teil auf innerschulischen Förderunterricht, zum größten Teil aber auf familiäre Unterstützung und *bezahlte Nachhilfe* schon in der Volksschule stützt (Wagner et al. 2003: 238). LFLB wirkt in die Familien hinein und beeinflusst die *Beziehung zwischen Eltern und Kindern* (Streit über Schulleistungen gehört zu den häufigsten Konfliktthemen

3 Das birgt allerdings, so wird befürchtet, die Gefahr eines *learning to the test*.

zwischen Eltern und Kindern; gute Leistungen führen zu verstärkter Zuwendung). Und nicht zuletzt ist LFLB *Taktgeber für den Ablauf des Schuljahres*. Die emotionalen Begleiterecheinungen von Schularbeiten wirken sich beispielsweise negativ auf den Unterricht der vorher und nachher angesetzten Fächer aus; Schularbeits- und Prüfungstermine bestimmen das Leben der Schüler/innen, und sie setzen mit dem Lernen aus, wenn die entsprechenden Prüfungen abgelegt sind. Der Unterricht in den mittleren und höheren Schulen kann am Schuljahresanfang erst beginnen, wenn die Wiederholungsprüfungen absolviert sind, und er endet de facto, wenn die Jahresnoten feststehen (sodass in den letzten Wochen des Schuljahres auf nichtfachliche Schulveranstaltungen ausgewichen wird). LFLB blockiert damit über das für ihre Durchführung notwendige Erfordernis hinaus Zeit, die eigentlich für Unterricht bestimmt ist.

## 2 Leistungsbeurteilung in der Praxis: Rechtsgrundlagen – Umsetzung – Entwicklungstendenzen

### 2.1 Rechtliche Grundlagen und ihre Umsetzung

#### 2.1.1 Beurteilungsstufen und Beurteilungsmaßstäbe

Für die Beurteilung der Schülerleistungen sehen Schulunterrichtsgesetz (SchUG) und LBVO die Anwendung eines person- und bezugsgruppenunabhängigen, lehrzielbezogenen Maßstabes vor (sog. kriteriale oder curriculare Norm):<sup>4</sup> Die Einordnung der Leistungen in eine der fünf Beurteilungsstufen hat mit Blick auf die Forderungen des Lehrplanes unter Bedachtnahme auf den jeweiligen Unterrichtsstand zu erfolgen (§ 18 Abs. 1, letzter Satz SchUG). Für die Einordnung sind maßgeblich:

**Beurteilungsstufen und -maßstäbe: Rechtslage**

- (1) ein vorwiegend reproduktiver Bereich, der sich auf die Erfassung und die Anwendung des Lehrstoffes sowie die Bewältigung von (relativ vertrauten) Aufgaben bezieht; diesen differenziert § 14 LBVO weiter in (a) wesentliche und (b) über das Wesentliche hinaus gehende Bereiche;
- (2) Eigenständigkeit und Fähigkeit zur selbstständigen Anwendung des Wissens und Könnens auf neuartige Aufgaben (vgl. auch Jonak/Kövesi 2007: 818).

Die Heranziehung der lehrzielbezogenen Norm ist – v. a. wenn diese für individuelle Rückmeldungen um die individuelle Norm ergänzt wird – in der erziehungswissenschaftlichen Literatur unumstritten. Sie ist die vorzugswürdige Norm

**Beurteilungsstufen und -maßstäbe: Bewertung**

- im Interesse der Lehr-/Lern-Steuerung,
- unter dem Aspekt der Beurteilungsgerechtigkeit,
- mit Blick auf die Berichtsfunktion von Noten und Zeugnissen und
- unter dem Gesichtspunkt der Hinführung der Schüler/innen zur sachlich begründeten Selbsteinschätzung.

<sup>4</sup> In Ausnahmefällen (im Falle körperlicher Behinderungen, in den sog. „Begabungsfächern“ und bei mangelnder Kenntnis der Unterrichtssprache) ist die individuelle Norm mit zu berücksichtigen (vgl. § 18 Abs. 6, 8, und 9 SchUG). Die Anwendung der sozialen Norm ist ausgeschlossen (Erk. des VerwGH Slg. 10391A).

**Beurteilungsstufen  
und -maßstäbe:  
Umsetzungsprobleme**

Aus mindestens zwei Gründen spielt die Definition der Beurteilungsstufen durch den Gesetzgeber aber in der Schulpraxis eine geringe Rolle:

1. Die Definition der Beurteilungsstufen bedient sich hochgradig unbestimmter Rechtsbegriffe, v. a. des Begriffes des weder in SchUG oder LBVO noch in den Lehrplänen inhaltlich bestimmten „Wesentlichen“; auch die Unterscheidung von Kern- und Erweiterungsbereichen im „Lehrplan 2000“ erscheint dafür nicht konkret genug. Lehrpläne enthalten insbesondere keine hinreichend konkreten Kompetenzkataloge oder Lehrzielangaben, die eine auch nur annähernd einheitliche Operationalisierung durch die Lehrer/innen erwarten lassen. Die Anwendung *einheitlicher* Kriterien zur Gewährleistung von Gerechtigkeit und Objektivität ist damit nicht möglich. Ihr Fehlen begünstigt zudem die Anwendung der sozialen Norm und damit überhaupt ein Abgehen von einer rein sachorientierten Beurteilung, weil die von den Schüler/inne/n faktisch erbrachten Durchschnittsleistungen zu einer wichtigen Orientierungshilfe bei der Wahl des Beurteilungsmaßstabes werden.
2. Die Definition der Notenstufen stellt sehr hohe Ansprüche nicht nur an die Schüler/innen, sondern auch an den Unterricht. Denn schon der Genügend-Schüler braucht Anwendungskompetenz, der Befriedigend-Schüler muss „das Wesentliche zur Gänze“ beherrschen<sup>5</sup> und für bessere Bewertungen sind Transferfähigkeit und Eigenständigkeit gefordert. Diese Forderungen finden nicht immer Entsprechung im Unterricht; Stoff- und Vollständigkeitsorientierung verdrängen manchmal die Orientierung am Exemplarischen und das Üben, die Anwendung des Gelernten an neuen Beispielen kommt oft zu kurz.<sup>6</sup> „Dem Stand des Unterrichts“ entsprechend dürften dann Lernziele höheren Niveaus nicht eingefordert werden – fairerweise gegenüber dem Lernenden und korrekterweise angesichts der legislativen Vorgaben.

**2.1.2 Ermittlung von Gesamtbeurteilungen****Die Gesamtbeurteilung:  
Rechtsslage**

Lehrkräfte müssen zusammenfassende Prüfungsnoten geben und aus den Prüfungsnoten wiederum zusammenfassende Jahresnoten bilden. Da die Beurteilungsstufen auf Einzelprüfungen wie auch auf die Beurteilung der Leistungen einer ganzen Schulstufe Anwendung finden, hat der Lehrer bei der Vergabe von Prüfungsnoten Einzelaufgaben und bei der Vergabe von Jahresnoten Einzelprüfungen im Hinblick auf die Wesentlichkeit der jeweiligen Anforderungen und auf das Niveau der geforderten intellektuellen Operationen zu wägen und in eine Gesamtbeurteilung zu überführen. Zudem sind Stoffumfang, Schwierigkeitsgrad und, vor allem in „aufbauenden“ Fächern, Zeitpunkt der Prüfung (vgl. § 3 Abs. 5 und § 20 LBVO) gewichtungsrelevante Faktoren bei der Ermittlung der Jahresnote.

**Die Gesamtbeurteilung:  
Bewertung und  
Umsetzungsprobleme**

Für die „Verrechnung“ von Leistungen zu Noten und von Einzelnoten zu Gesamtnoten bietet die Rechtslage keine praktikablen Anhaltspunkte. Praxisgängige Verfahren, v. a. die Anwendung von Punktesystemen und die Mittelwertberechnung von Noten, bleiben in SchUG und LBVO nicht nur unerwähnt, sie werden durch die geltende Rechtslage (und das Skalenniveau der Noten) im Grunde sogar ausgeschlossen. Im Einzelnen (vgl. auch Neuweg 2006a: 79–83, 2006b):

- Die Verpflichtung zur Benotung jeder Einzelprüfung ist schon deshalb problematisch, weil nicht jede Prüfung jene Aufgabenvielfalt aufweist und aufweisen kann, die eine Einordnung der gezeigten Leistung unter die fünf Notenstufen erlaubt.
- Verfahren, in denen Prüfungsnoten auf die Addition von Punkten oder Fehlern zurückgehen, unterscheiden kaum zwischen Reproduktionsleistungen einerseits und eigenständigen bzw. selbstständigen Leistungen andererseits sowie zwischen wesentlichen und dar-

<sup>5</sup> Und damit alles, wenn man unterstellt, dass nichts anderes als Wesentliches zum Gegenstand des Unterrichts wird (Vierlinger 1993: 56).

<sup>6</sup> Hinweise auf den sehr unterschiedlichen Umgang von Lehrer/inne/n mit diesem Problem finden sich bei Thonhauser 2005b: 29.

über hinausgehenden Bereichen. Das ermittelte Punkte- oder Fehlerkonglomerat kann nicht begründet auf die inhaltlich bestimmten Notenkategorien der LBVO bezogen werden.

- Vor diesem Hintergrund erscheint die verbreitete Usance, eine positive Note ab Erreichen der 50%-Punktemarke zu vergeben, besonders problematisch. Für diese Marke existiert in der LBVO kein Anhaltspunkt. Auch ist immer inhaltlich zu fragen, in welchen Leistungsbereichen ein Schüler diese Punkte erzielt hat und ob dieses Leistungsbild der qualitativen und eben nicht quantitativen Beschreibung einer Notenstufe entspricht.
- Das Gebot der differenzierten inhaltlichen Wägung besteht auch zwischen den Leistungsfeststellungen. Auch hier dürfte die übliche Bildung eines arithmetischen Mittels über mehrere Leistungsfeststellungen hinweg nur selten rechtskonform und sachgerecht sein. Abgesehen davon, dass mit ordinal skalierten Daten Mittelwertberechnungen grundsätzlich nicht vorgenommen werden sollten,<sup>7</sup> kann die Bildung des Mittelwerts zudem lediglich das Kriterium der Anzahl der Leistungsfeststellungen berücksichtigen, weder aber die anderen in § 3 Abs. 5 LBVO angeführten Kriterien (Stoffumfang, Schwierigkeitsgrad) noch das Gebot, nach Maßgabe der fachlichen Eigenart des Unterrichtsgegenstandes und des Aufbaues des Lehrstoffes dem auf einer Schulstufe zuletzt erreichten Leistungsstand das größere Gewicht zuzumessen (§ 20 LBVO).

