

## Bildungsstandards und zentrale Kompetenztests:

### Ansprüche, Probleme, Perspektiven<sup>1</sup>

Ms. für: Recht und Bildung (2005)

Kerncurricula, „Bildungs“standards und Kompetenztests haben Konjunktur in der Reformdebatte. Aber können sie die Ansprüche erfüllen, die mit ihrer Einführung verbunden werden? Gehen wir die zentralen Erwartungen der Reihe nach durch.

#### 1. Vereinheitlichung des Unterrichts mit dem Ziel höherer Chancengleichheit

Immer wieder kritisieren Eltern(verbände) die regionalen Unterschiede des Unterrichts, wobei interessanterweise die Differenzen *zwischen* den Bundesländern, weniger dagegen die Streuung *innerhalb* der einzelnen Länder Beachtung finden<sup>2</sup>.

Allerdings stellt sich die Frage, ob eine stärkere zentrale Kontrolle tatsächlich die entscheidenden Bedingungen vor Ort in den Griff bekommen kann. Erfahrungen in anderen Bildungssystemen stimmen skeptisch.

Nach der politischen „Wende“ 1989 ergab sich die seltene *Gelegenheit*, Schulleistungen von SchülerInnen aus zwei Bildungssystemen zu vergleichen, die sich in der Intensität zentraler Kontrolle deutlich unterschieden. In unserem Schreibvergleich BRD-DDR haben wir dies für die Grundschule am Beispiel der Rechtschreibung im Diktat und in freien Texten (Klassen 1 bis 4) getan. Dabei haben wir in den Diktaten aus dem geübten Grundwortschatz in der DDR-Stichprobe tatsächlich eine geringere Streuung der Fehlerquote festgestellt als in den gleichen Diktaten in der BRD. Aber in freien Texten

---

<sup>1</sup> In den folgenden Beitrag habe ich Textstücke übernommen aus: Brügelmann, H. (2004): Kerncurricula, Bildungsstandards und Leistungstests: Zur unvergänglichen Hoffnung auf die Entwicklung der guten Schule durch Evaluation „von oben“. In: Vierteljahrsschrift für wissenschaftliche Pädagogik, 80. Jg., H. 4, 415-441; und: Brügelmann, H. (2005): Schule verstehen und gestalten - Forschungsbefunde zu Kontroversen über Erziehung und Unterricht. Libelle: CH-Lengwil, Kap. 46-49.

<sup>2</sup> Vgl. dazu die starke Überlappung der Verteilungen selbst von Bayern und Bremen in Bos u. a. (2004, 61)

war die Bandbreite der Rechtschreibleistungen in beiden Systemen etwa gleich groß<sup>3</sup> - obwohl der Rechtschreibunterricht in der DDR bis ins Detail vorgeschrieben und kontrolliert wurde. Auch die internationale Lesestudie der IEA von 1991 stellte für die 9-Jährigen in der DDR eine vergleichbare Streuung fest wie für die gleichaltrige Vergleichsgruppe in der BRD<sup>4</sup>.

Im englischen Bildungssystem, das die Einhaltung seines hoch strukturierten „national curriculum“ mit regelmäßigen landesweiten Tests kontrolliert, ist die Streuung der Schülerleistungen noch größer als in Bremen, dem deutschen Bundesland, das bei IGLU-E die größte Streuung aufzuweisen hatte (vgl. Bos u. a. 2004, 63).

Die Analysen von PISA zeigen außerdem, dass zentrale Prüfungssysteme nicht generell zu höheren Leistungen oder einer geringeren Streuung führen (Baumert u. a. 2000, 345-346).

