

Tanja Westfall-Greiter

Leistungsbeurteilung in der NMS: Geschärfte Konturen von Praxisproblemen

Summary: Leistungsbeurteilung ist ein wesentlicher Faktor, der die Lehr- und Lernkultur bestimmt. Sie ist somit eine wesentliche Lehrerkompetenz, die entsprechende Aufmerksamkeit im Rahmen von Professionalisierungsprozessen verdient. Obwohl seit längerem Expert/innen auf die Kluft zwischen Rechtslage und Praxis im österreichischen Schulwesen hinweisen, wird eine tradierte Praxis zum Gesetz. Mit der Einführung der Neuen Mittelschule (NMS) wurden die Konturen dieser systemimmanenten Praxisproblemen geschärft. Gleichzeitig wurden neue Dimensionen der Leistungsbeurteilung in der Gesetzgebung eingeführt, die international keine Entsprechung finden und somit zur Entwicklungsnot in der Praxis sowie zur Notwendigkeit wissenschaftlicher, fachdidaktischer Auseinandersetzung führen. Im Rahmen der NMS-Entwicklungsbegleitung wurden das Kompetenzfeld Leistungsbeurteilung analysiert und Impulse für eine gerechte und rechtskonforme Praxis gesetzt. Nichts desto trotz stellen manche Aspekte der NMS-Gesetzgebung eine Herausforderung dar.

Einleitung

Das Spannungsfeld zwischen Selektion und Förderung im Rahmen der Leistungsbeurteilung ist insbesondere in Bildungssystemen mit einem hohen Grad an systembedingter Selektion ausgeprägt. Obwohl Selektionsmechanismen in Form von Berechtigungen erst nach der Beurteilung der Lehrperson in Gang gesetzt werden, nehmen viele Lehrpersonen Selektion als ihre Aufgabe wahr. Diese Tiefenstruktur im System prägt gleichermaßen die Praxis, die Fort- und Ausbildung sowie die Bildungspolitik. Mit der Einführung der Neuen Mittelschule (NMS) in September 2012 wurden einerseits die Selektionsmechanismen in Form von Leistungsgruppen aufgehoben, andererseits wurde eine neue Dimension in der Benotung in den 7. und 8. Schulstufen hinzugefügt, indem die Leistungsentwicklung der Schüler/innen unter den Gesichtspunkten „vertieft“ und „grundlegend“ bei der Benotung beurteilt werden. Um entsprechend sinnvolle Vorschläge für den Umgang mit dieser Anomalie zu machen, ist zunächst eine Begriffsklärung notwendig. In Folge stellt sich die Qualitätsfrage für Leistungsbeurteilung im Rahmen der NMS-Gesetzgebung mit dem Ziel, eine gerechte, sachliche, akkurate Leistungsbeurteilung in der Praxis zu sichern. Möglichkeiten, den neuen Anforderungen professionell und sachlich gerecht zu werden, werden aufgezeigt, sowie Dilemmata, die die Praktikabilität einschränken.

Zwischen Selektion und Förderung – halb schwanger?

Die Einführung der NMS in der Sekundarstufe I wurde als epochales Moment in der Bildungsgeschichte Österreichs beschrieben. Diese Einschätzung ist berechtigt, insofern dass der Grad systembedingter Selektion durch die Abschaffung der Leistungsgruppen der Hauptschule reduziert sowie die Möglichkeit der Beteiligung für allgemein bildenden höheren Schulen (AHS) gesichert wurde. So gesehen ist die NMS genau genommen keine neue Schulform, sondern eine Form, Schule in der Sekundarstufe I zu gestalten. Sie bleibt weiterhin der „Angelpunkt der Diskussion um Struktur und Qualität des Schulsystems“ (Eder 2009) und Österreich bleibt weiterhin u.a. auf dem Radar der OECD (2013), weil die frühe systemische und systematische Selektion am Ende der Volksschule aufrecht bleibt.

Schon während der NMS-Pilotierungsphase wurden die Konturen von systemimmanenten Praxisproblemen in der Leistungsbeurteilung geschärft, die eine große Kluft zwischen Rechtslage und Praxis belegen (vgl. Eder, Thonhauser & Neuweg 2009). Die Selektionsfunktion der Leistungsbeurteilung dominiert. Sozialnormen steuern intransparente, variable Maßstäbe; Beurteilungen von Leistung sind mit Zuschreibungen und personenbezogenen (Vor)Urteilen verstrickt. Verhalten fließt in die Beurteilung der fachlichen Kompetenz durch Missverständnisse im Hinblick auf den Begriff „Mitarbeit“ ein, und die prüfende Funktion von Schule, wo Schüler/innen sich reüssieren müssen, steht vor ihrer fördernden oder gar bildenden Funktion.

Die Antwort auf diese Misstände seitens der NMS-Entwicklungsbegleitung (Schley, Schratz, Hofbauer & Westfall-Greiter 2010) war die Säule Leistungsbeurteilungskompetenz im „Haus der NMS“, der Professionalisierungsrahmen für die Qualifizierung von Lerndesigner/innen (Westfall-Greiter & Hofbauer 2012), welcher des Lehrgangcurriculum für die Lerndesigner/innen-Qualifizierung, die seit 2012 an Pädagogischen Hochschulen in Zusammenarbeit mit dem Bundeszentrum für lernende Schulen – NMS-Entwicklungsbegleitung (ZLS), weiterentwickelt wurde. Die obengenannten Misstände können allerdings von einer einzigen Lehrperson an einer Schule nicht bereinigt werden; es braucht die Kohärenz und Konsequenz auf allen Systemebenen sowie hochqualitative Fort-, Weiter- und Ausbildung, um die Leistungsbeurteilungspraxis im österreichischen System zu transformieren.

