

# Leistungsstandards und Kompetenztests: eine pädagogische Sicht

Plädoyer für eine „pädagogische Leistungskultur“ durch *dialogische Lernberatung* und für eine Schulentwicklung mit Hilfe des „Blicks über den Zaun“ von *kritischen Freunden*

von Hans Brügelmann\*

Über die Funktion, das Format und die Folgen von Bildungsstandards kann man heutzutage nicht urteilen, ohne ihre enge Verbindung mit dem Einsatz von Kompetenztests mit zu diskutieren. Ich nehme deshalb einen kleinen Umweg zu meinem Thema und stelle es in einen weiteren Kontext.

## Bildungsstandards oder Leistungsstandards?

Die OECD begründet die Bedeutung von PISA und seinen Ergebnissen vor allem mit zwei Argumenten:

- *gesellschaftlich* mit der Notwendigkeit, „Bildung“ als Voraussetzung für die ökonomische Konkurrenzfähigkeit von Staaten zu verbessern,
- *individuell* mit der Annahme, „Bildung“ sei Voraussetzung für den persönlichen Berufs- und Lebenserfolg.

Beide Argumente erscheinen fragwürdig, betrachtet man die Entwicklung der Wirtschaftssysteme und des Arbeitsmarkts in den letzten Monaten bzw. Jahren.

Zur ersten Begründung: Die dramatische Finanzkrise im Herbst 2008 ist nicht die Folge unzureichender Fachqualifikation von Lehrlingen und damit den 15-25 Prozent sog. „RisikoschülerInnen“ anzulasten, die nach den PISA-Studien 2000, 2003 und 2006 in aller Munde sind. Zuzurechnen ist sie vielmehr der fehlenden Verantwortung von Hochqualifizierten, deren mathematische Kompetenz und Lesefähigkeiten in den üblichen Tests mit hoher Wahrscheinlichkeit in den oberen Prozenträngen liegen dürften.

An diesem Beispiel wird die immer wieder beklagte Verkürzung des Bildungsbegriffs in den Versuchen einer empirischen Kontrolle der „Qualität“ von Schule deutlich - etwa im Vergleich mit den sieben Perspektiven der Allgemeinbildung, wie Heymann (1996) sie formuliert hat:

- Lebensvorbereitung auf Alltagsanforderungen
- Sicherung des gesellschaftlichen Zusammenhalts („kulturelle Kohärenz“)
- Weltorientierung (materiales Wissen)
- Kritischer Vernunftgebrauch (methodisches Können)
- Verantwortungsbereitschaft
- Fähigkeit zur Verständigung und Kooperation
- Persönliche Identität („Stärkung des Schüler-Ichs“)

Die Begründung dieser sieben Dimensionen macht deutlich, dass jede von ihnen notwendig, aber nicht hinreichend für Bildung ist, so dass es um ihre Verbindung und nicht um die

-----

\* In: Hofmann, B./ Valtin, R. (Hrsg.) (2009: Projekte, Positionen, Perspektiven. 40 Jahre Deutsche Gesellschaft für Lesen und Schreiben DGLS. Deutsche Gesellschaft für Lesen und Schreiben: Berlin, 66-81.

plumpe Alternative geht: entweder „intelligenter Egoist“, der keine Verantwortung kennt, oder „tumber Gutmensch“, der nicht zureichend lesen, schreiben, rechnen kann. Meine Kritik zielt auf die Einseitigkeit der dominierenden instrumentellen Vernunft - sowohl in den *Kriterien* für den Erfolg der Schule als auch in den *Verfahren* seiner Überprüfung, d.h. *woran* und *wie* ihre Qualität „gemessen“ werden.

Vor allem die Übersetzung in standardisierte Tests führt (schon aus methodischen Gründen) dazu, dass die letzten drei, vier der Heymann'schen Perspektiven ausgeblendet werden. Bildung in diesen Dimensionen kann nicht - oder nur unzureichend - in standardisierten Aufgaben erfasst werden. Denn: WIE etwas gelernt wird, prägt mit, WAS tatsächlich gelernt wird.

Spezifische Fachleistungen zu untersuchen und zu bewerten ist nicht ehrenrührig - nur sollte man diese Aktivitäten dann begrifflich angemessen ausweisen und die Reichweite ihrer Ergebnisse entsprechend begrenzen. Auch bei den sog. „Bildungs“-Standards der KMK handelt es sich um „Leistungs“-Standards - und so sollten sie auch genannt werden.

