

# Chancen und Grenzen der Leistungsmessung an Gymnasien<sup>1</sup>

Urs Moser

*Institut für Bildungsevaluation,  
assoziiertes Institut der Universität Zürich*

Sehr geehrte Herren Regierungsräte  
Sehr geehrte Rektorinnen und Rektoren  
Sehr geehrte Damen und Herren

«Das Gymnasium lebt» lautete der Titel eines Kommentars der Neuen Zürcher Zeitung zum Jubiläum «175 Jahre Zürcher Mittelschulen». Darin wird festgehalten, dass es den Gymnasien immer noch gelinge «aus pubertierenden Kindern junge Leute zu machen, die in der Auseinandersetzung mit anspruchsvollen Stoffen, im Ausprobieren ihrer körperlichen und musischen Talente, in der Arbeit an ihrer Urteilsfähigkeit die Reife erlangt haben, ihren Weg auf eigenen Füßen weiterzuverfolgen». Wahrlich ein Kompliment, dessen Gültigkeit anhand testtheoretisch basierter Leistungsmessung allerdings kaum überprüft werden kann.

Zum Leben des Gymnasiums gehört aber auch – so die Auffassung der NZZ – die gemeinsame Arbeit der Mittel- und Hochschulen an der Definition der für ein Studium erforderlichen Qualifikationen: «Soll die Matur das Ticket zum Eintritt zur Hochschule bleiben, sind mess- und vergleichbare Anforderungen unentbehrlich».

Ich bin aus zwei Gründen glücklich über diesen Kommentar in der NZZ zum künftigen Leben des Gymnasiums. Erstens wird mit zwei Sätzen auf das Dilemma der standardisierten Leistungsmessung an Gymnasien hingewiesen. Ein grosser Teil dessen, was an Mittelschulen vermittelt wird, entzieht sich der standardisierten Leistungsmessung; vor allem dann, wenn diese auf Test- oder Prüfungsaufgaben reduziert wird. Zugleich können sich die Mittelschulen in der Schweiz künftig nicht mehr einer Standardisierung der Anforderungen verschliessen und bei der Beurteilung vollständig auf die Leistungsmessung verzichten. Zweitens bestreite ich heute ein Auswärtsspiel – nicht etwa, weil ich in Zürich beheimatet bin; einer Stadt, die heute noch nicht dem Bildungsraum Nordwestschweiz angehört –, sondern weil ich als empirischer Bildungsforscher auf Unterstützung von aussen angewiesen bin, wenn es darum geht, Sie davon zu überzeugen, dass die standardisierte Leistungsmessung das Leben an Gymnasien durchaus bereichern kann, vielleicht sogar lebenserhaltend wirken kann.

Wer die Literatur zum Thema «Leistungsmessung an Gymnasien» verfolgt, dem ist vermutlich wie mir aufgefallen, dass sich die Begeisterung der Betroffenen in Grenzen hält und nahezu jeder Aufsatz zum Thema sinngemäss wie jener von Jürgen Oelkers über gemeinsames Prüfen an Gymnasiums beginnt: «Den Ausdruck Bildungsstandards hören Gymnasialkollegien nicht gerne. Die Reaktion der Lehrerschaft auf Standards wird oft

---

<sup>1</sup> Referat an der Auftaktveranstaltung «Bildungsraum Nordwestschweiz und Programm Mittelschulen» vom 17. November 2008 in Solothurn

von Skepsis und Abwehr begleitet ... Schule nach der Industrienorm halten zu wollen, sei ein gefährlicher Irrweg ...»<sup>2</sup>.

Ich werde Ihnen im ersten Teil meiner Ausführung drei Gründe anführen, weshalb die Leistungsmessung auch an Gymnasien bei angemessenem Einsatz durchaus sinnvoll, ja sogar notwendig ist und nicht zwingend auf Irrwege führen muss. Im zweiten Teil werde ich auf das Dilemma der Leistungsmessung an Gymnasien und auf Schwierigkeiten eingehen, die die Grenzen der Leistungsmessung an Gymnasien aufzeigen. Abschliessend werde ich notwendige Rahmenbedingungen für die erfolgreiche Einführung der standardisierten Leistungsmessung an Gymnasien zur Diskussion stellen.