Begrifflichkeiten

Anlehnend an Eder, Neuweg und Thonhauser (2009) fungiert „Leistungsbeurteilung“ in diesem Beitrag als Überbegriff für die gesamte Praxis. Im Einklang mit der Leistungsbeurteilungsverordnung (LBVO) steht „Leistungsfeststellung“ für das Messen der Qualität einer Leistung; die Leistungsfeststellungsergebnisse dienen als Belege für die qualitative Benotung („Leistungsbeurteilung“ im engeren Sinn der LBVO). Leistungsfeststellung und Benotung brauchen jeweils unterschiedliche Prozesse und Werkzeuge: während Leistungsfeststellung Kriterien und Maßstäbe für den Vorgang des Messens voraussetzt, ist die Benotung ein qualitativer Vorgang, der die Lehrperson im Hinblick auf die Beurteilungsstufen der Notenskala der LBVO vollzieht. Dieser Vorgang setzt eine Entscheidungsgrundlage und klare Minimalanforderungen für jede Beurteilungsstufe voraus. Aus einer Leistungsfeststellung ergibt sich ein Ergebnis („score“), aus einer Benotung ergibt sich eine Note („grade“). Noten sind bei besonderen Prüfungsformen wie etwa einer Schularbeit sowie im Zeugnis notwendig. Eine beispielhafte Darstellung dieser Prozesse und entsprechender Werkzeuge wird von Rothböck in diesem Heft präsentiert (vgl. Westfall-Greiter 2012).

Spielregeln für eine qualitätsvolle Leistungsbeurteilungspraxis

Die Auswirkungen der LBVO auf die pädagogische Praxis werden von *Neuweg* (2009) ausführlich ausgelegt. Es seien hier nur wesentliche Aspekte für die folgende Diskussion erwähnt:

- Das Ziel von Leistungsbeurteilungsprozessen ist eine akkurate Beurteilung zu sichern; dafür braucht es relevante, korrekte Information als Grundlage für die Ermittlung der Note (vgl. *Stiggins, Arter, Chappuis & Chappuis* 2006).
- Es gilt das Prinzip, „so viel wie nötig, so wenig wie möglich“ bei der Leistungsfeststellung, um die Note auf Basis ausreichender, valider Belege zu ermitteln.
- Die Mitarbeit umfasst unterrichtsbegleitende Leistungsfeststellungen, die sich von punktuellen Leistungsfeststellungen unterscheiden, indem Schüler/innen nicht merken, dass die Qualität ihrer Leistung gemessen wird (vgl. *Eder, Neuweg & Thonhauser* 2009).
- Die Leistungsfeststellung ist ein Vorgang, der an entsprechenden Kriterien und transparenten Maßstäben orientiert ist (vgl. *Marzano* 2009).
- Kontinuierliche Informationsfeststellungen dienen der Lehrperson als Grundlage für Entscheidungen über die weitere Vorgehensweise im Unterricht (vgl. *Earl* 2013; *Hattie* 2012; *Beywl, Bestvater & Friedrich* 2011).
- Die Bildung von Mittelwerten zur Ermittlung der Note bzw. Prozente- bzw. Punktesysteme sind nicht nur rechtswidrig sondern „worst practice“ für eine gerechte, akkurate Beurteilung, weil sie Ergebnisse verzerren (vgl. *Marzano* 2009).

Auswirkung der NMS-Gesetzgebung auf Leistungsbeurteilung

Durch die Aufhebung der Leistungsgruppen gelten für alle Schüler/innen in den 5. und 6. Schulstufen gleiche Fachlehrpläne als inhaltliche Grundlage sowie die fünfstufige Notenskala für Leistungsbeurteilung. Lehrkräfte, die an einer Stratifizierung der Lehr- und Lerninhalte (unterschiedlicher Input für unterschiedliche Leistungsgruppen) in den Gegenständen Deutsch, Mathematik und Englisch jahrelang gewöhnt waren, stehen vor einem scheinbar unlösbaren Problem. Im ersten Moment scheint es, als ob irgendetwas nachgeben muss: entweder eine Nivillierung des Niveaus nach unten oder eine Ausdehnung der Notenskala. Beide Optionen sind kaum zu rechtfertigen, wenn die Ziele der NMS aufrecht bleiben sollen.

Vielmehr braucht es zunächst die inhaltliche Klärung fachspezifischer Anforderungen entsprechend jeder Schulstufe, die für alle gelten. Die Bildungsstandards für die 8. Schulstufe dienen dabei als Orientierung für die Kompetenzentwicklung; diese Entwicklung muss allerdings zuerst rückwärts von der 8. auf die 5. Schulstufe heruntergebrochen werden, um darauf hin die Anforderungen und Inhalte Schulstufe für Schulstufe aufzubauen. Dann müssen Aufgaben für die Leistungsfeststellung mit den Anforderungen übereinstimmen (*alignment of assessments and expectations*), damit das, was erzielt wird, tatsächlich beurteilt wird. Erst dann können die qualitativen Beurteilungsstufen der Notenskala in Beziehung zu den Anforderungen einer Schulstufe gesetzt werden. Dieser Prozess stellt einen hohen fachdidaktischen Anspruch an Lehrpersonen dar, ist aber überhaupt nicht neu. Er ist Kern der lehramtlichen Tätigkeit.