## 2. Sicherung einer gemeinsamen Grundbildung für alle

Die VerfechterInnen eines Kanons übersehen, dass derselbe Inhalt für verschiedene SchülerInnen mit ihrer je individuellen Lernbiografie *Unterschiedliches bedeutet* - und umgekehrt, dass sich an verschiedenen Inhalten durchaus vergleichbare Erfahrungen gewinnen lassen. Warum kann ein Deutschunterricht 17-jährigen SchülerInnen nicht die Wahl lassen, welches Liebesgedicht (Barockgedicht oder ...) sie für sich lernen wollen - solange er sie darauf verpflichtet, sich über ihre Gründe und über ihre Lernerfahrungen auszutauschen. Dasselbe gilt für Sachthemen in der Grundschule. Statt konkreter Inhalte wären also Kategorien oder Kriterien festzulegen, die unterschiedlich abstrakt denkbar sind. Es geht dabei um eine hohe Qualität der *Lernmöglichkeiten*, also primär um die Bildungsansprüche der Kinder, nicht nur um Qualifikationsanforderungen der Gesellschaft (vgl. unten den Vorschlag des Grundschulverbands 2003).

Je konkreter und enger Inhalte vorgegeben werden, um *Gemeinsamkeit* zu sichern, umso schärfer stellt sich zudem das Problem des Transfers auf andere Inhalte und Situationen.

---

<sup>3</sup> In Diktaten streuten die Fehlerquoten auf den verschiedenen Jahrgangstufen der Grundschule in der DDR (11.1 - 19.6) weniger als in der BRD (14.7 -26.6.), in freien Texten war die Streuung etwa gleich (5.6, 10.5, 13.5, 22.9 vs. 6.8, 8.6, 13.2, 23.4 in den vier Schuljahren der Grundschule), s. Brügelmann/ Richter (1994, 137).

<sup>4</sup> Vgl. Lehmann u.a. (1995, 216); in den 8. Klassen der (Ost-)Gesamtschule war die Streuung allerdings geringer als im dreigliedrigen Schulsystem West (a. a. O., 219) – ein Effekt, der nach den Befunde in der Grundschule aber wohl eher auf die unterschiedliche Schulorganisation als auf das Steuerungssystem zurückzuführen ist.

Die aktuelle Diskussion über empirische Studien zum Bildungswert des Lateinischen macht dieses Dilemma exemplarisch deutlich<sup>5</sup>.

### 3. Gewährleistung von Mindestniveaus der Leistungen in zentralen Lernbereichen

Das Problem unzureichender Leistungen vieler SchülerInnen, z. B. im Lesen, haben PISA und IGLU offenkundig gemacht<sup>6</sup>. Aber kann der propagierte Transfer von Systemelementen aus anderen Ländern dieses Problem wirklich lösen? Bei IGLU hat beispielsweise Bremen in Deutschland am schlechtesten abgeschnitten - bedingt vor allem durch besonders schwache Leistungen im unteren Bereich. International entspricht aber das unterste Niveau (Prozentrang 5) den Leistungen, die z. B. auch in Ländern wie Neuseeland (das internationale Vorzeigesystem für Leseförderung<sup>7</sup>) und in England (Platz 3 bei IGLU) bzw. den USA (Platz 9) erreicht werden - obwohl es in den letzteren beiden Staaten seit 10-30 Jahren Standards und regelmäßige Leistungstests auf überregionaler Ebene gibt.

Leistungsstreuung ist ein Phänomen, dass sich in allen Bildungssystemen findet (s. oben 2.). Unsere Lesestudie LUST in NRW<sup>8</sup> zeigt, dass von Klasse 2 bis 4 die schwachen 10% eines Jahrgangs den starken 10% um drei bis vier Schuljahre hinterherhinken - trotz Zurückstellung von 5-10% am Schulanfang, trotz des Sitzenbleibens von weiteren rund 5% pro Schuljahr und trotz der Überweisung von rund 5% eines Jahrgangs auf Sonderschulen. Die IGLU-Daten machen deutlich, dass die Streuung in einigen anderen Ländern noch wesentlich höher ist als in deutschen Grundschulen<sup>9</sup>. Studien zum Schulanfang belegen, dass diese Bandbreite bereits in den unterschiedlichen Vorerfahrungen der Kinder begründet ist<sup>10</sup>. Wie will man angesichts dieser Heterogenität sinnvolle Schwellenwerte für Ende 2. oder 4. Klasse festlegen, ohne die einen heillos zu unterfordern und die anderen von vornherein ins Abseits zu stellen?<sup>11</sup>

Ein weiteres Problem kommt hinzu. Niemand kann empirisch begründet sagen, wo die Schwellenwerte für „tragfähige Grundlagen“ liegen, ohne die ein Schüler in der Sekundarstufe nicht erfolgreich weiterlernen kann („lernbiografische Validität“ von

---

<sup>5</sup> Vgl. die Kontroverse zwischen Haag/ Stern (2000), Strunz (2003) und Westphalen (2003) – s. auch unten (5.).