## **1. Chancen der Leistungsmessung an Gymnasien**

Mit der standardisierten Leistungsmessung werden in der Regel drei Funktionen erfüllt. Standardisierte Tests und Prüfungen werden erstens zur individuellen Beurteilung eingesetzt, zweitens zur Evaluation und Entwicklung von Schule und Unterricht und drittens als Teil des Monitorings von Bildungssystemen.

### *Individuelle Beurteilung*

Leistungstests zur individuellen Beurteilung werden vor allem an den Schnittstellen des Bildungssystems eingesetzt. Tests sind Bestandteil von Zulassungsverfahren oder von Abschlussprüfungen am Ende der Schulzeit. Wie schwierig es ist, die Fähigkeiten von Schülerinnen und Schülern fair und mit gleichem Massstab zu beurteilen, zeigt eine kleine Auswertung zur zentralen Aufnahmeprüfung für die Langgymnasien des Kantons Zürich. Bis 2006 wurden die Aufnahmeprüfungen an die Mittelschulen des Kantons Zürich von jeder Mittelschule selbstständig durchgeführt. Seit 2007 werden einheitliche Aufnahmeprüfungen eingesetzt<sup>3</sup>. Jeder Schüler, jede Schülerin löst die gleichen Prüfungsaufgaben eines Mathematiktests und eines Sprachtests, der Textverständnis, Wortschatz und Grammatik prüft, und jeder Schüler, jede Schülerin verfasst einen Text. Die Korrektur wurde standardisiert, Mathematik- und Sprachtest sogar einer teststatistischen Gütekontrolle unterzogen. Bei der Bewertung der Texte wurden – mit Murren – einheitliche Beurteilungskriterien angewendet, allerdings ohne die individuellen Beurteilungsmassstäbe der korrigierenden Lehrpersonen oder Kollegien zu berücksichtigen. Die einzige Vorgabe zur Verhinderung allzu grosser Benachteiligungen oder Bevorzungen bei der Korrektur war, dass die Durchschnittsnoten aller beurteilten Texte einer Schule zwischen 3,3 und 3,8 liegen mussten.

In einer Analyse der Durchschnittsnoten pro Schule konnten wir dann feststellen, dass jene Schule, die die höchsten Prüfungsergebnisse im Mathematik- und im Sprachtest erreichte, bei den Texten mit einer Durchschnittsnote von 3,4 unerwartet schlecht abschnitt. In anderen Schulen war es genau umgekehrt. Die Durchschnittsnoten der beurteilten Texte waren wesentlich höher, nahe bei 3,8. Die Durchschnittsnoten im Mathematik- und Sprachtest gehörten hingegen zu den tiefsten im Kanton. Der hinlänglich

---

<sup>2</sup> Oelkers, J. (2008). Gemeinsames Prüfen als Standard des Gymnasiums. In: W. Mahler (Hrsg.), *Gemeinsame Prüfungen. Absichten – Erfahrungen – Perspektiven* (S.17–39). Bern: hep.

<sup>3</sup> Keller, H. (2008). Gemeinsam Prüfen: Aufnahmeprüfung für das Gymnasium. In: W. Mahler (Hrsg.), *Gemeinsame Prüfungen. Absichten – Erfahrungen – Perspektiven* (S.40–44). Bern: hep.

bekannte positive Zusammenhang zwischen Wortschatz, Grammatik, Leseverständnis einerseits und produktiven Sprachkompetenzen andererseits wurde durch die willkürliche Beurteilung boykottiert. Je besser die durchschnittlichen Prüfungsergebnisse im Sprachtest an einer Schule waren, desto schlechter wurden an dieser Schule die Texte bewertet. Oder anders formuliert. Je besser die Prüfungsergebnisse im Sprachtest waren, desto strenger wurden die Texte bewertet.