### 2.1.3 Unterrichtsbegleitende versus punktuelle Leistungsfeststellungen

Der Gesetzgeber priorisiert die Feststellung der Mitarbeit gegenüber den punktuellen Prüfungsformen deutlich (vgl. auch Erl. Bem. In RV 431 der Beil. [XVIII. GP]):

- § 3 Abs. 1 LBVO nennt die Mitarbeitsfeststellung als erste Prüfungsform, kennzeichnet alle anderen Formen als „besondere“ und signalisiert auf diese Weise, dass der Mitarbeitsfeststellung der Status einer tragenden diagnostischen Säule zukommt.
- Die Mitarbeitsfeststellung ist die einzige zwingend in allen Unterrichtsgegenständen anzuwendende Prüfungsform; für alle anderen Formen existieren Einsatzbeschränkungen.
- Die Mitarbeitsfeststellung ist den anderen Prüfungsformen gleichwertig (§ 3 Abs. 5 LBVO).
- Soweit lehrplanmäßig keine Schularbeiten vorgesehen sind, kann sich die Jahresbeurteilung ausschließlich unterrichtsbegleitend ergeben.
- Für punktuelle Prüfungen existiert ein generelles Sparsamkeitsgebot (nur so viele, wie unbedingt notwendig, § 3 Abs. 4 LBVO).

Grundsätzlich ist eine Verlagerung des Gewichtes weg von punktuellen zu unterrichtsbegleitenden Formen der Leistungsfeststellung wünschenswert aufgrund

- einer erhöhten Zuverlässigkeit (Einbeziehung einer größeren Zahl von Leistungsdaten),
- einer erhöhten Validität durch Diagnose in „natürlichen“ Situationen,
- einer gleichmäßigeren Belastung der Schüler/innen<sup>8</sup> und
- der Erfassung eines vielfältigeren Leistungsspektrums.

**Begleitende  
vs. punktuelle  
Leistungsfeststellung:  
Rechtslage**

**Begleitende  
vs. punktuelle  
Leistungsfeststellung:  
Bewertung und  
Umsetzungsprobleme**

<sup>7</sup> Harwell/Gatti (2001: 113–114) sehen zwei Möglichkeiten für den Umgang mit rangskalierten Daten: (a) die Beschränkung auf geeignete, nämlich nonparametrische, Verfahren; (b) eine Transformation der Daten, z. B. basierend auf der Item-Response-Theorie, wodurch parametrische Verfahren anwendbar werden. Die weit verbreitete Vernachlässigung des tatsächlichen Skalenniveaus beeinträchtigt die Validität von Ergebnissen.

<sup>8</sup> Allerdings steigt damit auch die Gefahr, dass es faktisch keine Lernphasen gibt, die beurteilungsfrei sind. Eine (zeitweilige) Entmischung von Lern- und Leistungssituationen ist durchaus vorteilhaft (vgl. Neuweg 2006a: 33, Jäger et al. 2008: 12).

Ursprünglich wurden von der Priorisierung der Mitarbeitfeststellung zudem Rückwirkungen auf die Unterrichtsgestaltung erwartet, weil „diese Form der Leistungsfeststellung bei dem aus pädagogischer Sicht zu bevorzugenden Arbeitsunterricht, nicht jedoch beim Frontalunterricht verwirklichtbar ist“ (Jonak 1984: 34).

Rechtslage und Praxis klaffen aber auseinander:

- Die Mitarbeitfeststellung gilt, insbesondere in Schularbeitengegenständen, den anderen Prüfungsformen keineswegs immer als gleichwertig (Buschmann/Thonhauser 2000) und hat häufig nur den Status eines „Züngleins an der Waage“.
- Sie umfasst, wie auch schon vom Verwaltungsgerichtshof gerügt (Erk. vom 9. 3. 1981, Zl. 10/3420/80), oft nur Teile des in § 4 Abs. 1 LBVO dargestellten, die gesamte Unterrichtsarbeit umgreifenden Leistungsspektrums, z. B. lediglich „Bankfragen“.
- Ihre Aufgabe wird – im Widerspruch zu § 18 Abs. 5 SchUG (Verbot der Einbeziehung des Verhaltens des Schülers in die Leistungsbeurteilung) – häufig verkannt, wenn durch sie nicht Leistungen beurteilt werden, sondern sie als Disziplinierungsinstrument genutzt wird.
- Mitarbeitfeststellungen erweisen sich, etwa in Form testähnlicher „schriftlicher Mitarbeitüberprüfungen“, oft als Miniaturen besonderer Leistungsfeststellungen. Als „kleine Tests“ oder „kleine mündliche Prüfungen“ konterkarieren sie die Grundidee der Mitarbeitfeststellung: Mitgearbeitet werden kann, wenn und solange unterrichtet, nicht mehr aber, wenn der Unterricht zu Prüfungszwecken still gelegt wird.

Besonders problematisch ist die Regelungsasymmetrie zu den besonderen Prüfungsformen hin. Die Durchführung der Mitarbeitfeststellung ist weitgehend unregelt. § 4 LBVO sieht weder Durchführungsbestimmungen noch Restriktionen vor. Die vergleichsweise sehr restriktiven Bestimmungen zu den Tests, Diktaten und mündlichen Prüfungen (§§ 5 ff. LBVO) laden dazu ein, Mitarbeitfeststellung zu nennen, was dem Wesen nach eigentlich Test, Diktat oder mündliche Prüfung wäre und entsprechend unter die dafür geltenden Schutzbestimmungen fiel. Umgangen werden können auf diese Weise zum Beispiel Bestimmungen zur Beschränkung der Anzahl punktueller Prüfungen, zu ihrer sachlichen bzw. zeitlichen Unzulässigkeit sowie zur Ankündigung und Durchführung punktueller Prüfungen (vgl. bereits Fankhauser 1992/93: 27 und jüngst Neuweg 2006b: 35–36).

#### 2.1.4 Berufung

##### Berufung: Rechtslage

Berufungen sind gegenwärtig nur gegen das Nicht-Aufsteigen in die nächste Schulstufe möglich, nicht jedoch gegen einzelne Noten (das gilt selbst für ein Nicht genügend im Jahreszeugnis, wenn trotzdem das Aufsteigen in die nächste Stufe gewährleistet ist), und haben keine aufschiebende Wirkung.

##### Berufung: Bewertung und Umsetzungsprobleme

Das Verfahren ist für die Behörden, auch angesichts der zeitlichen Massierung im Sommer bzw. Herbst, relativ aufwändig, einem großem zeitlichen Erledigungsdruck ausgesetzt und auf schulrechtliche Aspekte zentriert, obwohl häufig zwischenmenschliche und pädagogische Aspekte den Problemerkern bilden. Die erstaunlich geringe Verfahrenszahl hat schon früh die Vermutung begründet, dass, möglicherweise aufgrund von Repressionsängsten, die Zahl der erhobenen Rechtsmittel in keinem Verhältnis zur Zahl der Personen steht, die mit schulischen Entscheidungen tatsächlich unzufrieden sind (Fankhauser 1988: 65).

#### 2.1.5 Notenumrechnung in den Leistungsgruppen

##### Notenumrechnung: Rechtslage

Nach Abschaffung der Aufnahmeprüfung für die berufsbildenden mittleren und höheren Schulen wurde die Berechtigung zum Übertritt an ein Notenkalkül aus Deutsch, Englisch und Mathematik gebunden. Schüler/innen mit positivem Abschluss der AHS-Unterstufe oder der ersten Leistungsgruppen der Hauptschule sowie Schüler/innen, die in den zweiten Leistungsgruppen mindestens die Note „Gut“ aufweisen, sind ohne Aufnahmeprüfung

zum Übertritt berechtigt. Auf Beschluss der Klassenlehrerkonferenz kann diese Übertrittsberechtigung auch mit einem „Befriedigend“ in der 2. Leistungsgruppe erteilt werden. Für die Aufnahme in (mindestens dreijährige) mittlere Schulen ist mindestens der positive Abschluss der 2. Leistungsgruppe erforderlich. Für Hauptschulen ohne Leistungsgruppen gelten gesonderte Regelungen.

Im Hinblick auf die Berechtigungen bedeutet dies, dass alle Noten der dritten Leistungsgruppe einem „Nicht genügend“ gleichzuhalten sind. Durch die Absolvierung einer Aufnahmeprüfung kann zwar trotzdem eine Übertrittsberechtigung erworben werden; in der Praxis wird das Bestehen dieser Prüfung als „Genügend“ gewertet, sodass faktisch immer eine Schlussreihung der betreffenden Schüler/innen erfolgt. Tatsächlich sind die Leistungsüberschneidungen zwischen den Leistungsgruppen jedoch so groß, dass eine generelle Abwertung der dritten Leistungsgruppe ungerechtfertigt ist (Eder 1998, 2001, Haider/Schreiner 2006, Thonhauser/Pointinger, im Druck).

**Notenumrechnung:  
Bewertung und  
Umsetzungsprobleme**

B

## 2.2 Leistungsfeststellung und Leistungsbeurteilung aus der Sicht der Schüler/innen

Die gesetzliche Regelung der Leistungsbeurteilung zielt in hohem Ausmaß darauf ab, Prüfungssituationen zu versachlichen und transparent zu machen, die Zahl der Prüfungen zu verringern, die Verwendung von Noten als Disziplinierungsmittel hinten zu halten und das Prüfungsgeschehen angstfreier zu gestalten. Inwieweit diese Intentionen erreicht werden, kann aus den Befindensuntersuchungen 1994 (Eder 1995) und 2005 (Eder 2007) abgeschätzt werden.