Die wesentlichste und auffälligste Neuigkeit zur Leistungsbeurteilung in der NMS-Gesetzgebung ist die Leistungsbeurteilung in den differenzierten Pflichtgegenständen nach „grundlegenden“ und „vertieften“ Gesichtspunkten in den 7. und 8. Schulstufen. Das Gesetz ist in diesem Punkt allerdings ambivalent, weil „grundlegend“ und „vertieft“ als Begriffe nicht definiert sind. An unterschiedlichen Stellen des Schulorganisationsgesetzes werden zwei Hinweise gegeben, die zur Konkretisierung für die Praxis dienen sollen (kursiv von mir):

„...eine Unterscheidung nach grundlegender und vertiefter Allgemeinbildung erfolgt, wobei die Inhalte der vertieften Allgemeinbildung *eine Auseinandersetzung mit den grundlegenden Bildungsinhalten in einer über die Grundanforderungen hinausgehenden Art auf einem höheren Komplexitätsgrad vorzusehen haben.* (§ 8 n SchOG, BGBl Nr. 242/1962 idF BGBl I Nr. 36/2012)

„Die Anforderungen der Vertiefung haben jenen der Unterstufe der allgemein bildenden höheren Schule zu entsprechen.“ (§ 21b. (2) SchOG)

Weiters präzisiert die NMS-Lehrplanverordnung (LPVO), dass „die Zuordnung der Leistungen in grundlegende und vertiefte Allgemeinbildung *im Rahmen der Benotung* auszuweisen“ sind (BGBl. II, Nr. 185, S. 10). Hinweise sowohl auf *Inhalt* als auch *Anforderungen* (*Auseinandersetzung* über *etwas* hinaus) sind in den Gesetzestexten enthalten. Im ersten Moment mögen diese Ausführungen klar erscheinen. Aber was ist „vertiefte Allgemeinbildung“? Was sind „grundlegende Bildungsinhalten“ bzw. „Grundanforderungen“? Was ist „Komplexitätsgrad“? Welche „Anforderungen der Vertiefung“ gibt es in den AHSen?

Diese Fragen lösen eine Suchbewegung aus, die Detektivarbeit im Labyrinth der Gesetzgebung ähnelt. Zunächst führt der Hinweis auf AHS zu den Lehrplanverordnungen. Obwohl idente Fachlehrpläne für die gesamte Sekundarstufe I gelten, gibt es Unterschiede im gesetzlichen Auftrag. Die NMS-LPVO schafft eine inklusivere Schule mit dem gesetzlichen Auftrag, die Schüler/innen „je nach Interesse, Neigung, Begabung und Fähigkeit für den Übertritt in mittlere oder in höhere Schulen zu befähigen und auf das Berufsleben vorzubereiten“ (BGBl. II, Nr. 185, S. 1). Die AHS hat hingegen den gesetzlichen Auftrag, den Schüler/innen „eine umfassende und vertiefte Allgemeinbildung zu vermitteln und sie zugleich zur Hochschulreife zu führen“ (LPVO-AHS, S. 1). Zugrunde liegt diesem Auftrag die Selektion auf Basis von Noten, die sichern soll, dass nur die schulisch erfolgreichsten Schüler/innen Zugang zur AHS und somit zur Vorbereitung auf Hochschulreife haben.

Fragen zur Aussagekraft der Noten in der 4. Schulstufe sowie zur Sinnhaftigkeit der Feststellung von Eignung für Universitätsreife bei 10-Jährigen beiseite gelassen, reichen diese Unterschiede nicht für die Konkretisierung von „vertieft“ und „grundlegend“ aus – ob im Bezug zum Inhalt oder zu den Anforderungen. Weil die Fachlehrpläne ident sind und die Bildungsstandards für die 8. Schulstufe für die gesamte Sekundarstufe I gelten, stellt sich die Frage, was der Unterschied sein könnte. Auch hier gerät man in eine Sackgasse, wenn man nach einem „Niveau“ sucht: Leistungsmessungen und soziometrische Analysen können seit den 1980er keinen klaren Unterschied im „Niveau“ zwischen AHS und Hauptschule bestätigen (vgl. Eder 2009). Vielmehr zeigt sich einen Zusammenhang zwischen Unterricht und Leistungsentwicklung der Schüler/innen, wobei die genaue Praxis in diesen Schulformen weitgehend unbeforscht bleibt (vgl. Eder 2006).

Auch bei der Suche nach „grundlegenden Bildungsinhalten“ in den Fachlehrplänen scheitert man. Nun wendet sich die Suche in eine andere Richtung. Was ist „Komplexitätsgrad“? Die Beurteilungsstufen der LBVO geben indirekt Orientierung für wesentliche Faktoren zur Beurteilung von Komplexität erbrachter Leistungen: 1) Eigenständigkeit und die Fähigkeit, Wissen und Können bei neuartigen Aufgaben anzuwenden (Transfer), 2) Erfassung des Lehrstoffes (von Reproduktion zum Verstehen, Wesentlichkeit). Allerdings fehlt ein Instrument für die sachliche Bestimmung von Komplexitätsgrad in der Praxis.