Was bedeutet dieses Ergebnis für eher sprachschwache Schülerinnen und Schüler, die aufs Gymnasium möchten? Die Erfolgsaussichten sind an jenem Gymnasium am grössten, bei dem die Durchschnittsnoten im Sprachtest am tiefsten ausfallen.

Und was zeigt dieses Beispiel? Trotz gleicher Prüfungen wird der einzige Spielraum, der bei der zentralen Aufnahmeprüfung den Schulen bei der Beurteilung von Texten noch bleibt, willkürlich ausgenutzt. Tragisch ist dieser Zustand – nicht für die Gymnasien, aber für die betroffenen Schülerinnen und Schüler. Mit einer höheren Standardisierung der Korrektur könnte dieses Faktum einfach aus der Welt geschafft werden. Für die Lehrpersonen einer Schule könnte die Standardisierung allerdings zur Folge haben, dass weniger Schülerinnen und Schüler die Aufnahmeprüfung bestehen und weniger Stellenprozent als vorgesehen zur Verfügung stehen. Es ist zu hoffen, dass dieses Kartell endlich aufgegeben wird zugunsten fairer Aufnahmeprüfungen und eines redlichen Wettbewerbs unter den Gymnasien.

Nicht nur die Zulassung zu den Gymnasien, auch die Zulassung zu den Universitäten und Hochschulen liegt in der Hand der Gymnasien, die die Maturitätsprüfungen durchführen und die Maturitätsnoten festlegen. Gymnasien haben das Privileg zu bestimmen, was Hochschulreife ist und was nicht. Dieses Privileg wird zwar nicht grundsätzlich bestritten, jedoch in seiner radikalen Form vermutlich auch nicht unantastbar bleiben, wie das Jürgen Oelkers in seiner Expertise über die Qualität der Schweizer Gymnasien festhält<sup>4</sup>. Die Wahrscheinlichkeit, dass sich diese Prognose in Bälde erfüllt, stuft ich aufgrund von empirischen Daten als relativ hoch ein.

Die Grafik zeigt Ihnen den Zusammenhang zwischen dem Anteil an Schülerinnen und Schülern, die in einem Kanton in der 9. Klasse den Schultyp mit hohen Anforderungen besuchen, und den Leistungen im PISA-Mathematiktest. Höhere Anforderungen bedeutet in der 9. Klasse meistens Gymnasium, manchmal Bezirksschule oder Sekundarschule mit progymnasialem Niveau, was in der Romandie als «Voie secondaire de baccalauréat» oder «exigence élevée» bezeichnet wird. Die Position der Kantone ergibt sich aus dem Anteil Schülerinnen und Schüler im Schultyp mit hohen Anforderungen (X-Achse) und dem kantonalen Mittelwert im PISA-Test (y-Achse).

Ein schönes Bild, das man in den Sozialwissenschaften selten antrifft. Wer glaubt, dass der Besuch eines Schultyps zugleich etwas über die Leistungsfähigkeit eines Schülers oder einer Schülerin aussagt, der irrt.

Eine Standardisierung der Hochschulreife ist nur schon deshalb gerechtfertigt, weil sich die Maturitätsquoten zwischen den Kantonen zum Teil beträchtlich unterscheiden. Diese Unterschiede können zum einen auf Unterschiede bei der Nachfrage, zum andern auf Unterschiede bei den Anforderungen zurückgeführt werden. Bereits die Thirt Internati-

---

<sup>4</sup> Oelkers, J. (2008). *Die Qualität der Schweizer Gymnasien. Eine Expertise zuhanden der Bildungsdirektion des Kantons Zürich*. Bern: hep.

onal Mathematics and Science Study TIMSS hat vor mehr als 10 Jahren gezeigt, dass zwischen den durchschnittlichen Mathematik- und Physikleistungen der Maturandinnen und Maturanden und der Maturandenquote eines Kantons ein negativer Zusammenhang besteht. Je höher die kantonale Maturandenquote war, desto tiefer waren die durchschnittlichen kantonalen Testleistungen<sup>5</sup>.