In der Befindensuntersuchung 1994 wurden Schüler/innen gefragt, ob sie am vorangegangenen Schultag eine Leistungsfeststellung gehabt haben. 62% verneinten, 32% gaben eine Leistungsfeststellung, die verbleibenden 6% gaben zwei bis vier Leistungsfeststellungen an. Da die Befragungen an allen Wochentagen stattfanden, lässt sich daraus ein grobes Bild der Prüfungshäufigkeit ermitteln: An einem durchschnittlichen Schultag absolviert jedenfalls mehr als ein Drittel der Schüler/innen mindestens eine Leistungsfeststellung. Umgerechnet auf die – damals noch sechs Tage umfassende – Schulwoche haben Schüler/innen pro Woche etwa zwei Leistungsfeststellungen zu absolvieren. In der Volksschule ist diese Quote deutlich niedriger, zwischen den übrigen Schultypen bestehen keine großen Unterschiede (Sonderauswertung aus dem Datensatz der Befindensuntersuchung 1994).

**Anzahl der Prüfungen**

20% (1994) bzw. 19% (2005) der Schüler/innen von der 4. bis zur 12. Schulstufe stimmten der Aussage „Die Noten, die ich bekomme, sind oft ungerecht“ mit „stimmt ziemlich“ oder „stimmt genau“ zu. Mehr als 30% der Schüler/innen der Sekundarstufe II stimmten 2005 der Aussage „Wenn man sich schlecht benimmt, muss man damit rechnen, dass man zur Strafe streng geprüft wird“ zu; die Vermischung von Leistungsbeurteilung und Disziplinierung ist also in der Wahrnehmung der Schüler/innen keine Seltenheit.

**Gerechtigkeit und  
Disziplinierung**

Angst ist nach wie vor eine Begleiterscheinung des Schulbesuchs. Etwas mehr als 20% der Schüler/innen berichten, sie hätten „oft“ Prüfungsangst (Eder 2007: 35–37). Sie steigt beim Übertritt von der Volksschule in die Sekundarstufe I deutlich an und ist dann in allen Schultypen ungefähr gleich hoch ausgeprägt.

**Schul- und  
Prüfungsangst**

## 2.3 Aktuelle Initiativen zur Verbesserung und Weiterentwicklung der Leistungsbeurteilung

Zu den bedeutsamsten aktuellen Entwicklungen im Rahmen der Weiterentwicklung von LBLF gehören das Projekt Bildungsstandards, die Initiative qibb, Bemühungen im Rahmen der „Neuen Mittelschule“ und die Teilzentralisierung der Reifeprüfung.

**Bildungsstandards**

Das Verhältnis der *Bildungsstandards* zu Lehrplänen und zur Leistungsbeurteilung auf mittlere bis längere Sicht ist unklar. Unmittelbare Auswirkungen auf die Leistungsbeurteilung sind derzeit weder intendiert noch zu erwarten. Indirekt könnte die verbindliche Formulierung von Standards am Ende der einzelnen Schulabschnitte (Volksschule, Sekundarstufen I und II) freilich einen Beitrag zur Verähnlichung der Anforderungen – und damit zur Urteils-gerechtigkeit – in jenen Fächern leisten, in denen Standards formuliert werden (Deutsch, Mathematik, Englisch). Auf längere Sicht wird vermutlich außerdem die Nichteinbeziehung der Testergebnisse in die Leistungsbeurteilung kaum zu halten sein.

Die bisherige Praxis der Rückmeldungen der Ergebnisse (Lucyshyn 2007) erschwert selbst eine mittelbare Nutzbarmachung der Testergebnisse für die Weiterentwicklung des Unterrichts oder der Schulen und lässt damit ein erhebliches Potenzial ungenützt:

- Individuelle Ergebnisse werden nur den betroffenen Schüler/inne/n zurückgemeldet, Lehrer/innen erfahren lediglich Klassenergebnisse.
- Schulleiter/innen erfahren nur Daten, die auf Schulebene aggregiert sind, nicht aber, wie die einzelnen Klassen abgeschnitten haben. Die Setzung treffsicherer Maßnahmen wird dadurch stark beeinträchtigt.

**qibb**

In der 2004 vom Bildungsministerium gestarteten Initiative „*QIBB – Qualität in der Berufsbildung*“ wurde für die Schuljahre 2006/07 und 2007/08 Leistungsbeurteilung als bundesweiter Arbeitsschwerpunkt für die berufsbildenden Schulen festgelegt. An den kaufmännischen und humanberuflichen Schulen wurde die Arbeit systematisch durch einen breit angelegten Raster unterstützt, der fünf Maßnahmenbereiche vorsieht: reflektierte Auswahl von Prüfungsanforderungen, professionelles Prüfen, objektives Beurteilen, Transparenz-sicherung und angemessene Ergebniskultur (Neuweg 2008).

Die in den Bundesqualitätsberichten der einzelnen Schultypen dargestellten Evaluationsergebnisse zeigen verhältnismäßig positive Ergebnisse in den Bereichen Transparenz des Zustandekommens von Einzel- und Gesamtnoten und Termintransparenz; deutlich skeptischer äußern sich die Schüler/innen aber, was die Lernberatung durch die Lehrkräfte und vor allem die Antwort auf die Frage betrifft, wofür der jeweilige Lernstoff benötigt wird. Anlass zu Besorgnis geben die eklatanten Wahrnehmungsunterschiede zwischen Schüler/inne/n und Lehrer/inne/n. Letztere beurteilen ihre eigene Praxis ganz erheblich positiver als die Schüler/innen.

**LFLB in der „Neuen Mittelschule“**

Die Versuche zur „*Neuen Mittelschule*“ sehen im Wesentlichen eine Erweiterung der gesetzlich vorgesehenen Beurteilung mit Ziffernnoten durch verbale Zusätze und Ergänzungen vor, die in Form eines „Begabungsprofils“ ausgearbeitet werden und eine Grundlage für die Rückmeldung an Schüler/innen und Eltern sowie für eine Schullaufbahnberatung bilden sollen. Einige Standorte beabsichtigen, spezielle Konzepte für motivierende Leistungsrückmeldungen zu entwickeln, einige verzichten auf Klassenwiederholungen. Das Konzept der Europaschule Linz beispielsweise sieht – ähnlich wie jenes der Praxishauptschule der PH Klagenfurt – vor, in den ersten drei Jahren auf Noten vollständig zu verzichten und statt dessen zweimal jährlich Gespräche mit den Eltern über Leistungsstand und -entwicklung sowie Fördermaßnahmen zu führen. Zusätzlich erhalten die Schüler/innen am Ende jedes Semesters ein „Dokument der Lern-, Leistungs- und Sozialentwicklung“.

**Teilstandardisierte Matura**

Nach dem Konzept einer *teilstandardisierten Matura*<sup>9</sup>, dem derzeit noch die gesetzliche Grundlage fehlt, soll die Reifeprüfung an AHS künftig aus drei Teilen bestehen: einer schriftlichen vorwissenschaftlichen Arbeit, einem schriftlichen Teil mit zentral vorgegebenen Aufgaben in vier Fächern und einem schul- bzw. klassenspezifischen mündlichen Teil in zwei Fächern. 2008 fand in Form eines Pilotprojekts für Englisch und Französisch (Hörverständnis und Leseverständnis) an 57 AHS ein erster Durchgang statt. Eine Ausweitung des Projekts auf BHS und weitere Fächer ist im Gange.

9 Vgl. näher <http://www.bife.at/standardisierte-reifepruefung>



Die Lehrer/innen sind in das Projekt aktiv eingebunden, insofern ihnen die Auswertung der Aufgabenlösungen übertragen wurde. Um die Objektivität zu gewährleisten, stehen detaillierte Handreichungen und eine Hotline zur Klärung strittiger Fragen zur Verfügung. Man erhofft sich aus der Einbindung der Lehrer/innen einen positiven Backwash-Effekt auf die Kultur der Leistungsbeurteilung insgesamt und damit eine Verbesserung des Unterrichts.<sup>10</sup>

## 2.4 Alternative Formen der Leistungsbeurteilung

In der Praxis kommen, insbesondere in Schulversuchen an Volksschulen (§ 78a SchUG), hauptsächlich drei alternative Formen der Leistungsbeurteilung vor: verbale Beurteilung, lernzielorientierte Beurteilung sowie Beurteilung direkter Leistungsvorlagen. Zusätzlich existieren Modifikationen bestehender Formen der Leistungsfeststellung.

Bei der *verbalen Beurteilung* wird die Ziffernote durch eine qualitative Leistungsbeschreibung ergänzt oder ersetzt. Diese soll Aufschluss geben über den Grad der Lernzielerreichung, die Art und Weise, in der Aufgaben bearbeitet wurden, die dabei sichtbar gewordenen charakteristischen Fähigkeiten und Arbeitsweisen und insbesondere auch den Lernzuwachs. Diese Beurteilungsform stellt hohe Anforderungen an die diagnostischen und sprachlichen Fähigkeiten der Lehrpersonen, ihr Reformpotenzial ist wissenschaftlich umstritten (Sacher 2001: 141–148).<sup>11</sup> Sie hat ihren Schwerpunkt an Volksschulen, insbesondere in der Grundstufe I. Ein entsprechender Schulversuch wird an ca. 100 Volksschulen Wiens geführt (vgl. Grubich-Müller et al. 2008: 23–24), aber auch in anderen Bundesländern ist verbale Beurteilung in der Grundschule möglich und durch entsprechende Erlässe geregelt.

*Lernzielorientierte Beurteilung* macht den Lernfortschritt auf Basis von Lernziellisten sichtbar. Im Laufe eines Schuljahres entsteht ein Kompetenzprofil, das als Basis für Gespräche zwischen Lehrpersonen und Schüler/innen bzw. Eltern dienen kann. Ein wichtiges Ziel ist dabei die Förderung der Selbsteinschätzungskompetenz und Eigenverantwortung der Schüler/innen. Erfahrungen mit dieser Art von Leistungsbeurteilung stammen vor allem aus der Montessori- und der Integrationspädagogik. Mit entsprechend kindgemäß formulierten Zielen kommen „Pensbuch“ und „Studienbuch“ inzwischen auch in Volksschulen zur Anwendung (für Wien vgl. Grubich-Müller et al. 2008: 22–23). In der Modellschule Graz erhalten die Schüler/innen (bzw. Eltern) am Ende des Schuljahres statt eines Notenzeugnisses ein Bündel von Lernziellisten („Lernzielbuch“), in denen das Ausmaß der Zielerreichung für die einzelnen Fächer ausgewiesen ist (vgl. Michaelis 1992). Neben fachlichen Lernzielen werden mit den Schülern gemeinsam auch Verhaltensziele für jede einzelne Klasse bzw. Stufe erarbeitet. Ihre Erreichung wird von Schüler/innen und Lehrer/innen eingeschätzt.