<sup>6</sup> Dass dies nicht mit einem „Leistungsverfall“ in den Schulen gleich zu setzen, sondern die Folge erheblich gestiegener schriftsprachlicher Anforderungen in Beruf und Alltag ist, habe ich an anderer Stelle dargestellt (vgl. Brügelmann 1999, 10-20, 120-126).

<sup>7</sup> Vgl. Rossa/ Rossa (1995).

<sup>8</sup> Brügelmann, H. (2003s).

<sup>9</sup> Bei IGLU streuten Ende 4. Klasse schon die Leistungen der mittleren zwei Drittel über rund zwei Schuljahre, die obersten 5% vs. die untersten 5% über 5-6 Schuljahre, vgl. Bos, W., u. a. (2004, 62-65)

<sup>10</sup> Brügelmann (1983, 201), Rabenstein u. a. (1989) und Richter (1992).

<sup>11</sup> Ähnlich die Kritik am Bundesgesetz „No child left behind“ in den USA; vgl. Marshak (2003).

Leistungsprädiktoren). Lernen ist zwar ein kumulativer, aber kein linearer Prozess, und die Testleistungen in einem Fach sind unzureichende Prädiktoren für die zukünftigen Fähigkeiten eines Menschen, da diese auch von deren Entwicklung als Person abhängen<sup>12</sup>.

Ebenso wenig ist geklärt, auf welchem Niveau *Basisqualifikationen* für Berufs- oder Lebenserfolg zu definieren sind („ökologische Validität“ von Leistungsprädiktoren<sup>13</sup>). In einer Pilotstudie haben wir unseren Test grundlegender Lesefähigkeiten in einer Handwerkskammer durchgeführt. Nach den dort erhobenen Leistungen könnten von der SchülerInnen der vierten Klasse schon fast 90% in der Grundfähigkeit mit HandwerkerInnen in Meister(!)kursen mithalten - selbst wenn man dort die schwächsten 5% außer Acht lässt.

#### 4. Verbesserung der Qualität von Unterricht

US-Bundesstaaten, die in den 90er Jahren die Vergabe von Abschlusszeugnissen, erfolgsabhängige Gehaltszuschüsse für LehrerInnen und die finanzielle Ausstattung der Schulen an die Ergebnisse in ihren jährlichen Tests gebunden haben, berichten immer wieder über einen Anstieg der Leistungen. Anfang 2000 hatten sich 28 Staaten, also mehr als die Hälfte diesem Konzept verpflichtet. Eine Studie von Amrein/ Berliner (2002) zeigt aber, dass der berichtete Leistungsanstieg in der Regel nur für den engen Bereich der im jeweiligen Bundesstaat etablierten Tests galt. In unabhängigen Tests, z. B. für die Zulassung zu den Hochschulen, zeigte sich für 2/3 der 28 Staaten eine *Abnahme* der Testleistungen. Zusätzlich wurden wachsende Dropout-Raten berichtet, d. h. dass leistungsschwächere SchülerInnen aus dem System ganz herausfielen, was die gemessenen Testleistungen zusätzlich in die Höhe trieb, ohne dass sich der „Output“ tatsächlich verbesserte<sup>14</sup>.

Auch zentrale Tests werden mit Erwartungen überfrachtet, Chancen der Selbstevaluation werden übersehen. Systemevaluationen sind wichtig, aber in einem differenzierten Verbund müsste einer unterrichtsnahen Evaluation, die neben der Erhebung von Wirkung auch die Qualität der Prozesse erfasst<sup>15</sup>, ein höheres Gewicht zugemessen werden. Auch

---

<sup>12</sup> Ausführlicher zum Prognoserisiko von Risikoprognosen: Brügelmann (2003d).