Die durchschnittlichen Leistungen der Gymnasien im Kanton Basel-Stadt werden immer tiefer sein als jene der Gymnasien im Kanton Aargau, weil die Maturandenquote im Kanton Basel-Stadt bei knapp 30 Prozent, im Kanton Aargau bei 14 Prozent liegt. Hätte das Gesetz «Je höher die Maturitätsquote, desto tiefer die durchschnittlichen Leistungen der Gymnasien» keine Gültigkeit, dann würde die Schweiz aufgrund des im internationalen Vergleich geringen Anteils an Akademikerinnen und Akademikern nicht zu den hoch entwickelten und wirtschaftlich erfolgreichen Dienstleistungsländern Europas gehören.

Die Leistungsmessung könnte zu einem leistungsgebundenen und somit gerechteren Übergang von den Maturitätsschulen an die Hochschulen führen. Je mehr zudem die Ausbildung und die Vergabe von Zugangsberechtigungen an Maturitätsschulen standardisiert sind, desto weniger wird das Prinzip, «dass nicht prüft, wer ausbildet», kritisiert werden. Dies ist einer der Hauptgründe, weshalb Tests auch an Maturitätsschulen eingeführt werden sollen, allerdings vor allem mit dem Ziel der individuellen Beurteilung. Die kompetenzbezogene Zertifizierung der Leistungen wird auch in der Expertise über die Qualität der Schweizer Gymnasien als wichtiger Entwicklungsschritt für Gymnasien betrachtet<sup>6</sup>.

### *Individuelle Beurteilung*

Zur individuellen Beurteilung gehören nicht nur Tests als Teil von Zulassungs- oder Abschlussprüfungen, sondern auch Tests, mit denen Sprachdiplome erworben werden können, oder Tests, die für das «International Baccalaureate» durchgeführt werden<sup>7</sup>. Es gibt in der Schweiz bereits eine stattliche Anzahl an Gymnasien, die ihre Attraktivität dank der Nutzung von Angeboten der standardisierten Leistungsmessung erhöhen. Die Diplomprüfungen für das «International Baccalaureate» finden zwar an den Maturitätsschulen statt, die Aufgaben werden aber zentral entwickelt und die Arbeiten von Korrektoren der Organisation in Genf beurteilt. Es ist folglich durchaus möglich, standardisierte, schriftliche Tests am Ende der Sekundarstufe II durchzuführen, die auch bei den Gymnasien auf hohe Akzeptanz stossen und mit der Schweizer Maturität vereinbar sind. Zentrale Abschlussprüfungen kennen beispielsweise auch Länder wie Holland, Finnland oder Frankreich<sup>8</sup>. Sprachdiplome wie DELF oder DALF (Diplôme approfondi de langue française) beziehungsweise FIRST oder ADVANCED (First Certificate in English; Certificate in Advanced English) werden bereits an einigen Maturitätsschulen

---

<sup>5</sup> Ramseier, E., Keller, C. & Moser, U. (1999). *Bilanz Bildung. Eine Evaluation am Ende der Sekundarstufe II auf der Grundlage der «Third International Mathematics and Science Study»*. Chur, Zürich: Rüegger.

<sup>6</sup> (ebd., S. 16).

<sup>7</sup> ([www.lgr.ch/programm/ib/info\\_ib.pdf](http://www.lgr.ch/programm/ib/info_ib.pdf))

<sup>8</sup> van Ackeren, I. (2003). *Evaluation, Rückmeldung und Schulentwicklung. Erfahrungen mit zentralen Tests, Prüfungen und Inspektionen in England, Frankreich und Niederlanden*. Münster: Waxmann.

erfolgreich eingesetzt. Sie mögen aus der Sicht einiger Fremdsprachehrpersonen mit einer gymnasialen Bildung wenig zu tun haben. Ein Vorteil im Wettbewerb mit Privatschulen sind sie auf jeden Fall.