Die *direkte Leistungsvorlage* macht die konkreten Leistungen der Schüler/innen direkt sichtbar und „verdichtet“ diese nicht durch Noten. In der Regel entstehen Sammelmappen (Portfolios) mit Produkten der Schüler/innen, die als Leistungsnachweis und Basis für Gespräche zwischen den Lehrpersonen und den Schüler/innen bzw. Eltern dienen. Eine direkte Leistungsvorlage ist aber auch gegeben, wenn Schüler/innen konkrete Kompetenzen in flüchtiger Form demonstrieren, zum Beispiel Volksschüler/innen ihre Lesekompetenz dadurch zeigen, dass sie ihren Eltern vorlesen. Die in Österreich vor allem von Vierlinger (1999) entwickelte und propagierte direkte Leistungsvorlage und verwandte Formen wie „Lernfortschrittsdokumentationen“ werden inzwischen vor allem im Grundschulbereich häufig als Alternative herangezogen (vgl. Grubich-Müller et al. 2008: 17–18).

10 Mündliche Mitteilung von Projektleiter Lucyshyn, August 2008; dazu vgl. auch die Argumente von Fend, 2008: 98 und 356.

11 Bei der Untersuchung von Wagner/Valtin, die keine Vorteile verbaler Beurteilung gegenüber Ziffernnoten belegen konnte, wurde leider nicht sicher gestellt, dass die verbale Beurteilung von den Lehrkräften im Sinne des Konzepts praktiziert wurde (Wagner/Valtin 2003: 33 f.); hingegen berichtet Tarnai Vorteile der verbalen Beurteilung hinsichtlich der angewandten Bezugsnormen und der steigenden Akzeptanz bei Eltern (Tarnai 2006).

### Verbale Beurteilung

### Lernzielorientierte Beurteilung

### Direkte Leistungsvorlage

**Modifikationen der Schularbeiten**

Vor allem Hauptschulen haben in der 3. Leistungsgruppe die Anzahl der *Schularbeiten* reduziert oder überhaupt auf solche verzichtet, um die Schüler/innen zu entlasten (Turic 2001). Hinsichtlich der Arbeitsweise gibt es Versuche mit einer „Zweiphasenschularbeit“ (Blüml 1994); den verschiedenen Varianten ist gemeinsam, dass Schüler/innen Aufgaben (z. B. ein Aufsatzthema) nicht in einem Zug bearbeiten, sondern nach einer ersten Phase unterstützende Kommentare von den Lehrer/inne/n einholen oder sich anderweitige Hilfe organisieren können. Danach verbessern sie ihre Arbeit in einer kürzeren zweiten Phase. Die Zweiphasenschularbeit fördert die Selbstbeurteilungskompetenz und hat vielfach eine höhere ökologische Validität als eine unter strikter Zeitbegrenzung geschriebene Arbeit.

**Beurteilung von Schlüsselkompetenzen**

Bislang wenig Erfahrungen und brauchbare Instrumente gibt es für die Beurteilung fachübergreifender *Schlüsselkompetenzen*. Bei der Erprobung der Wiener Bildungsstandards wurden in zwölf Klassen der 6. bis 8. Schulstufe auch „dynamische Fähigkeiten“ sowie das Arbeitsverhalten erfasst. Zur Überprüfung der Lernfortschritte wurden (I-)Can-do-Statements (z. B.: „Ich kann mir selbstständig Informationen beschaffen.“; „Ich verstehe es, bei Konflikten zu vermitteln.“) entwickelt, die sowohl von Schüler/inne/n als auch von Lehrer/inne/n (übrigens recht unterschiedlich) eingeschätzt wurden (Annau et al. 2005). Die K.-R.-Popperschule in Wien bietet zur Ergänzung der forcierten Unterstützung im kognitiven Bereich das Fach „Kommunikation und soziale Kompetenz“ an und hat dafür ein eigenes Curriculum entwickelt, das auch Hinweise auf Möglichkeiten der Überprüfung der Lernfortschritte enthält (Wustinger/Braun 2008). An zahlreichen anderen Schulen gibt es Lernangebote im sozialen Bereich (z. B. Konfliktmanagement), die mit Zertifikaten bestätigt werden können.

**Neue Rückmeldeverfahren**

Veränderungen bei den *Rückmeldeverfahren* zielen häufig darauf ab, Schüler/innen zunächst zu einer Selbsteinschätzung zu veranlassen und diese dann mit der Fremdeinschätzung durch die Lehrpersonen zu konfrontieren. Dieser Idee folgen der im Volks- und Hauptschulbereich häufig verwendete „Feedbackstern“, aber auch die „Kindersprechtag“/„Kinder-Eltern-Lehrer-Gespräche“, an denen ein ausführliches Gespräch zwischen Lehrpersonen und Schüler/inne/n über deren Leistungen stattfindet. Beide Formen beziehen neben den Fachleistungen auch die personale Entwicklung und soziale Kompetenzen in die Rückmeldung mit ein.

### 3 Leistungsfeststellung und Leistungsbeurteilung im Spiegel der Forschung

#### 3.1 Validitätsaspekte

##### 3.1.1 Inhaltliche Validität

Inhaltliche Validität bedeutet, dass die bei Leistungsmessungen gestellten Aufgaben die gesamte Bandbreite der interessierenden Kompetenz- und Inhaltsbereiche angemessen repräsentieren.

**Leistungsfeststellung im Unterricht**

Für Österreich ist zu vermuten, dass aufgrund der großen Anzahl der Leistungsfeststellungen und der Forcierung des stofflichen Lernens die Breite der Inhalte angemessen abgedeckt wird; konkrete Untersuchungen dazu liegen jedoch nicht vor. Hinsichtlich der Komplexität und des Niveaus der Aufgaben zeigen vereinzelte Untersuchungen, dass vor allem die schriftlichen Überprüfungen häufig kognitive Aufgaben auf einem relativ niedrigen taxonomischen Niveau umfassen (Buschmann 2001), das Niveau der Messung also unter dem Niveau des Unterrichts oder jedenfalls dem der Lehrpläne bleibt.

**Externe Tests**

Leistungsfeststellungen im Gesamtsystem beschränken sich bis jetzt auf wenige Domänen (Lesen, Mathematik, Naturwissenschaften in PISA bzw. TIMSS; Deutsch, Englisch und Mathematik bei den Bildungsstandards), folgen teilweise einer anderen didaktischen Logik als

Lehrpläne und Unterricht und zeichnen auf diese Weise kein repräsentatives Bild der Leistungen des Gesamtsystems.

### 3.1.2 Auswertungsvalidität

Auswertungsvalidität bezieht sich darauf, dass bei der Aus- und Bewertung von Prüfungen Modelle umgesetzt werden, die den zu messenden Konstrukten angemessen sind und das Messergebnis nicht unangemessen verzerren oder verschleiern. Dies betrifft sowohl die Umsetzung der bestehenden rechtlichen Grundlagen als auch fachdidaktische Vorstellungen von Lernzieltaxonomien und Kompetenzmodellen. Rückschlüsse zu diesem Validitätsaspekt sind einerseits aus den Notenverteilungen möglich (a). Andererseits ist zu fragen, welche Beschränkungen der Aussagekraft durch die Ziffernnote an sich gegeben sind (b).

(a) Mit Ausnahme der Volksschule<sup>12</sup> zeigen sich in allen Schultypen und auf allen Schulstufen annähernd *Normalverteilungen*. Die lässt vermuten, dass entgegen den gesetzlichen Bestimmungen wesentlich auch sozial vergleichend beurteilt wird (vgl. Abb. 1).

Soziale statt kriteriale Norm

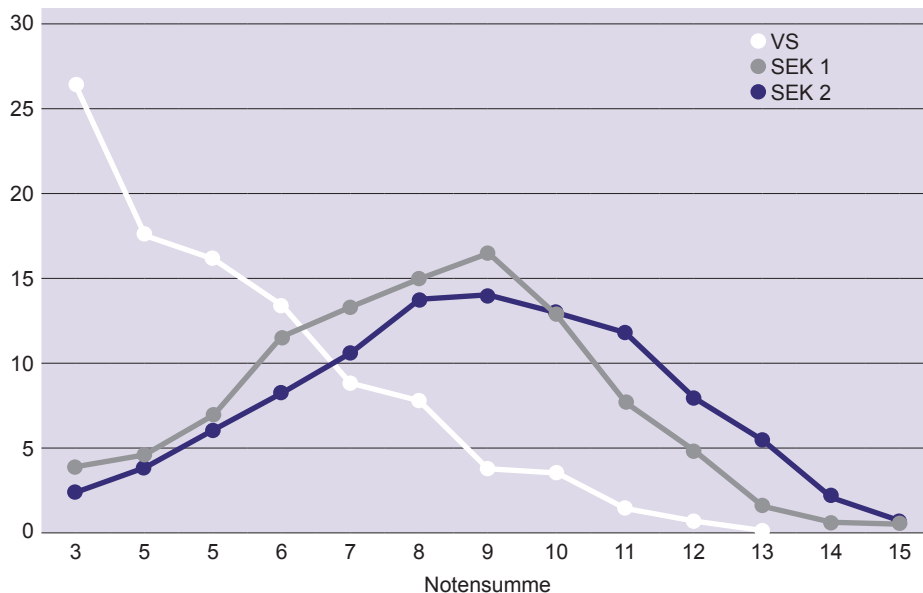


Abb. 1: Notenverteilungen auf den verschiedenen Ebenen des Schulsystems (Volksschule, Sekundarstufe I, Sekundarstufe II)

Anmerkung: Eingetragen ist die Notensumme aus Deutsch, Mathematik und Englisch (in der Volksschule: Sachunterricht). Eine Notensumme von 3 bedeutet, dass ein Kind in den drei genannten Gegenständen jeweils mit „sehr gut“ beurteilt wurde; 4 bedeutet, dass in einem Gegenstand ein „Gut“ vorliegt, usw.