<sup>13</sup> Die AutorInnen der PISA-Studie begründen nur allgemein, dass „Literacy“ für den Berufs- und Lebenserfolg wichtig ist (Baumert u. a. 2001, 29-32); sie haben aber weder die Validität ihrer Instrumente noch die der gesetzten Schwellenwerte empirisch untersucht.

<sup>14</sup> Vgl. dazu die Zusammenfassung und den Kommentar von Winter in der New York Times v. 28.12.2002 und ausführlicher unten (8.).

<sup>15</sup> Vgl. einerseits den „Aufruf für einen Verbund reformpädagogisch engagierter Schulen“ zu einer substanzielleren Selbstevaluation ☞ [www.BlickUeberDenZaun.de](http://www.BlickUeberDenZaun.de) und andererseits die *Prozess-Standards* zur *opportunity-to-learn* des NCTM in den USA (ref. u. a. bei Klieme u. a. 2003, 24, 25, 28-30 - dort auch

die neuen finnischen Bildungsstandards definieren neben groben Kompetenzstufen, die erst auf Schulebene konkretisiert werden, Anforderungen an die Lernumgebung und den Stil des Unterrichts. Diese werden normativ (mit Bezug auf Prinzipien menschlichen Zusammenlebens) und nicht nur instrumentell (im Blick auf die Ergebnisse) begründet. Konkretisieren lassen sich solche Standards für guten Unterricht in Form von Kriterien für die Organisationsform und die materielle Umgebung, für die Vereinbarung von Regeln und die Qualität von Impulsen der Lehrperson.

## 5. Und wo bleibt das Konstruktive?

Kritisieren lässt sich leicht. Aber was ist die Alternative? Denn die diagnostizierten Probleme bestehen in der Tat: Die Formen der Rechenschaftslegung sind im deutschen Bildungswesen unterentwickelt.

Gemeinsam mit KollegInnen im Grundschulverband habe ich deshalb in den letzten fünf Jahren konkrete Vorschläge entwickelt, die sich auf drei zentrale Bereiche beziehen:

- einen Strukturvorschlag für eine nach Ebenen und Funktionen differenzierte Organisation von Rechenschaft im Schulsystem<sup>16</sup>;
- ein inhaltliches Konzept für Standards, bewusst formuliert als „Bildungsansprüche der Kinder“<sup>17</sup>, die Ziele als Entwicklungsdimensionen für eine individuelle Förderung nicht als Leistungsniveaus zu festgelegten Termen beschreiben; und
- eine Sammlung von alltagstauglichen Beobachtungshilfen für die Kernbereiche des Grundschulunterrichts<sup>18</sup>, die förderorientiert ausgelegt sind und sowohl die Evaluationskompetenz der LehrerInnen als auch die der SchülerInnen stärken:
  - Lern-Stände feststellen und vergleichend einordnen
  - Lern-Entwicklungen dokumentieren und bestätigen
  - Lern-Gespräche mit anderen führen
  - über eigene Lern-Wege nachdenken.

---

hoch gelobt - und trotzdem entschied sich ohne nähere Begründung für eine Output-Orientierung der Standards...).

<sup>16</sup> Die folgenden Vorschläge sind übernommen aus Brügelmann (1999), die auch eingegangen sind in die Konzeption des Grundschulverbandes für ein System der Qualitätssicherung, vgl. Bartritzky u. a. (1999)

<sup>17</sup> Vgl. Grundschulverband (2003): Bildungsansprüche von Grundschulkindern - Standards zeitgemäßer Grundschularbeit. Empfehlungen des Grundschulverbandes zur aktuellen Standard-Diskussion. Grundschulverband - Arbeitskreis Grundschule e.V. : Frankfurt.

<sup>18</sup> Brinkmann, E./ Brügelmann, H. (2005): Pädagogische Leistungskultur - Materialien für Klasse 1 und 2: Deutsch. Teilband 3 der Beiträge zur Reform der Grundschule 119. Grundschulverband: Frankfurt.