### *Unterrichtsentwicklung*

Es sind nicht nur die Schülerinnen und Schüler, sondern auch Lehrpersonen und Schulen, die von der Leistungsmessung profitieren können. Leistungstests zur Evaluation und Entwicklung von Schule und Unterricht sind eine Folge der ergebnisorientierten Schulreform, deren Ziel insbesondere die Stärkung der Einzelschule ist. Schulen werden mit mehr Kompetenzen ausgestattet und übernehmen mehr Verantwortung für pädagogisches Handeln, was mit der Auflage verbunden ist, Transparenz in Bezug auf Curricula und Schulentwicklung sowie über die Lernergebnisse der Schülerinnen und Schüler zu schaffen. Qualität muss im Sinne der Rechenschaftslegung ausgewiesen werden. Dahinter steht die Annahme, dass Testsysteme, die aus Informations- und Anreizsystemen bestehen, nicht nur positiv zur Leistungssteigerung eines Bildungssystems beitragen, sondern geradezu notwendig erscheinen, damit sich das schulische Personal dem Ziel der Schulentwicklung verpflichtet fühlt<sup>9</sup>. In den meisten Ländern, die diesen Modernisierungsprozess bereits vollzogen haben, hat deshalb die Leistungsmessung zur Optimierung der individuellen Förderung sowie zur Schul- und Unterrichtsentwicklung eine wichtige Stellung eingenommen. Die Evaluation wird zusehends mit der Entwicklung von Schulen verbunden und Sanktionen werden nicht strafend, sondern unterstützend ergriffen<sup>10</sup>. Erfolgreiche Beispiele lassen sich auch an Schweizer Gymnasien finden, wie uns heute Nachmittag noch gezeigt wird.

### *Systemmonitoring*

Schliesslich können Leistungstests als Teil des Monitorings zu wichtigen Informationen über ein Bildungssystem führen und Grundlagen für bildungspolitische Entscheide und öffentliche Diskussionen schaffen<sup>11</sup>. Am besten hat diese Funktion der Leistungsmessung bis anhin der internationale Schulleistungsvergleich PISA erfüllt, der nicht nur Schwächen unserer Jugendlichen im Lesen aufgedeckt hat, sondern auch eine bedeutende politische und öffentliche Diskussion über Bildungssysteme und Schule in Gang brachte. Ein Bildungsmonitoring auf Sekundarstufe II auf nationaler Ebene ist jedoch noch nicht in Sichtweite. Ganz so einfach ist die Sache mit der Leistungsmessung doch nicht, womit ich beim Dilemma und den Grenzen angelangt bin.

## **2. Grenzen**

### *Reduktion auf das Messbare*

Relativ einfach lässt sich die Leistungsmessung an Gymnasien einführen, wenn die gängige individuelle Beurteilung durch standardisierte Testergebnisse mit dem Ziel ergänzt werden soll, offizielle Diplome oder Zertifikate auszustellen. Die Möglichkeit für

---

<sup>9</sup> (ebd., S. 182)

<sup>10</sup> Klieme, E. (2007). Bildungsstandards, Leistungsmessung und Unterrichtsqualität. In Labbudde, R. (Hrsg.), *Bildungsstandards am Gymnasium. Korsett oder Katalysator?* (S. 75–84). Bern: hep.

<sup>11</sup> Bildungsbericht Schweiz (2006). Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.

die Schülerinnen und Schüler der Sekundarstufe II, international anerkannte Sprachdiplome abzulegen, führt zweifelsohne zu einer angemessenen Beurteilung der Fremdsprachenkompetenzen und zu einer Aufwertung der Ausbildungsgänge.

Für die übrigen Kompetenzen fehlen aber Instrumente zur standardisierten Leistungsmessung und es sind auch keine solchen absehbar, sieht man einmal von einzelnen kantonalen Orientierungsarbeiten oder vom Testsystem «eprolog» ab. So gut diese Initiativen gemeint sind, so gross ist die Gefahr, dass solche Prüfungen oder Tests allzu stark durch ökonomische Faktoren bestimmt werden. Ich erinnere Sie an die Bewertung der Texte im Rahmen der zentralen Aufnahmeprüfungen an die Langgymnasien des Kantons Zürich. Die Standardisierung der Mathematik- und Sprachtests wurde ohne «Wenn und Aber» eingeführt, auf eine gleichwertige Standardisierung bei der Beurteilung der Texte wurde hingegen verzichtet. Die Gründe dafür sind nahe liegend. Der Aufwand wäre zu gross. Es müssten Korrekturteams geschult werden, die Schulen könnten die Texte ihrer Prüflinge nicht mehr selbst korrigieren, eine bestimmte Anzahl der Texte müsste doppelt korrigiert werden, damit Verzerrungen aufgrund zu milder oder zu strenger Massstäbe rechnerisch neutralisiert werden könnten. Dies alles gehört zum «State of the Art» und wird beispielsweise von den meisten zertifizierenden Testinstituten angeboten, ist aber aufwändig und teuer.