Bestimmung von „vertieft“ und „grundlegend“ nach Grad der Komplexität

„Komplexitätsgrad“ fungiert als Fachbegriff in mehreren Bereichen und ermöglicht eine sachliche Orientierung für die Bestimmung von Komplexität. Wir blenden den transdisziplinären Bereich der Komplexitätstheorie aus und gehen zu schulrelevanten

Disziplinen: einerseits Kognitionspsychologie, andererseits Testing and Assessment. Insbesondere im Bereich der standardisierten Testung, wo die Bestimmung vom Komplexitätsgrad der Aufgaben essentiell ist, gibt es im Zeitalter der Kompetenzorientierung rege Entwicklung. Wir wenden uns aber zunächst der Kognitionspsychologie zu, aus der die Bloom'sche Taxonomie (*Taxonomy of Educational Objectives*, 1956) entstanden ist. Auch wenn die überarbeitete Taxonomie (*Anderson und Krathwohl* 2001) mehrere Problemstellen des veralteten Modells beseitigt, gibt es zwei Kritikpunkte: Erstens begreift die Kognitionstheorie, worauf die Taxonomie baut, Kognition als mentale Prozesse im Kopf und blendet dabei den regen Diskurs zu „erweiterter Kognition“ (vgl. *Rowlands* 2010) aus. Aktuelle Theoriebildungsversuche in der philosophischen Psychologie legen aus, warum Kognition nicht bloß als Gedankengänge im Kopf zu begreifen ist, sondern als gesamtheitlicher Vorgang, der mit Körper, Umfeld und Situation verstrickt ist. Sowohl pädagogisch-wissenschaftlich als auch bildungstheoretisch scheint es sinnvoll, bei der Bestimmung von Allgemeinbildung diese Frage zu überlegen. Ist der Mensch, der gebildet wird, im cartesianisch-dualistischem Sinn bloß ein Kopf, oder liegt ein ganzheitliches Menschenbild pädagogischer Überlegungen zugrunde?

Der zweite Kritikpunkt liegt im Wesen einer Taxonomie als Klassifikationsschema. In den Naturwissenschaften dienen Taxonomien der Klassifizierung und dem Umgang mit Ausnahmen, aber Pädagogik ist keine naturwissenschaftliche Praxis. Anthropologisch betrachtet entstehen im Kulturleben Taxonomien im Dienst der sozialen Ordnung bzw. der Erklärung von Phänomenen. Es besteht in der Schulpraxis das Risiko, dass die Bloom'sche Taxonomie zu einem quasi-wissenschaftlichen Instrument der Klassifizierung von Schüler/innen wird, wenn es der Lehrperson nicht gelingt, sich von der Selektionsdenkweise zu befreien. In der Fachdidaktik, in der Lernziele in einer linearen Progression nach der Taxonomie formuliert werden, stellt sich die Frage, ob diese Progression aktuellen Erkenntnisse über die Entwicklung von Expertenwissen (Breite und Tiefe) gerecht wird. Auch im Hinblick auf lebensweltlichen Handlungssituationen ist eine saubere lineare Stufung nicht gegeben: Expert/innen müssen manchmal Informationen bloß wiedergeben und „Nicht-Expert/innen“ sind mit kreativen, komplexen Aufgaben konfrontiert - oft mit erstaunlichem Erfolg.

Nun geht die Suche nach einem sinnvollen Instrument für die Bestimmung von Komplexitätsgrad in Richtung *Testing and Assessment*. Weil die Bestimmung von „vertieft“ und „grundlegend“ sowohl mit Aufgabenstellung als auch mit Leistungsergebnis zu tun hat, scheint *Webbs* Modell „*Depths of Knowledge*“ („Tiefe des Wissens“) vielversprechend zu sein. Entwickelt auf Basis seiner Expertise für das Council of Chief State School Officers (*Webb* 1997) und mit Förderung der National Science Foundation (*Webb et al.*, 2005), wird *Webbs* Modell zunehmend im Prüfungs- und Schulwesen bevorzugt und verwendet. *Webb* analysierte die Komplexität von Aufgaben in Standardstestungen und entwickelte daraus vier Bereiche für die Bestimmung von Komplexitätsgrad unter Berücksichtigung der Handlungssituation, die eine Aufgabe erzeugt.

Webbs „Depths of Knowledge“ Bereiche	
Bereich 1: Abrufen/Wiedergeben	Fakten, Informationen, Begriffe, einfache Verfahren wiedergeben; vertraute Prozesse oder Formeln verwenden
Bereich 2: Fertigkeiten/ Schlüsselkonzepte	Informationen bzw. Schlüsselkonzepte anwenden; zwei oder mehrere Schritte; Überlegungen über Lösungswege anstellen
Bereich 3:	Logisch denken, einen Plan entwickeln, Belege/Daten verwenden,

Strategisches Denken	mehrere Lösungswege zur Verfügung, begründen, Schritte in Reihenfolge setzen, Abstraktion
Bereich 4: Erweitertes Denken	Untersuchen, erkunden, nachdenken, mehrere Bedingungen bei der Problemanalyse und Lösungsfindung berücksichtigen, vernetzen, in Beziehung setzen, eine Lösungsstrategie aus vielen möglichen entwickeln und anwenden

Das DOK-Modell hat zwei Vorteile: Einerseits ist es auf Basis Aufgabenstellungen und Standardsformulierung im Zeitalter der Kompetenzorientierung entstanden, andererseits ermöglicht es eine systematische Bestimmung vom Komplexitätsgrad. Der Fokus bleibt auf der Aufgabenstellung und der daraus resultierenden Leistung. Aus diesen Gründen scheint es für die Bestimmung von „vertieft“ und „grundlegend“ entlang dem Grad der Komplexität gut geeignet zu sein und bietet darüber hinaus Lehrkräften ein praktikables Werkzeug, um die Anforderungen ihrer Aufgabenstellung mit Standards im Einklang zu bringen.