Das **Feststellen der Lernstände** hilft zum einen, die Bandbreite der Leistungen innerhalb der eigenen Klasse auszuloten, die in der Regel eine Verteilung über drei bis vier Entwicklungsjahre innerhalb einer Klasse zeigt. An den Randwerten dieser Bandbreite wird erkennbar, wie breit gefächert das Angebot an Aufgaben und Materialien sein sollte, damit alle Kinder die Chance erhalten, etwas Passendes zu finden, das ihnen einen Lernfortschritt ermöglicht. Zum anderen wird es im Vergleich mit einer standardisierten Stichprobe möglich, die eigene Einschätzung des Leistungsstandes der Klasse *insgesamt* und die persönliche Wahrnehmung *einzelner* Kinder zu überprüfen und eventuell zu korrigieren oder auch genauer hinzuschauen, wenn die eigene Einschätzung nicht zu den Auswertungsergebnissen passen.

Die **Lernentwicklungen der Kinder** werden nur durch den Vergleich von Beobachtungen des *einzelnen* Kindes zu *verschiedenen* Zeitpunkten sichtbar. Erst durch die Zusammenschau der dokumentierten Beobachtungen und einer Vielzahl von Arbeitsergebnissen, die unter ganz bestimmten Gesichtspunkten betrachtet werden, lassen sich Entwicklungen der Kinder schlüssig beschreiben und daraus Konsequenzen für eine spezifische Förderung oder Herausforderung ziehen. Wichtig ist dabei, dass die Kinder in die Beobachtung der Lernentwicklung mit einbezogen werden: Sie müssen erkennen können, dass sie Fortschritte machen, dass sich also ihre Mühe lohnt, auch wenn noch nicht alles perfekt ist. Wichtig ist dabei auch, dass sie gemeinsam mit ihrer Lehrerin festlegen, welche Entwicklungsaufgabe sie sich in einem nächsten Schritt vornehmen sollten.

Diese Anforderung führt schon zum nächsten Aspekt der Lernbeobachtung:

**Lerngespräche führen**, z. B über den „harten Brocken des Tages“ (Christa Erichson), ein orthografisch schwieriges Wort, an dem die Kinder ihre Strategien erproben, über die sie sich hinterher austauschen. Auf einer Metaebene können sich Kinder untereinander und mit ihrer Lehrerin darüber verständigen, was ihnen aus ihrer Sicht schon gut gelungen ist, was noch nicht so gut geklappt hat, welche Ziele sie sich gesetzt haben und ob sie mit ihrer Arbeit zufrieden sind.

Solche reflexiven Gespräche helfen mit, nicht nur das unterschiedliche Können zu würdigen, sondern auch zu begreifen, dass verschiedene Wege zum Ziel führen können. Die Kinder beginnen, über die **eigenen Lernwege nachzudenken**, sie mit denen von anderen Kindern zu vergleichen, neue Ideen auszuprobieren und sich bewusst mit dem eigenen Lernen auseinander zu setzen. Ein Portfolio mit der persönlichen Auswahl „meiner besten

Texte" oder Lerntagebücher fordern die Kinder heraus, über ihre eigenen Stärken und Schwächen nachzudenken.

Evaluation wird hier also als dialogischer Prozess verstanden - nicht als ein Urteil „von oben“. Mit den vorgeschlagenen Maßnahmen lässt sich nicht sichern, dass *alle* Kinder an *allen* Orten *alles* Wichtige lernen. Aber das ist eh' eine Schimäre, wie ich im ersten Teil dieses Beitrags gezeigt habe - von den zu erwartenden negativen Nebenwirkungen ganz abgesehen. Die Konzeption des Grundschulverbands zielt auf eine Qualifizierung aller Beteiligten - vor allem der LehrerInnen und SchülerInnen. Damit trägt sie zur Demokratisierung der Schule bei - und bildet ein bedeutsames Gegengewicht zur Expertokratie von PISA & Co.