Häufig wird die standardisierte Leistungsmessung zudem mit einem Aufgabenformat gleichgesetzt, beispielsweise mit Multiple-choice-Aufgaben. Dies ist Unsinn. Messen heisst, einer Person in Bezug auf eine spezifische Eigenschaft in systematischer Weise Zahlen zuzuordnen. Diese systematische Zuordnung lässt sich auch bei Texten anwenden, die von Schülerinnen und Schülern verfasst werden. Messen ist nicht an bestimmte Aufgabenformate gebunden. Aufgabenformate sind aber ein ökonomischer Faktor bei der Leistungsmessung. 1000 Texte von Schülerinnen und Schülern zuverlässig zu korrigieren ist wesentlich teurer, als die Angaben von 1000 Tests mit je 20 Multiple-Choice-Aufgaben elektronisch zu erfassen und auszuwerten. Noch billiger wird es, wenn die Tests gleich vom Computer korrigiert werden. Nur ist der Computer bei der Korrektur von Schülerlösungen immer noch sehr eingeschränkt, weshalb Onlinetests zwar eine billige Alternative zu aufwändigen Testverfahren sein können, nicht aber zur Entschärfung des eingangs erwähnten Dilemmas führen werden.

#### *Fehlende Referenzsysteme*

Tests können dann einen Nutzen für die Unterrichtsentwicklung haben, wenn durch sie Klarheit, Transparenz und Strukturiertheit im Unterricht erreicht werden<sup>12</sup>. Diese Forderung von Klieme, dem Verfasser der Expertise über Bildungsstandards in Deutschland, kann momentan auf der Sekundarstufe II nicht eingelöst werden.

Wenn standardisierte Tests mit dem Ziel eingesetzt werden sollen, den Unterricht zu verbessern, dann müssen diese zumindest auf den Unterricht – besser noch auf die Lehrpläne und die Lehrmittel – abgestimmt sein. Doch wer kann konkret sagen, welches Wissen und Können an Gymnasien unterrichtet und von den Schülerinnen und Schülern erwartet wird? Am ehesten die Lehrperson, die unterrichtet, vielfach auch die Fachschaft einer Schule, vielleicht auch noch die Schülerinnen und Schüler.

---

<sup>12</sup> Klieme, E. (2007). *Bildungsstandards, Leistungsmessung und Unterrichtsqualität*. In Labbudde, R. (Hrsg.), *Bildungsstandards am Gymnasium. Korsett oder Katalysator?* (S. 75–84). Bern: hep.

In Anbetracht der zurzeit fehlenden Referenzsysteme und der geringen Standardisierung des Bildungsangebots können vorläufig nur regional begrenzte Tests für die Schul- und Unterrichtsentwicklung eingesetzt werden. Denn damit Standards mit Tests auf nationaler Ebene überprüft werden können, muss vorgängig mit einem aufwändigen Verfahren der Bezug zwischen Referenzsystemen oder Standards und Tests empirisch hergestellt werden. Dieser Bezug kann ohne stufengerechte fachwissenschaftliche und fachdidaktische Diskussion und ohne Anwendung psychometrischer Verfahren nicht erreicht werden<sup>13</sup>. Das regionale Vorgehen bei der Leistungsmessung kann als Nachteil betrachtet werden. Zurzeit ist es aber der einzige gangbare Weg und zudem eine ausgezeichnete Möglichkeit für den Bildungsraum Nordwestschweiz, sich zu profilieren und die Leistungsmessung in der Schweiz in die richtigen Bahnen zu lenken.