Schwierigkeit vs. Komplexität

Webbs Modell orientiert sich an Komplexität, nicht Schwierigkeit. „Schwierigkeit“ als Fachbegriff bezieht sich auf die Häufigkeit von korrekten Antworten zu einer Frage, z.B.: Wenn viele Prüflinge die Frage, „Was bedeutet ‚unklar‘?“ beantworten können, ist sie leicht. Wenn wenig eine Antwort für „Was bedeutet ‚Ambiguitätstoleranz‘?“ parat haben, ist die Frage schwierig. In beiden Fällen ist allerdings die kognitive Leistung die Gleiche – Begriffe wiedergeben, d.h. Bereich 1. Zunehmender Schwierigkeitsgrad in einfacheren Aufgaben bildet Differenz, die insbesondere in Tempo sichtbar wird. Dies passiert weniger bei Aufgaben mit komplexeren Ansprüchen, weil *zunächst alle denken müssen*. Die Devise lautet: Einfache Aufgaben führen zu einfachen Leistungen, komplexe Aufgaben zu komplexen Leistungen. Unterschiede in Schwierigkeitsgrad dienen nicht der Bestimmung von vertieft/grundlegend nach Komplexitätsgrad und vor allem nicht der Bildungsqualität. Die Herausforderung in der Praxis ist entsprechend den Anforderungen der Schulstufe anspruchsvolle, d.h. komplexe, Aufgaben zu stellen (etwa „Argumentieren“ in Deutsch oder Mathematik). Dies ist besonders schwer, wenn wenige Beispiele vorhanden sind.

Sind komplexe Aufgaben für alle?

Die kurze Antwort auf die Frage, ob komplexe Aufgaben für alle sind, lautet: Ja! Es ist nicht nur fragwürdig im Hinblick auf Gerechtigkeit, sondern auch nicht rechtlich zulässig, das Leistungspotential von Schüler/innen durch die Zuteilung von unterschiedlichen Aufgaben vorwegzunehmen bzw. eine Person mit „grundlegend“ oder „vertieft“ zu etikettieren. Der Krux des Paradigmenwechsels in dieser Lehr- und Lernkultur liegt darin, dass sich *alle* mit komplexen Aufgaben auseinandersetzen, damit ihnen selbst und der Lehrperson ihr volles Leistungspotential sichtbar gemacht wird. Aus diesem Grund wird in den „Richtlinien für die NMS-Entwicklungsarbeit“ des BMUKKs präzisiert, dass jegliche Formen der Leistungsfeststellungen „sowohl die grundlegende als auch die vertiefte Allgemeinbildung abbilden“ müssen, damit „**allen Schüler/innen alle Komplexitätsstufen über das ganze Schuljahr hinweg**“ eröffnet werden (2012, S. 6). „Daraus resultiert, dass **unterschiedliche Aufgabenstellungen** nach grundlegender und vertiefter Allgemeinbildung bei Leistungsfeststellungen **nicht vorzusehen** sind. Zuweisungen von Aufgabenstellungen dürfen nicht auf Grund von Prognosen über das Leistungsvermögen eines Schülers /einer Schülerin getroffen werden“ und die Zuweisung von Schüler/innen in Gruppen nach grundlegender/vertiefter Allgemeinbildung ist unzulässig (Ibid.).

Wenn nicht alle im Rahmen von Lern- und Lehrprozessen mit komplexen fachspezifischen Aufgaben konfrontiert werden, a) haben sie keine Chance, ihre Fähigkeiten bei komplexen

bzw. herausfordernden Aufgabenstellen weiter zu entwickeln bzw. zu beweisen, und b) gibt es in den Aufzeichnungen keine Grundlage für eine Beurteilung nach allen Beurteilungsstufen der LBVO – mit oder ohne „vertieft“ und „grundlegend“. Bereits die Beurteilungsstufen selbst stellen einen hohen Anspruch hinsichtlich Vertiefung in einem Fach, und zwar bei den Faktoren *eigenständige Anwendung* und *Wesentlichkeit*, dar. Reproduktive Aufgaben, die einen einfachen kognitiven Anspruch stellen (Webb-Bereich 1), schränken die Leistungsentwicklung sowie die Leistungsfeststellungen als Basis für die Ermittlung der Note auf ein „Genügend“ ein.

Kriteriengestützte Leistungsbeurteilung

Die Leistungsbeurteilung in der NMS erfolgt weiterhin im Rahmen der Rechtslage. Leistungsfeststellungen und –beurteilungen (Benotungen) sind sachnormorientiert, d.h. auf Basis Kriterien, vorzunehmen. Transparenz ist gefragt: „Die Anforderungen sind den Schülerinnen und Schülern einsichtig zu machen, vor allem über transparente Beurteilungskriterien mit Bezug zu den jeweiligen Kompetenzen“ (NMS-LPVO, S. 10). Dies findet Zustimmung von Expert/innen: „Es wäre ausgesprochen wünschenswert, wenn der Lehrer schon am Beginn des Schuljahrs völlige Transparenz in der Notengebung schafft“ (Neuweg 2009, S. 102).