Quelle: Befindensuntersuchung 2005, N = 7.625 (Eder 2007: 69)

(b) Während die LBVO für die einzelnen Notenstufen qualitativ unterschiedliche Leistungen einfordert (vgl. Abschnitt 2.1.1), bildet die gängige Praxis lediglich quantitative Unterschiede ab, d. h. selbst wenn Aufgaben mit unterschiedlichem Anforderungsniveau gestellt werden, erhalten „schwierige“ Aufgaben bei der Auswertung allenfalls mehr Punkte, begründen aber keine qualitativ andere Einstufung. Allerdings fehlen auch dazu systematische Untersuchungen.

Grenzen der Aussagekraft von Ziffernnoten

Zu fragen ist, ob Ziffernnoten überhaupt in der Lage sind, das Leistungsniveau von Schüler/innen in Hinblick auf die (impliziten) Kompetenzmodelle einzelner Fächer angemessen abzubilden. Ansätze für angemessenere Darstellungen finden sich in den internationalen Leis-

<sup>12</sup> Die unverkennbar andere Form der Notenverteilung in der Volksschule legt nahe, dass eine kriteriale Bezugsnorm angewandt wird; es ist aber auch denkbar, dass einfach auf einem „milderen“ Niveau sozial vergleichend beurteilt wird.

tungsvergleichen, die für die einzelnen Schüler/innen Kompetenzstufen ausweisen (obwohl auch diese mit einer gewissen Willkürlichkeit über quantitative Grenzen definiert sind).

### 3.1.3 Externe Validität

Externe Validität orientiert sich an Außenkriterien. Während zu relevanten Außenkriterien (z. B. Ergebnisse objektiver Tests, spätere Erfolge in Ausbildung und Beruf) enge Zusammenhänge bestehen sollten, sollten umgekehrt keine systematischen Zusammenhänge mit irrelevanten Merkmalen (z. B. Wohnort eines Schülers) bestehen. Zu diesem Aspekt erscheinen folgende Ergebnisse von Bedeutung:

#### Übereinstimmungsvalidität

(a) Der Zusammenhang zwischen von außen gemessenen Leistungen und den Noten der Schüler/innen ist nicht hinreichend hoch.

Analysen an österreichischen Daten ergeben ein differenziertes Bild. Eder (2003: 134) fand bei Schüler/innen der Sekundarstufe I immerhin eine Korrelation von  $r = -.69$  zwischen Testleistung und Note, zu erwartende Unterschiede (z. B. nach Geschlecht, Schultyp) werden von Leistungstests aber differenzierter abgebildet als von Noten. Analysen auf Basis der PISA-Daten 2003 wiederum zeigten für Schüler/innen der AHS eine außerordentlich große Varianz der Leistungen innerhalb der einzelnen Notenstufen. Von den Schüler/innen mit „Nicht genügend“ in Mathematik entfielen 7% auf die höchsten Kompetenzstufen 5 und 6, 32% auf die Kompetenzstufe 4. Umgekehrt entfielen von den Schüler/innen mit „Sehr gut“ 10% auf die Kompetenzstufen 0 bis 2. In jeder Notenstufe waren alle Kompetenzlevels vertreten (Haider/Schreiner 2006: 233). Analysen auf Basis der Pilot-Testungen für die Bildungsstandards ergaben in Mathematik und in Englisch sehr niedrige Korrelationen zwischen Testleistungen und Noten innerhalb der einzelnen Differenzierungsgruppen der Sekundarstufe I. Die relativ höchsten Korrelationen (bis  $r = -.36$ ) finden sich in der AHS, in der 3. Leistungsgruppe zeigen sich praktisch durchgehend kaum noch nennenswerte Zusammenhänge zwischen Note und gemessener Leistung; die höchste Korrelation betrug hier  $r = -.15$  (Datensatz Pilot-Untersuchungen Standards 2007).

Diese Ergebnisse finden zum Teil eine Erklärung darin, dass in das Lehrerurteil zahlreiche andere Faktoren eingehen, insbesondere die bei den Schüler/innen vermutete Anstrengung, aber auch Disziplinierungs- und Motivierungsmotive, Fehler bei der Bildung von Gesamtnoten, Vorurteile usw. (Ingenkamp 1985; vgl. auch Neuweg 2006b: 118–121). Es ist daher nicht zu erwarten, dass Noten das Leistungsniveau der Schüler/innen angemessen wiedergeben (vgl. Bos et al. 2007: 281). Zugleich erscheinen aber einmalige, objektive Messungen, wie sie mit Schulleistungstests erfolgen, weniger reliabel und valide als Beurteilungen, die sich auf eine Vielzahl von Beobachtungen stützen.

#### Prognosevalidität in Volks- und Hauptschule

(b) Die prognostische Qualität der Beurteilung in Volks- und Hauptschule ist gering.

Nach einer Längsschnittuntersuchung von Eder und Thonhauser (2006) erlangten Volksschüler/innen mit AHS-Reife, die in eine ländliche Hauptschule übertraten, am Ende der Sekundarstufe I nur zu 70% die Berechtigung zu einem Übertritt in höhere Schulen; andererseits wurde dieses Ziel auch von 30% der Volksschüler/innen erreicht, die nicht über die AHS-Reife verfügten. Haider/Schreiner (2006) zeigten auf Basis der PISA-Leistungen, dass von den ehemaligen Schüler/innen der 3. Leistungsgruppe mehr als 5% Leistungen über dem PISA-Mittelwert erbrachten und es zu erheblichen Überschneidungen mit den Schüler/innen aus der AHS-Unterstufe kommt: „So erreichen die rund 10% besten Schüler/innen aus der untersten, der 3. Leistungsgruppe der Hauptschule mit Noten von Befriedigend und darunter vergleichbare Leistungen im PISA-Test wie die untersten 10% in den AHS/HS 1. Leistungsgruppe mit der Note Sehr gut (!)“ (Haider/Schreiner 2006: 235).

(c) Es gibt deutliche Geschlechtsunterschiede in den Noten, die allerdings in den gemessenen Leistungen nicht immer eine Entsprechung finden.

Geschlechterunterschiede

Mädchen haben auf allen Ebenen und Stufen des Schulsystems zumindest in den Schularbeitsfächern Deutsch, Englisch und Mathematik etwas bessere Noten (vgl. Eder 2007: 70 und bereits Eder 1995: 85). In den PISA-Mathematikleistungen zeigen sich jedoch im Gegenteil leichte Vorsprünge der Burschen (Bacher/Paseka 2006: 221). In den von Eder et al. (2002) berichteten Analysen mit Matkomp I zeigten sich bei den direkt gemessenen Mathematikleistungen deutliche, in den von den Lehrpersonen erbetenen Kompetenzbeurteilungen nur geringe Unterschiede zu Gunsten der Burschen, in den Noten sogar Vorteile für die Mädchen. Vergleichbar differenzierte Analysen zu anderen Fächern liegen nicht vor.

(d) Regionale Gesichtspunkte können einen erheblichen Einfluss auf die Praxis der Notengebung haben.

Regionale Aspekte

Eder/Thonhauser (2006: 281) zeigten, dass Veränderungen in der Notengebung der Volksschule in den letzten zehn Jahren fast ausschließlich in den Städten erfolgten. Hier entwickelt sich offenbar massiver Druck auf die Lehrpersonen, den Kindern nicht durch schlechte Noten den Zugang zur AHS zu erschweren. Tab. 2 bestätigt die Tendenz zu guten Noten in den Städten auf Basis von Daten aus PIRLS 2006 (Thonhauser/Pointinger, im Druck; für Deutschland vgl. Trautwein et al. 2007).

Größe des Schulstandortes	Fächer		
	Deutsch	Mathematik	Sachunterricht
< 3.000	27	39	59
3.000 – 15.000	26	37	59
15.000 – 500.000	29	41	60
> 500.000 (= Wien)	35	49	67

Tab. 2: Prozentanteil der Note „Sehr gut“ in 4. Volksschulklassen nach Größe des Schulstandortes

### 3.2 Entsprechung mit sozialen Werten

Leistungsmessungen unterliegen einem erhöhten Anspruch auf Realisierung sozialer Werte, insbesondere Gerechtigkeit, Fairness, Vergleichbarkeit und Transparenz. Untersuchungen belegen, dass die gegenwärtige Praxis den diesbezüglichen Erwartungen nicht in ausreichendem Maße entspricht.

(a) Die Noten sind in jenen Schulfächern am schlechtesten, für die die meiste Unterrichts- und Lernzeit verwendet wird (Spiel/Wagner 2002: 336–338). Tab. 3 zeigt Notenverteilungen für Mathematik und naturwissenschaftliche Fächer in der 7. und 8. Schulstufe.

Lernaufwand und Noten

	HS				AHS			
	Mathematik	Biologie	Geographie	Physik/ Chemie	Mathematik	Biologie	Geographie	Physik
Sehr gut	7	31	28	22	10	38	32	32
Gut	27	30	29	29	25	31	31	30
Befriedigend	40	26	27	30	34	21	24	24
Genügend	23	12	15	16	26	8	12	12
Nicht genügend	3	2	2	3	5	2	2	2
N	3.302	3.280	3.282	3.271	2.524	2.494	2.523	2.527

Tab. 3: Notenverteilungen in Mathematik und in den naturwissenschaftlichen Fächern

Anmerkung: Eingetragen ist der Prozentsatz der Schüler/innen auf den einzelnen Notenstufen.

Datenquelle: TIMSS 1995, Population 2. Die Daten umfassen die 7. und 8. Schulstufe.

Mathematik ist auf diesen Schulstufen mit vier bis fünf Wochenstunden dotiert und erfordert als Schularbeitenfach sowie durch regelmäßige Hausübungen erhebliche häusliche Lernzeit. Die Notenverteilung ist jedoch deutlich ungünstiger als beispielsweise in Biologie und Umweltkunde, das in der Regel zwei Wochenstunden umfasst und wenig häusliche Lernzeit erfordert. Ähnlich ungünstige Notenverteilungen wie Mathematik weisen Englisch und – in etwas abgemilderter Form – Deutsch auf (Datensatz Eder 2007).

Trotz steigenden Lernaufwandes der Schüler/innen kommt es während der Schullaufbahn zu einer kontinuierlichen Verschlechterung der Noten (vgl. Abb. 2).