Damit dies gelingt, müssen nicht nur Ressourcen für schwer messbare Kompetenzen zur Verfügung gestellt und Referenzsysteme entwickelt werden, sondern auch die Ziele der Leistungsmessung und die Rahmenbedingungen geklärt werden.

### 3. Rahmenbedingungen

Die erfolgreiche Einführung der Leistungsmessung in der Schule verlangt aufgrund der Erfahrungen aus anderen Ländern wie den USA<sup>14</sup> in einem ersten Schritt klare (politische) Entscheidungen über

- (1) die Ziele, die mit den Tests verfolgt werden (individuelle Beurteilung, Unterrichtsentwicklung, Schulevaluation, Monitoring),
- (2) die Verwendung der Testergebnisse (Von wem werden die Daten eingesehen und wozu werden sie verwendet?) und
- (3) über die Konsequenzen, die mit den Testergebnissen verbunden werden (Was geschieht, wenn die Ergebnisse die Erwartungen nicht erfüllen?).

Bevor diese Fragen nicht beantwortet und die wichtigsten Abläufe von der Testentwicklung bis zur Verwendung der Ergebnisse geklärt sind, könnten Projekte zur standardisierten Leistungsmessung auf der Sekundarstufe II kontraproduktiv sein und von der Praxis erfolgreich korrumpiert oder verhindert werden.

Kaum Regelungsbedarf besteht, wenn an den Gymnasien Tests zur Ergänzung der individuellen Beurteilung eingesetzt werden. In einer ersten Phase sollten am Gymnasium deshalb vor allem bestehende Tests genutzt werden, die es allen Schülerinnen und Schülern ermöglichen, Diplome oder Zertifikate zu erreichen.

Werden Tests hingegen zur Qualitätssicherung und zur Evaluation von Schulen eingesetzt, dann sind sehr viele offene Fragen in Bezug auf die Ziele der Tests, die Verwendung der Testergebnisse und die mit den Testergebnissen verbundenen Konsequenzen

---

<sup>13</sup> Köller, O. (2007). Bildungsmonitoring und Vergleichsarbeiten in Deutschland. In Labudde, R. (Hrsg.), *Bildungsstandards am Gymnasium. Korsett oder Katalysator?* (S. 65–73). Bern: hep.

<sup>14</sup> Baker, E.L. (2004). *Aligning Curriculum, Standards, and Assessments: Fulfilling the Promise of School Reform*. CSE Report 645. Center for the Study on Evaluation. National Center for Research on Evaluation, Standards, and Student Testing. University of California. Verfügbar unter: <http://cresst96.cse.ucla.edu/index4.htm> (14.3.2005).

zu klären, zumal die Definition von zu testenden Fähigkeiten und entsprechenden Testaufgaben anspruchsvoll und der Aufwand für die Entwicklung guter Testinstrumente gross ist. Zudem ist es nicht ganz einfach, Testergebnisse mit Merkmalen guter Schulen und guten Unterrichts in Beziehung zu setzen. In diesem Bereich dürfen Lehrpersonen zu Recht erwarten, dass die mit der Leistungsmessung verbundenen Ziele und die Konsequenzen transparent sind.

Klare Regelungen, die politisch festgelegt und mit den Schulen vereinbart werden, entscheiden darüber, ob die mit der Leistungsmessung angestrebten Ziele tatsächlich erreicht werden können. Sofern die Rahmenbedingungen geklärt sind, kann sich das Gymnasium getrost an die Arbeit machen und seine Lebendigkeit auch im Bereich der Leistungsmessung unter Beweis stellen. Dies war eine der Forderungen des eingangs erwähnten NZZ-Kommentars mit dem Titel «Das Gymnasium lebt». Der Kommentar bezog sich auf einen Beitrag mit dem Titel «Die Mittelschulen in Revision». Denn der Mut zur Revision ist für all jene Institutionen eine Notwendigkeit, die von sich behaupten, junge Menschen auf dem Weg auf ihre eigenen Füsse unterstützen zu können.