Die Lehrperson muss im Rahmen der Leistungsfeststellung Belege in ihren Aufzeichnungen sichern, die der Ermittlung einer akkuraten Note dienen. Das ist nicht neu, weder in den gesetzlichen Rahmenbedingungen noch in „good practice“. Akkuratheit setzt Kriterien voraus. Als Werkzeuge fungieren Beurteilungsraster (*rubric*) jeglicher Form. Diese beschreiben konkrete Performanzen entlang von Kriterien auf unterschiedlichen Qualitätsniveaus, damit Leistungsqualität objektiv beurteilt werden können. Sie sind insbesondere bei komplexen Aufgaben unverzichtbar, weil diese Leistungen nicht mit „richtig-falsch“ beurteilt werden können. Der Beurteilungsraster dient der Transparenz und löst Schüler/innen von ihrer Herkunft, indem Kriterien und klar definierte Qualitätsstufen fachliche Leistungen an- und besprechbar machen. Sie sind wirksame Katalysatoren für die Entwicklung einer inhaltlich-fachlichen, d.h. „akademischen“ Sprache, die im *Nationalen Bildungsbericht 2012 Band 2* von Herzog-Punzenberger & Schnell (2012) als wesentlicher differenzvermeidender Faktor hervorgehoben wird.

Aus diesen Gründen werden in der Schulwirksamkeitsforschung sowie in der forschungsbasierten Systementwicklung Beurteilungsraster jeglicher Form (analytisch/holistisch, „4.0-Skala“ in der NMS, „Kriterienkatalog“ im IKM-Deutsch, usw.) als zentrale Grundlage für lernwirksame Leistungsbeurteilungspraxis betont. Weil Verwirrung durch die inflationäre Verwendung von Rastern in der Schulpraxis herrscht, ist der Unterschied zwischen Beurteilungsraster und Kompetenzraster (*skills inventories*) an dieser Stelle zu erwähnen. Kompetenzraster sind Kataloge von Fertigkeiten, wie etwa die Bildungsstandards, und stellen *eine curricular-inhaltliche Entwicklung* dar. Sie sind kein Instrument für kriteriengestützte Leistungsbeurteilung, sondern bestimmen was beurteilt wird. Als Selbeinschätzungsinstrument werden sie erst wirksam, wenn die Kriterien entsprechend der Schulstufe verstanden werden. Sonst bleiben sowohl Fremd- als auch Selbstbeurteilung in der Subjektivität und Beliebigkeit verhaftet.

Die entstehende Leistungsbeurteilungspraxis in der NMS

Das Ziel in der Leistungsbeurteilung ist eine Orientierung an Kompetenzen, Komplexitätsgrad und Kriterien, die sogenannte „3-K Orientierung“. Dieses Ziel ist kein NMS-Spezifikum. Die rechtliche Grundlage diesbezüglich gilt für das gesamte Schulwesen. Konturen des Problems, vor allem die Kluft zwischen Rechtslage und Praxis sowie zwischen good practice und realer

Praxis, wurden durch die NMS-Entwicklungsarbeit geschärft. Für NMS-Lehrer/innen und damit für die Fort- und Ausbildung ergeben sich folgende Ansprüche für die curriculare-inhaltliche Entwicklung, Gestaltung von Lehr- und Lernprozessen sowie Leistungsbeurteilung:

- *Kompetenzen* und *Kriterien* werden im Einklang mit dem Fachlehrplan und den Bildungsstandards im Vorfeld des Unterrichts nach dem Prinzip „vom Ende her“ festgelegt.
- Die *Lernzielformulierungen* stellen dar, was am Ende beurteilt wird.
- *Kriteriengestützte Beurteilung* bedeutet, dass relevante *Kriterien* im Einklang mit den Kriterien für die Bildungsstandards festgelegt und in Qualitätsstufen in einem *Beurteilungsraster* konkretisiert werden.
- Ein *Spektrum vom Komplexitätsgrad* entsprechend den Lernzielen der Schulstufe wird in allen Lehr- und Lernaktivitäten *für alle* gesichert.
- *Transparenz* wird durch Verständigungsprozesse über Lernziele in Verbindung mit Kriterien und Qualitätserwartungen entsprechend der Schulstufe gesichert, damit alle Beteiligten Lern- und Lehrprozesse zielgerecht steuern können.

Das „No-Go“ Mittelwertbildung bei der Ermittlung der Note

Die größte Barriere zu einer besseren Leistungsbeurteilungspraxis ist die Tradition der Mittelwertbildung sowie die Verwendung von Punkte- bzw. Prozentsysteme. Diese verbreiteten Praxen sind „weder messtechnisch noch rechtlich gerechtfertigt“ (*Neuweg* 2009, S. 103). Außerdem ist zu beachten, dass Noten keine Zahlen sind, „sondern in ihrer Abbildungsqualität eher Schätzungen der Leistung darstellen. Es ist daher – von der rechtlichen Fragwürdigkeit abgesehen – eine äußerst zweifelhafte Praxis, Notendurchschnitte auf zwei Dezimalen genau zu ermitteln und etwa die Entscheidung für eine Zeugnisnote ‚Befriedigend‘ oder ‚Genügend‘ davon abhängig zu machen, ob der Durchschnitt nun 3,42 oder 3,61 lautet.“ (*Ibid.*, S. 104).