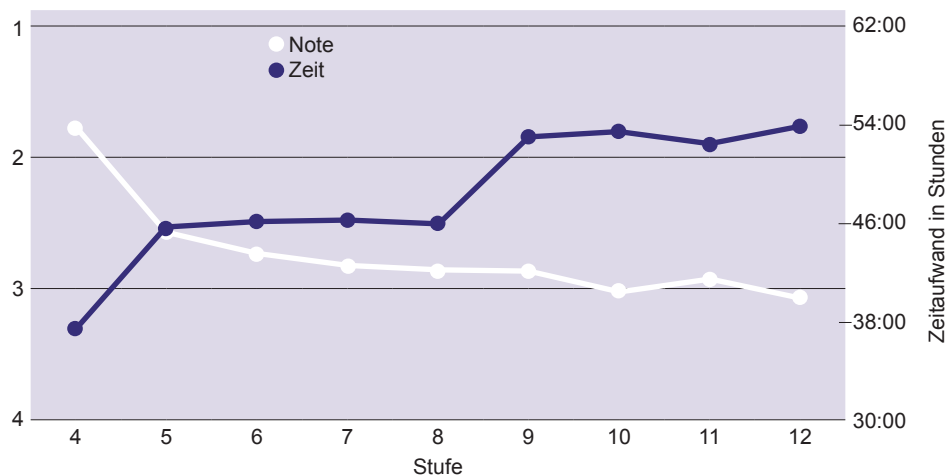


Abb. 2: Zeitaufwand für die Schule und Notendurchschnitte

Anmerkung: Eingetragen ist der durchschnittliche Zeitaufwand der Schüler/innen in Stunden sowie der Notendurchschnitt<sup>13</sup> aus Deutsch, Englisch (in der Volksschule: Sachunterricht) und Mathematik.

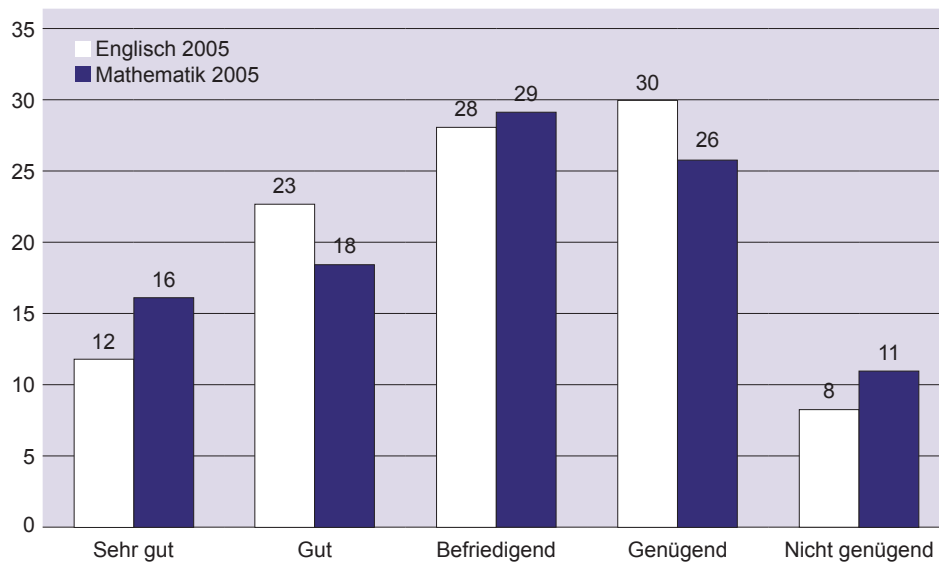
Quelle: Datensatz der Befindensuntersuchung 2005 (Eder 2007). Die Daten beziehen sich auf Schüler/innen in Vollzeitschulen.

Auffallend sind die abrupten Veränderungen beim Übergang auf die Sekundarstufe I. Es ist zu vermuten, dass der hier sichtbar werdende Mechanismus – steigender Aufwand bei zunehmend schlechteren Noten – eine wesentliche Ursache für die ebenfalls in der Sekundarstufe I massiv einsetzende Schulunlust darstellt.

13 Obwohl Noten in der Regel rangskaliert sind, scheint angesichts der in Abbildung 1 dargestellten Normalverteilungen die Berechnung eines Durchschnittswertes zulässig, um Trends in Datenverläufen sichtbar zu machen.

(b) Es gibt – im Vergleich zu den Fähigkeiten der Schüler/innen – vor allem in den höheren Schulen erstaunlich viele schlechte Noten. Insgesamt besuchen beispielsweise in Österreich derzeit etwa 20% eines Altersjahrgangs die Oberstufe einer AHS. Auf Basis mehrerer Untersuchungen (z. B. Eder 1998, 2001) lässt sich die durchschnittliche Intelligenz dieser Gruppe auf  $IQ = ca. 115$  schätzen, d. h. sie repräsentiert ungefähr das oberste Drittel der kognitiven Leistungsfähigkeit der Alterskohorte. Die Notenverteilung bspw. für Englisch (vgl. Abbildung 3) vermittelt jedoch den Eindruck einer mehrheitlich wenig leistungstüchtigen Gruppe. In Mathematik liegen die Verhältnisse nur geringfügig besser.

**Gute Noten als knappes Gut**



**Abb. 3: Notenverteilung in Englisch und Mathematik, Oberstufe AHS**

Quelle: Datensatz der Befindensuntersuchungen 2005.

Schließt man nicht nur unzureichende Intelligenz, sondern auch mangelhaften Unterricht als Ursache für derartige Verteilungen aus, so liegt die Vermutung nahe, dass Prüfungsaufgaben oder Beurteilungskriterien zumindest unbewusst so kalibriert werden, dass sie zwingend eine gewisse Misserfolgsquote nach sich ziehen.

(c) Zwischen einzelnen Einheiten des Bildungssystems bestehen oft extreme Unterschiede in den erbrachten Leistungen, nicht aber in ihrer Bewertung. Abb. 4 zeigt die Gegenüberstellung von zwei Gymnasialklassen, in denen mathematische Kompetenzen mit einem auf TIMSS 1995 aufbauenden Instrument getestet wurden. Die schlechtere Klasse liegt um mehr als zwei Lernjahre zurück, in der Notenverteilung zeigen sich jedoch praktisch keine Unterschiede. In einer Hauptschulklasse mit einem vergleichbaren Niveau wären etwa ein Drittel der Schüler/innen in der 3. Leistungsgruppe.

**Ungleiche Leistungen, gleiche Noten**

B

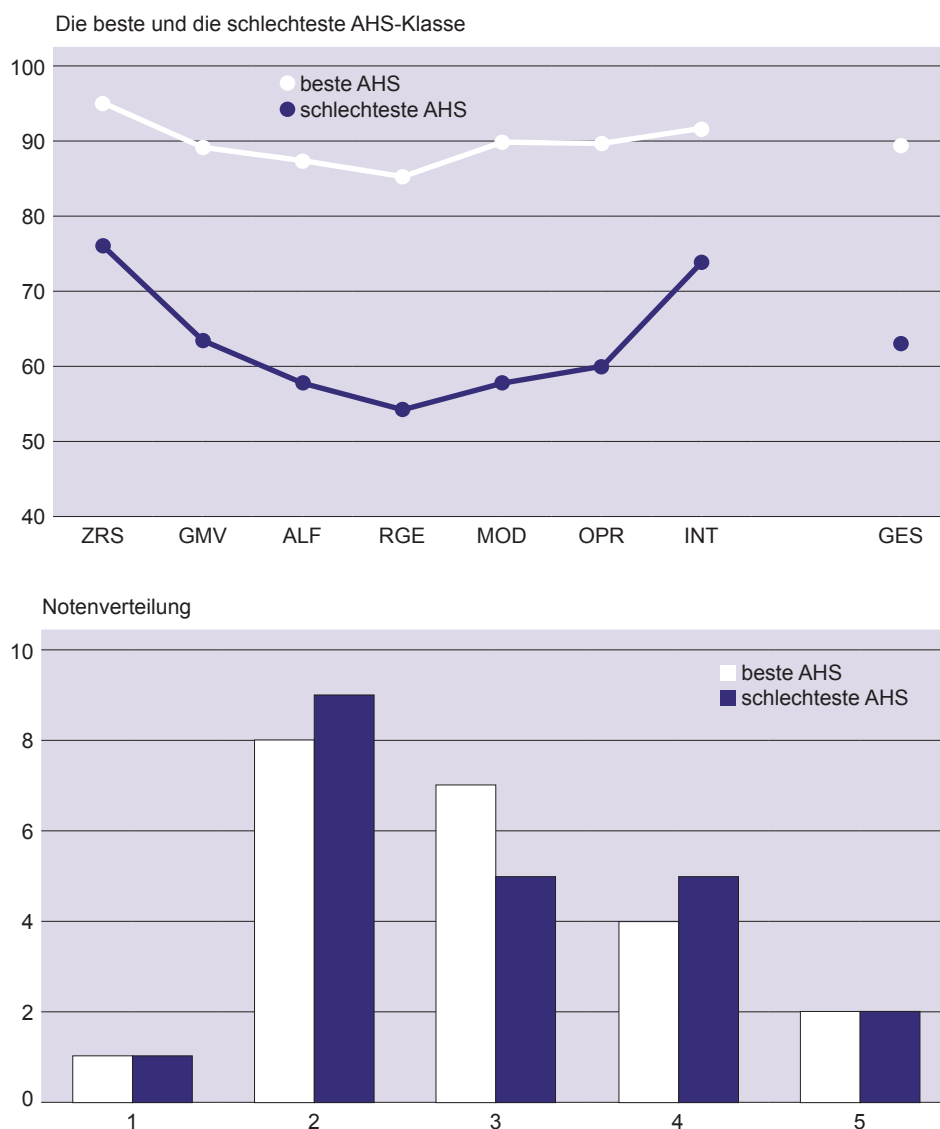


Abb. 4: Gemessene vs. bewertete Leistungen.

Anmerkung: Die obere Abbildung enthält durchschnittliche Leistungsprofile für mathematische Kompetenzen aus zwei Klassen der 7. Schulstufe (3. Klasse) AHS. Die untere Abbildung zeigt ihre Notenverteilungen in Mathematik.

Abkürzungen: ZRS: Zahlenverständnis, Rechnen, Statistik; GMV: Größen, Maße, Verhältnisse; ALF: Algebra und Funktionen; RGE: Raumvorstellung und Geometrie; MOD: Modellbildung; OPR: Operieren, Rechnen; INT: Interpretieren; GES: Gesamtwert. Die angeführten Prozentpunkte besagen, wie viele Punkte in einem Aufgabenbereich erreicht wurden.