Damit wird Evaluation also als Aufgabe von Spezialisten außerhalb des Unterrichts begriffen, sondern als Bestandteil von Lehr- und Lernaktivitäten im Unterrichtsalltag. Wir brauchen eine Differenzierung von Funktionen und Ebenen der Evaluation (vgl. Anm. 16):

**SchülerInnen** sollen sich um Klarheit über die eigenen Ziele und den Erfolg der eigenen Arbeit bemühen, z. B. durch die Verpflichtung auf bestimmte Ziele und durch die Selbstkontrolle von Aufgaben, aber auch durch regelmäßige Rechenschaftsberichte zum eigenen Lernfortschritt und zu den Zielen für die nächste Zeit.

**LehrerInnen** sollen im Blick auf einzelne SchülerInnen und die Klasse insgesamt **Lernprozesse** beobachten und auswerten, z. B. indem sie über die Entwicklung der SchülerInnen in einer Kartei mit Beobachtungsnotizen und Beispielarbeiten führen und Lernprozesse kontinuierlich dokumentieren, um mit den SchülerInnen individuell Wahrnehmungen und Erklärungsversuche zu ihren Fortschritten und Schwierigkeiten auszutauschen und gemeinsam Ziele für die weitere Arbeit zu vereinbaren.

**LehrerInnen** sollen zudem den eigenen **Unterricht** beobachten, um die Wirkungen ihrer Arbeit zu erfassen und zu überprüfen, z.B. die eigenen Ansprüche klären und ihre Realisierung überprüfen

- zunächst im „Selbstgespräch“ und durch Selbstbeobachtung (z.B. in Form eines Tagebuchs),

- danach durch Einladung eines „Fremdblicks“ (einer Kollegin, eines Praktikanten, einer Lehramtsanwärterin, von Eltern).

Das **Kollegium** soll im Blick auf die **Schule** das Schulprogramm entwickeln und fortschreiben, die **Schulleitung** die entsprechenden Aktivitäten anregen, fördern, absichern, z. B. indem sie periodisch verschiedene Beteiligte in der Schule befragt:

Die **Schulaufsicht** soll im Blick auf **Schulen** den Fremdblick stärken, Entwicklungen anerkennen, unterstützen, fordern, z.B. indem sie bei Schulbesuchen Gespräche mit der Schulleitung, LehrerInnen-Teams, einzelnen Lehrkräften zu folgenden Fragen führt:

- „Was sind Ihre eigenen Ansprüche?“
- „Wie weit können Sie diese realisieren?“
- „Welche Bedingungen behindern Sie dabei?“
- „Welche Ansprüche sind noch zu wenig im Blick?“
- „Was könnten realistische nächste Schritte sein?“

Das **Ministerium** soll mit Blick auf das **Gesamtsystem** allgemeine Problemstellen ausmachen und politische Prioritäten überprüfen, z.B. für die Sicherung der materiellen Voraussetzungen oder der rechtlichen Rahmenbedingungen Informationen gewinnen und aufbereiten, z. B. über Sekundärauswertungen von anfallenden Verwaltungsstatistiken und über die periodische Untersuchung des Leistungs- und Entwicklungsstands in anonymisierten Stichproben von SchülerInnen mit Erhebung von Eingangsvoraussetzungen, Prozessmerkmalen und Rahmenbedingungen des Unterrichts.

Landesweite Tests sind als *ein* Baustein in einem umfassenderen System zu sehen. Ihre Funktion ist ganz klar auf das „System Monitoring“ und auf die Bereitstellung von Referenzdaten für Evaluation vor Ort beschränkt. Solche Bestandsaufnahmen müssen deshalb auch nicht jedes zweite, dritte oder gar jedes Jahr stattfinden. Ihr Rhythmus kann längerfristig gedehnt werden, wegen der Trägheit des Systems reichen Wiederholungen alle 5-10 Jahre. Außerdem können sie sich auf repräsentative Stichproben beschränken und müssen nicht flächendeckend durchgeführt werden. Damit werden Mittel frei für schulnahe Evaluationsaktivitäten. Nur von ihnen könne wir eine tatsächliche Verbesserung des Unterrichts erwarten.