Ausblick: Dilemmata auf dem Weg zur Chancengerechtigkeit

Ungeklärt bleibt weiterhin Fragen zur Beziehung zwischen den Noten. Wie ist „befriedigend“ in vertiefter Allgemeinbildung mit „sehr gut“ in „grundlegender Allgemeinbildung“ inhaltlich gleich zu stellen? Kann eine einzige Aufgabenstellung alle Komplexitätsgrade beinhalten und sichtbar machen? Es scheint Staub von Gestern in der Gesetzgebung zu geben. Man kann darüber spekulieren, ob solche Dilemmata durch das Beharren auf zwei Schulformen im Interesse gesellschaftlich wirksamer Differenzbildung entstehen, aber dies hilft den Lehrer/innen nicht, die im rechtlichen Rahmen in der Praxis jetzt handeln müssen. Die beste Strategie für die Praxis scheint zu sein, stets Aufgaben und Ansprüche auf dem gesamten Spektrum von Komplexitätsgrad entsprechend den Anforderungen der Schulstufe für alle zu sichern, um überhaupt Lernanlässe sowie Belege für eine gerechte Benotung zu sichern.

Wenn Chancengerechtigkeit gewährleistet werden soll, heißt das für die Praxis: alle Schüler/innen werden in Richtung vertieftes Wissen bereits ab der 5. Schulstufe gefördert und gefordert. Dies scheint sowieso die Intention zu sein, wenn einerseits Inhalte entsprechend der Bildungsstandards und der Fachlehrpläne ausgelegt werden und andererseits die Anforderungen nach den Beurteilungsstufen der Notenskala in der LBVO verstanden werden. Letztendlich ist dies auch die Botschaft der Gesetzgebung: „Eine unterschiedliche Gestaltung der Lehrpläne der Unterstufe der allgemein bildenden höheren Schule und der Hauptschule oder der Neuen Mittelschule darf den Übertritt von Schülern der Hauptschule oder der Neuen Mittelschule in die allgemein bildende höhere Schule (§ 40 Abs. 2 bis 3a) nicht erschweren“ (SchOG § 39 Abs. 2).

Weitere Reform?

Die Reform der Leistungsbeurteilungsverordnung wird seit längerem eingefordert. Neue Fachlehrpläne und die Bildungsstandards sichern im curricularen Bereich der Sekundarstufe I die Kompetenzorientierung, der die LBVO nicht mehr gerecht wird. Die Verordnung braucht Aktualisierung im Hinblick auf den Erfordernissen einer kompetenzorientierten Beurteilung. Dazu gehören die kriterienbasierte Leistungsfeststellung von Kompetenzen; Prüfungsformen und -aufgaben, die die erzielten Kompetenzen sichtbar und messbar machen und folglich grundsätzliche Überlegungen dahingehend, ob die geregelten „besondere Prüfungsformen“ wie Schularbeiten überhaupt noch zweckmäßig sind.

Es scheint weniger sinnvoll, die Notenskala in der LBVO an der 7-stufigen Notenskala der NMS-Gesetzgebung anzupassen. Die fünf gültigen Beurteilungsstufen korrelieren erstaunlicherweise mit den Beschreibungen von Transfergraden von Kompetenz. Problematisch bleibt nach wie vor die Ermittlung der Note, sowohl bei punktuellen Leistungsüberprüfungen als auch als zusammenfassendes Gutachten über Leistungen, die während eines längeren Beobachtungszeitraumes erbracht wurden. In dieser Black Box geschieht die subjektive Selektion. Diesem Praxisproblem auszuweichen, indem Noten einfach abgeschafft werden, scheint auch nicht sinnvoll. Um von der rechtswidrigen Praxis der Mittelwertbildung weg zu kommen, braucht es für die Ermittlung der Note die Präzisierung geeigneter Instrumente (etwa wie die „Entscheidungsgrundlage“ von Stiggins et al., 2006, die derzeit in der NMS in Erprobung ist, s. Beitrag von Rothböck in diesem Heft) sowie die Konkretisierung von Minimalanforderungen für jede Beurteilungsstufe. Empfehlenswert wäre auch als Unterstützung die zur Verfügungstellung von Beurteilungsraster für wesentliche Kompetenzen in jedem Fach für jede Schulstufe, damit Lehrpersonen das Rad nicht neu erfinden müssen. Allerdings muss behutsam und wachsam mit der Lehrfreiheit in dieser Frage umgegangen werden.