### 3.3 Effizienz und Kosten

#### Ebene Unterricht

Leistungsmessungen sollen ein sinnvolles Verhältnis von Ressourceneinsatz und Ertrag aufweisen. Wegen der für das Schulsystem charakteristischen geringen Sichtbarkeit von Kosten sind tiefer greifende Analysen derzeit ausgesprochen schwierig.

Das in Österreich praktizierte System der Beurteilung innerhalb des Unterrichts ist grundsätzlich kostengünstig. Ohne besonderen Zusatzaufwand zum laufenden Unterricht kann eine Klassifikation der Schüler/innen auf Basis von Ziffernnoten erstellt werden. Diesen kommt trotz ihrer Schwächen eine gewisse Aussagekraft zu. Baron-Boldt et al. (1988) berichten als Ergebnis einer umfangreichen Meta-Analyse, dass der prognostische Wert von kombinierten



Schulnoten für den Ausbildungs- und Studienerfolg mit einer Korrelation von  $r = .37$  in einer vergleichbaren Höhe liegt wie jener von Intelligenztests. Insgesamt ist anzunehmen, dass die Schulnoten aufgrund ihrer einfachen Erfassung auch weiterhin einen gewissen Stellenwert beim Zugang zu Berufen behaupten werden.

Laufende Entwicklungen in Schule, Gesellschaft und Wirtschaft zeigen andererseits jedoch, dass die Klassifikation auf der Basis von Noten zunehmend weniger nachgefragt wird:

- Einschränkung des Repetierens und zunehmende Förderorientierung in der Schule reduzieren die Berechtigungsfunktion von Noten und lassen die Häufigkeit von Leistungsbeurteilungen in Schuljahren ohne anschließenden Übergang in ein anderes System als fragwürdig erscheinen.
- In der Wirtschaft zählen Schulnoten zwar zu den standardmäßig im Rahmen von Personalausleseverfahren erhobenen Daten, genutzt werden sie jedoch meist nur im Rahmen der Vorselektion von Bewerbern.
- Im Gegensatz zu Deutschland, wo mit dem Numerus clausus der Zugang zur Hochschulbildung explizit an ein Kalkül aus Schulnoten geknüpft ist, spielen sie in Österreich dafür praktisch keine Rolle. In den wenigen Fächern, in den Zulassungsbeschränkungen gegeben sind, werden entweder standardisierte Tests (z. B. Medizin) oder fachinhaltlich ausgerichtete Zulassungsprozeduren durchgeführt (z. B. Psychologie, Kommunikationswissenschaft); sehr gute Noten bringen in diesem Zusammenhang bestenfalls Bonuspunkte.

Die Kosteneffizienz von Maßnahmen des Systemmonitorings lässt sich derzeit schwer beurteilen, weil sich dieser Bereich erst in der Aufbauphase befindet. Mit der Schaffung gesonderter Einrichtungen zur laufenden Durchführung von Leistungsmessungen ergibt sich jedenfalls ein nicht unerheblicher Kostenfaktor, der erst dann gerechtfertigt ist, wenn durch die Rückführung von Ergebnissen eine Steigerung der Effizienz des Unterrichts erreicht wird.

**Systemmonitoring**

### 3.4 Die österreichische Leistungsbeurteilung in internationaler Perspektive

Im internationalen Vergleich nimmt Österreich mit der praktisch vollständigen Durchführung von LFLB durch die gleichen Personen, die auch den Unterricht erteilt haben, eine Extremposition ein; die kommissionelle Abwicklung der Matura ist, von Ausnahmefällen abgesehen (z. B. Beisitz bei Wiederholungsprüfungen, kommissionelle Prüfung im Berufungsverfahren) die einzige Situation, in der nicht die jeweils unterrichtenden Lehrer/innen allein für die Beurteilung zuständig sind. Selbst in Deutschland, das insgesamt ein relativ ähnliches System der Leistungsbeurteilung aufweist, existieren objektivierende Elemente (z. B. Zentralmatura). Die Randposition Österreichs wird auch daraus ersichtlich, dass sämtliche Kategorien des OECD-Indikators D5 („Wie werden Evaluationen und Leistungsmessungen in den Bildungssystemen eingesetzt?“) für Österreich als „nicht anwendbar“ ausgewiesen werden.<sup>14</sup>

**Objektivierung**

In Deutschland wurden zur Schaffung besserer Vergleichsmöglichkeiten in den letzten Jahren eine Reihe von Schulleistungstests mit schulnahen Aufgabenstellungen durchgeführt, um die direkte Umsetzung der Ergebnisse in die Verbesserung des Unterrichts zu unterstützen. Ein Beispiel dafür ist das Projekt VERA (Vergleichsarbeiten in der Grundschule), eine flächendeckende Lernstandserhebung in Mathematik und Deutsch auf der 3. (ursprünglich 4.) Schulstufe, die inzwischen vom Großteil der deutschen Bundesländer durchgeführt wird. Die teilnehmenden Schulen erhalten eine internetbasierte Rückmeldung, die sie für die Weiterentwicklung des Unterrichts nützen können. VERA verfolgt mehrere Ziele: Standort-

**Schulleistungstests**

<sup>14</sup> Vgl. OECD (2006). *Education at a Glance. Annex 3: Sources, methods and technical notes. Chapter D: The learning environment and organisation of schools*. Internet: <http://www.oecd.org/dataoecd/54/30/37382809.pdf> [Stand 2009-02-20].

bestimmung und Qualitätssicherung, Förderung diagnostischer Kompetenzen der Lehrer/innen und damit Verbesserung der Unterrichtsqualität, Stimulierung gezielter Förderprojekte und fundierte Schullaufbahnberatung. Die Aufgaben orientieren sich an Kompetenzstufen, die Ergebnisse werden für die unterschiedlichen Ebenen des Schulsystems (Schüler/in, Klasse, Schule, Region, Bundesland) aggregiert und zurückgemeldet. Die Durchführung in der 3. Schulstufe eröffnet die Möglichkeit gezielter Interventionen noch vor Abschluss der Grundschule (Helmke/Hosenfeld 2004).

#### Externalisierung und Förderorientierung

In den bei PISA besonders erfolgreichen Ländern erfolgte schon in den 1970er Jahren eine Umstellung auf externe Formen der Leistungsbeurteilung sowie eine Verstärkung der diagnostischen und förderbezogenen Maßnahmen. Zahlreiche Länder (z. B. Finnland, Frankreich, England, Niederlande) führen am Ende der wichtigsten Abschnitte des Schulsystems (Grundschule, Sekundarstufe I, Sekundarstufe II) unterschiedlich organisierte, extern konzipierte und teilweise auch extern durchgeführte Leistungsfeststellungen durch, mit denen teilweise weitere Berechtigungen verbunden sind. Die meist gesamtschulartigen Systeme dieser Länder verzichten über relativ lange Zeiträume, in denen die Leistungsfeststellungen der Lehrpersonen praktisch ausschließlich förderdiagnostischen (formativen) Zielsetzungen dienen, auf eine selektive Beurteilung (Arbeitsgruppe „Internationale Vergleichsstudie“ 2003; vgl. auch Fend 2008: 356 f).

### 3.5 Forschungsfragen

Zu Fragen der externen Validität, der Realisierung sozialer Werte und ansatzweise auch zu den Folgewirkungen und Begleiterscheinungen von LFLB ist die Forschungslage vergleichsweise gut. Große Forschungslücken bestehen im Hinblick auf Fragen der inhaltlichen Validität, der Auswertungsvalidität und der Einbettung der Leistungsbeurteilung in den Unterricht.

Im Folgenden soll der Forschungsbedarf über die bereits erfolgten Hinweise hinaus exemplarisch illustriert werden:

- Wie reflektiert werden Prüfungsanforderungen definiert? Wie wird den Schüler/inne/n die objektive Bedeutsamkeit von Aufgaben vermittelt, so dass sie ihnen auch subjektiv als bedeutsam erscheinen?
- Wie eng sind Unterricht und Prüfungswesen tatsächlich aufeinander abgestimmt?
- In welchen Prüfungsformen, anhand welcher Aufgaben und unter welchen räumlichen, zeitlichen, materiellen und sozialen Rahmenbedingungen müssen bzw. dürfen die Schüler/innen ihre Leistungen demonstrieren? Inwiefern dürfen sie über thematische Schwerpunkte, Formen oder Rahmenbedingungen der Leistungserhebung mitentscheiden?
- Wie gehen Lehrer/innen bei der Zusammenstellung von Aufgaben, bei ihrer Auswertung und bei der Gewichtung von Teilergebnissen vor? Welche Vorstellungen von den zu messenden Kompetenzen haben sie und wie setzen sie diese in Prüfungsaufgaben um?
- Wie werden Schüler/innen durch individuelle Rückmeldungen unterstützt? Welche weiteren Unterstützungsangebote stehen ihnen zur Verfügung? Wie wirken außerschulische Nachhilfe und Unterricht zusammen?
- Inwieweit fördert die Praxis der Leistungsfeststellung und -beurteilung die Kompetenz der Schüler/innen zur sachlich begründeten Selbstbeurteilung?
- Wie werden in Übergangs- bzw. Anfangssituationen die von den Schüler/inne/n mitgebrachten Fähigkeiten und Einstellungen festgestellt, um das Entwicklungspotenzial durch Unterricht abschätzen zu können? Diese Frage ist bereits bei der Einschulung von Bedeutung, weil zu diesem Zeitpunkt am ehesten korrigierende Maßnahmen mit weitreichenden Effekten für die Bildungslaufbahn gesetzt werden können (Wößmann 2007: 139–140).

- Nach welchen subjektiven Theorien wird die Leistungsbeurteilung in den sog. „Nebenfächern“ gestaltet? Wie sind die gegenüber den „Hauptfächern“ offenbar ganz unterschiedlichen Beurteilungspraktiken zu erklären?
- Wie erfolgt eine erfolgreiche Klassenführung, wenn das Motivierungs- und Drohpotenzial von Noten wegfällt? Wie gelingt es Lehrpersonen in anderen Ländern, Schüler/innen ohne Notendruck zum Lernen zu motivieren?