Weiters müssen die Regelungen der LBVO in allen Schulformen und allen Fächern praktikabel sein. Einsemestrige Kurse mit wenig Kontaktstunden, die in Sekundarstufe II häufiger vorkommen, stellen besondere Herausforderungen dar, aber auch hier muss eine transparente und gerechte Beurteilung im Einklang mit der Gesetzgebung machbar sein, wenn das Papier wirksam und nicht nur geduldig sein soll. Letztlich soll die Sprache einer zeitgemäßen Verordnung im Einklang mit der Fachsprache der Beurteilung sein. Inzwischen gilt es, den rechtlichen Rahmen sinnvoll und gerecht zu nützen, damit die Leistungsbeurteilungspraxis nicht im Weg des Lehrens und Lernens steht. Es gibt noch viel zu tun, vor allem im Rahmen der Professionalisierung von Lehrkräften. Ein erster Schritt in Richtung Qualität in der Leistungsbeurteilung ist daher einerseits Anerkennung für die Wichtigkeit der Lehrperson in ihrer lehramtlichen Tätigkeit und andererseits Verantwortungsübernahme seitens Lehrpersonen für eine hochqualitative, gerechte Beurteilungspraxis, die Bildungsprozesse nicht hemmt, sondern stärkt. Dafür brauchen sie Unterstützung von allen Sysebenen.

LITERATUR

- Anderson, L. W. & Krathwohl, D.R. (2001): A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. New York: Addison-Wesley.
- Beywl, W., Bestvater, H., Friedrich, V. (2011). Selbstevaluation in der Lehre: Ein Wegweiser für sichtbares Lernen und besseres Lehren. Münster: Waxmann.
- Bloom, B.S. (Ed.) (1956): Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay.

- Bloom, B.S. (Hrsg.) (1972): Taxonomie von Lernzielen im kognitiven Bereich. 4. Auflage. Beltz Verlag, Weinheim und Basel.
- BMUKK. (2012): Richtlinien für die NMS-Entwicklungsarbeit. Stand September 2012. Abrufbar unter: <http://www.nmsvernetzung.at/mod/forum/discuss.php?d=2880>.
- Earl, L. (2013). *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*, 2. Auflage. Thousand Oaks, CA: Corwin Press.
- Eder, F., Neuweg, G.H. & Thonhauser, J. (2009): B6: Leistungsfeststellung und Leistungsbeurteilung. In: Specht, W. (Hrsg.) *Nationaler Bildungsbericht Österreich 2009, Band 2. Analysen*. Graz: Leykam, S. 247-268. Abrufbar unter: www.bifie.at/buch/1024.
- Hattie, J. (2012). *Visible Learning for Teachers: Maximizing impact on learning*. London: Routledge.
- Herzog-Punzenberger, B. & Schnell, P. (2012): In: Herzog-Punzenberger, B. (Hrsg.) *Nationaler Bildungsbericht Österreich 2012, Band 2*. Graz: Leykam, S. 229-268. Abrufbar unter: www.bifie.at/system/files/buch/pdf/NBB2012_Band2_gesamt_0.pdf.
- Marzano, R. (2009): When Students Track Their Progress. *Educational Leadership*, 67/4, S. 86-87. Online abrufbar: www.ascd.org/publications/educational-leadership/dec09/vol67/num04/When-Students-Track-Their-Progress.aspx
- Neuweg, G.H. (2009): *Schulische Leistungsbeurteilung: Rechtliche Grundlagen und pädagogische Hilfestellungen für die Schulpraxis*. 4. Auflage. Linz: Trauner Verlag.
- Rowlands, M. (2010). *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. Boston: Massachusetts Institute of Technology.
- Schley, W., Schratz, M., Hofbauer, C. & Westfall-Greiter, T. (2009): Das Konzept der NMS-Entwicklungsbegleitung als Transformationsprozess. *Erziehung und Unterricht*, 7-8, 686-696.
- Stiggins, R., Arter, J., Chappuis, J. & Chappuis, S. (2006) : *Classroom Assessment for Student Learning: Getting it right – doing it well*. Portland: Assessment Training Institute.
- Webb, N. (1997): Research Monograph No. 8, "Criteria for Alignment of Expectations in Assessments in Mathematics and Science Education." Council of Chief State School Officers. Abrufbar unter: <http://facstaff.wceruw.org/normw/WEBBMonograph6criteria.pdf>.
- Webb, N. L. und andere. (2005). "Webb Alignment Tool". Wisconsin Center of Educational Research, University of Wisconsin-Madison, funded by the National Science Foundation. Abrufbar unter: wat.wceruw.org/index.aspx.
- Westfall-Greiter, T. (2012): Orientierungshilfe Leistungsbeurteilung: Teil 1: Grundlagen und Begriffe. URL: www.nmsvernetzung.at/mod/glossary/view.php?id=2473&mode=entry&hook=1731.
- Westfall-Greiter, T. & Hofbauer, C. (2012). NMS-Entwicklungsbegleitung Neu. Einblick in das Bundeszentrum für lernende Schulen. *Erziehung und Unterricht*, 9-10, 836-842.
- Wiggins, G. (1998): *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: John Wiley & Sons.

ZUR AUTORIN/ZUM AUTOR

Tanja WESTFALL-GREITER, M.A. Bachelorsstudium der Anglistik/Amerikanistik an der University of Wisconsin-Stevens Point, lehramtliches Mastersstudium der Erziehungswissenschaft an CUNY City College of New York. Referententätigkeit in der LehrerInnenaus- und weiterbildung, Autorin von Lehrwerken sowie Fachbüchern und wissenschaftlichen Beiträgen zur Schulentwicklung und Lernen, Forschungstätigkeit am Zentrum für Lernforschung der Universität Innsbruck, seit 2008 Entwicklungsbegleiterin der Neuen Mittelschule. Seit 2012 Co-Leitung, Bundeszentrum für lernende Schulen – NMS-Entwicklungsbegleitung (ZLS).