## 4 Optionen für weitere Entwicklungen

Für eine evidenzbasierte Formulierung von Entwicklungsoptionen sollten vor allem zwei Quellen herangezogen werden: die Ergebnisse der Validitätsanalysen sowie Hinweise aus den bereits ansatzweise praktizierten alternativen Ansätzen zur Leistungsbeurteilung. Als besonders relevante Analyseergebnisse erscheinen hier

- die offensichtlichen Defizite der gegenwärtigen Praxis hinsichtlich der Vergleichbarkeit von Noten,
- die fehlende Umsetzung gesetzlich vorgesehener Auswertungsmodelle (vergleichende statt kriterienbezogener Beurteilung),
- die geringe Integration der Leistungsbeurteilung in den Unterricht und die damit verbundene mangelnde Nutzung von Rückmeldemöglichkeiten,
- die geringe Prognosekraft der Beurteilungen vor allem auf den unteren Ebenen des Schulsystems.

Die praktizierten Alternativansätze forcieren vor allem eine stärkere Ausrichtung auf die inhaltliche Validität der Leistungsfeststellungen, auf die Erfassung von Schlüsselkompetenzen, auf die Einbeziehung der Schüler/innen in den Beurteilungsprozess und die verstärkte Nutzung von Rückmeldemöglichkeiten.

Auf dieser Basis werden die folgenden Optionen zur Weiterentwicklung vorgeschlagen:

### 4.1 Externalisierung und Objektivierung

Für die Schul- und Unterrichtsentwicklung grundsätzlich sehr bedeutsam erscheint eine von den Lehrer/innen selbst organisierte „kollegiale Externalisierung“, bei der Teilfunktionen der Leistungsbeurteilung (z. B. Aufgabenstellung, Auswertung) von Kolleg/inn/en an der Schule wahrgenommen werden. Dadurch können Diskussionen über die Bedeutsamkeit von Prüfungsaufgaben, über das eigene Beurteilungsverhalten und über Möglichkeiten zur Verbesserung des Unterrichts stimuliert werden.

**Kollegiale  
Externalisierung**

Leistungsfeststellungen, mit denen Berechtigungen verbunden sind, müssen sich an einheitlichen Standards orientieren und in objektivierter Form durchgeführt werden. Das impliziert, dass die Aufgaben zentral entwickelt werden, ihre Vorgabe in standardisierter Form und nach Möglichkeit durch nicht involvierte Testleiter/innen geschieht und die Auswertung nach einheitlichen Richtlinien (Standards; definierte Kompetenzen) erfolgt. Dies würde einen erheblichen Beitrag zur Behebung regionaler und standortbezogener Unterschiede leisten und dazu beitragen, dass die Schüler/innen sachbezogene Rückmeldungen über ihr Können erhalten und so Fairness und Chancengleichheit im Bildungssystem deutlich erhöhen.

**Externalisierung i. e. S.**

Den unzweifelhaften Objektivierungsvorteilen einer (Teil-)Externalisierung der Leistungsbeurteilung stehen aber auch ernst zu nehmende Bedenken gegenüber (Neuweg 2005). Vor

diesem Hintergrund erscheinen regional und zeitlich zunächst beschränkte und sorgfältig evaluierte Pilotierungsphasen wünschenswert.

## 4.2 Stärkere Beachtung grundlegender prüfungsdidaktischer Prinzipien

Wichtige Prinzipien sollten konsequent(er) umgesetzt werden, insbesondere:

- der Einsatz bedeutsamer und anspruchsvoller Lern- und Prüfungsaufgaben (Thonhauser 2008) sowie die Vermittlung ihrer Bedeutsamkeit;
- die strikte Anwendung der curricularen Bezugsnorm, wofür einerseits das Projekt Bildungsstandards Hilfen anbieten kann, insbesondere durch eine Festlegung des jeweiligen Fundamentums (Mindeststandards), wofür andererseits aber auch die kollegiale Verständigung zwischen den Lehrer/inne/n darüber, worin „das Wesentliche“ im Fach zu sehen ist, stimuliert werden muss;
- die Steigerung der Frequenz und der Qualität der Rückmeldungen („Keine Aufgabe ohne Rückmeldung“);
- die Gewährung prüfungsfreier Lernphasen, die dem risikofreien Erproben dienen und in denen (auch missglückte) Leistungen als Erfahrungen interpretiert werden können, die das weitere Lernen positiv beeinflussen, wenn sie reflektiert werden;
- das Gebot der Prüfungsfairness, wonach für die Schüler/innen ein ausreichendes und zielorientiertes Unterrichtsangebot gegeben sein muss, um die Aufgaben zu bewältigen;
- ein möglichst intensives, auf Dauer gestelltes Gespräch zwischen Lehrperson und Schüler/inne/n über Prüfungsanforderungen, -modalitäten und -ergebnisse, insbesondere auch im Interesse der Förderung der Selbstbeurteilungskompetenz der Lernenden;
- die Gewährung auch von curricular unabhängigen Demonstrationschancen für die Lernenden, d. h. von Möglichkeiten, von sich aus Indikatoren für die Entwicklung ihrer Fähigkeiten und Interessen zu zeigen, auch wenn diese inhaltlich nicht den gerade aktuellen Lehrplanvorgaben entsprechen.

## 4.3 Erprobung innovativer Verfahren in der Praxis

Für die breitflächigere Erprobung alternativer Formen der Leistungsbeurteilung böte sich eine Fokussierung auf Schulstufen an, z. B. auf die 6. Schulstufe, auf der keine weitreichenden Schullaufbahnentscheidungen getroffen werden. Erprobenswerte Alternativen sind insbesondere:

- Direkte Leistungsvorlagen und Portfolios (insbesondere Entwicklungsportfolios)
- Pensenbücher und Lernziellisten zur Forcierung zielorientierten Lernens (insbesondere im Bereich der Fundamenta)
- verbale Beurteilung, insbesondere wenn die Anwendung einer individuellen Bezugsnorm geboten ist
- Mehr-Phasen-(Schul-)Arbeiten mit individuellen formativen Rückmeldungen
- Verwendung von (selbst angefertigten) Arbeitsunterlagen bei Prüfungen, um der Beschränkung auf Aufgaben, die auf die Wiedergabe von auswendig Gelerntem abzielen, entgegenzuwirken

Die Erprobung dieser Verfahren sollte in Verbindung mit Evaluationen erfolgen, in denen Forschung und Entwicklung systematisch verknüpft werden.

## 4.4 Revision der rechtlichen Grundlagen

(1) Die Orientierung an der curricularen Bezugsnorm setzt voraus, dass entsprechende Kriterien bzw. Standards vorliegen. Bezugspunkt der Leistungsbeurteilung sollten nicht die stark interpretationsbedürftigen Lehrpläne, sondern Kompetenzkataloge sein, die – möglichst operational – angeben, welche Kompetenzen für die Vergabe einer je bestimmten Note vorliegen müssen. Wenn solche für die Schulstufe und den Unterrichtsgegenstand maßgebliche Standards vorliegen, müssen einzelne Leistungsfeststellungen nicht mehr gesondert benotet und daher auch nicht arithmetisch in Jahresnoten überführt werden. Sie zeigen vielmehr, welche Kompetenzen der Schüler bereits besitzt und führen daher zu entsprechenden Einträgen in schülerindividuellen Kompetenzkatalogen (Listen mit „I can do“-Statements). Zu einem bestimmten Zeitpunkt nicht gezeigte Leistungen werden dem Schüler nicht „nachgetragen“, sofern er sie zu einem späteren Zeitpunkt zeigt.

**Kompetenzkataloge und kumulatives Lernen**

(2) Mit Blick auf die Mängel bei der Durchführung unterrichtsbegleitender Leistungsfeststellungen sollte verbindlich geregelt werden, in welchem Mindestmaß Leistungsfeststellung unterrichtsbegleitend zu erfolgen hat (z. B. mindestens ein Drittel unterrichtsbegleitend, maximal ein Drittel durch punktuelle Leistungsfeststellungen, ein Drittel durch standardisierte Tests) und was diese Form von den punktuellen Formen unterscheidet (Eingebundenheit in die Unterrichtsarbeit, Diagnose als Nebenprodukt von Aktivitäten, die primär auf Lehren und Lernen bezogen sind). Gleichzeitig ist die Regelungsasymmetrie – dichte Regelung der besonderen Formen bei gleichzeitig nahezu völligem Verzicht auf Regelungen bei der Mitarbeitsfeststellung – zu beseitigen, weil sie Umgehungshandlungen begünstigt.

**Mitarbeitsfeststellung**

(3) Für Berufungen ist die Einführung einer Schlichtungsstelle zu erwägen, die nicht nur bei der Beurteilung mit Nicht genügend angerufen werden kann. Für das Berufungsverfahren selbst könnte die derzeit nur für den Ausnahmefall nicht ausreichender Unterlagen vorgesehene kommissionelle Prüfung (§ 71 Abs. 4 und 5 SchUG) zum Standard erhoben werden. Jeder Berufungswerber bekäme auf diese Weise Gelegenheit, seinen Kompetenzstand zu beweisen. Der Berufung selbst sollte aufschiebende Wirkung zuerkannt bzw. ein vorläufiges Aufsteigen ermöglicht werden (vgl. bereits Fankhauser 1988: 68). Schüler/innen mit nur einem Nicht genügend sollten jedenfalls (einmal) zum Aufsteigen in die nächsthöhere Schulstufe berechtigt sein.

**Berufung und Aufsteigen**

(4) Die Bindung der Übertrittsberechtigung an die Noten in den Leistungsgruppen sollte angesichts deren geringer Validität und der sehr heterogenen Zuordnungspraxis zu den Leistungsgruppen durch eine an Standards orientierte Abgangs- oder Aufnahmeprüfung ersetzt werden. Eine solche Maßnahme würde gewährleisten, dass für alle Schüler/innen vergleichbare Kriterien angewendet werden.

**Übertrittsberechtigungen**

## 4.5 Intensivierung der Forschung

Die in diesem Beitrag, insbes. in 3.5, aufgezeigten Forschungsdesiderate sollten den Kern eines von der Bildungsadministration ausgerichteten und finanzierten Schwerpunktprogramms „Leistungsbeurteilung“ bilden, das Analyseprojekte sowie Entwicklungsprojekte vorsieht und für Universitäten und Pädagogische Hochschulen ausgeschrieben wird.

B