Zum zweiten Argument: In der Tat gibt es viele Betriebe, die für ihre Ausbildungs- und Arbeitsstellen keine qualifizierten BewerberInnen finden. Es gibt aber auch viele qualifizierte ArbeitnehmerInnen, die keinen Arbeitsplatz bekommen oder die „betriebsbedingt“ gekündigt werden, obwohl sie erfolgreich gearbeitet haben. Zur Versachlichung der Diskussion wären Statistiken hilfreich, die ausweisen, welche Anteile der Arbeitslosigkeit tatsächlich auf unzureichende Qualifikation der BewerberInnen zurückzuführen sind. Zudem ist zu beachten: Eine höhere Qualifikation mag zwar individuell die Chancen erhöhen, Arbeit zu finden - aber auf einem engen Arbeitsmarkt immer zu Lasten anderer, deren bisher ausreichende Kompetenz durch das allgemein zunehmende Zertifikatsniveau abgewertet wird. Insofern leistet das zweite OECD-Argument der Gefahr Vorschub, das Problem der Arbeitslosigkeit zu pädagogisieren, statt seine gesamtwirtschaftlichen und damit gesellschaftspolitischen Wurzeln anzugehen.

Meine Probleme mit den KMK-Standards lassen sich in fünf Thesen zusammenfassen:

- Es handelt sich um (Fach-)Leistungsstandards - nicht um Bildungsstandards.
- Sie beschränken sich auf die Zielebene („Output“) - liefern keine Kriterien für die Qualität von Lehr-/Lernprozessen.
- Die Lernstandserhebungen reduzieren sie weiter auf das Testbare - statt den Unterricht pädagogisch anzuregen und herauszufordern.
- Die Auswertung orientiert sich einseitig an der sozial vergleichenden Bezugsnorm - statt dass die individuelle Leistungsentwicklung ausgewiesen und honoriert würde.
- Als Regelstandards errichten sie Selektionshürden - statt Anstrengungen zur optimalen Entfaltung der individuell unterschiedlichen Möglichkeiten zu fördern.

## **Kompetenztests und Evaluation**

Das Kernproblem ist die enge Anbindung der Standards an Tests, und deren Einsatz in einem Kontrollkontext, so dass sowohl LehrerInnen als auch SchülerInnen zu

Täuschungsversuchen verleitet werden. Formulierungen wie das Versprechen einer Schulministerin an die Eltern der ViertklässlerInnen bei der Einführung von VERA „Jetzt wissen Sie, wo Ihr Kind wirklich steht!“ suggerieren einen Autoritätsanspruch, dem Tests nicht gerecht werden können. Der berechtigte Widerstand von Lehrerinnen gegen eine solche Entmündigung kann dann leider auch dazu führen, dass sinnvolle Funktionen von Standards und Tests nicht mehr gesehen und ihr Einsatz ganz verweigert wird.

Der inzwischen gängige Vorwurf fehlender Diagnosekompetenz (z.B. bei den Empfehlungen am Ende der Grundschulzeit) vereinfacht komplexe Wahrnehmungs- und Beurteilungsprozesse unzulässig. Wenn Lehrerurteil und Testergebnis differieren, wird fast selbstverständlich unterstellt, dass die Einschätzungen der PraktikerInnen falsch sind. Könnte es nicht auch umgekehrt sein, dass die

- nur punktuellen
- inhaltlich ausschnittshaften
- in der Aufgabenform künstlichen

Tests in vielen Fällen irren?

Dazu in Stichwortform einige Fragen (vgl. zum Folgenden Brügelmann 2008, wo auch die empirischen Befunde im Einzelnen nachgewiesen sind).

*Wie sicher ist die ökologische Validität von Tests?*

In unserer LUST-Studie mit dem Stolperwörter-Lesetest für GrundschülerInnen erreichen mehr als 10 Prozent der Handwerker in Meisterkursen nur das Durchschnittsniveau von ViertklässlerInnen; selbst von LehrerInnen in Fortbildungen lesen einige im Test nicht schneller und fehlerfreier als durchschnittliche ViertklässlerInnen. Was bedeuten solche Ergebnisse für die Relevanz von Ergebnissen in Leistungstests für Anforderungen in Alltagssituationen?

*Wie weit ist es her mit der prognostischen Vorhersagekraft von Tests?*

Von den 15-jährigen SchülerInnen, die bei PISA in Dänemark als „Illiterates“ klassifiziert wurden, haben 20 Prozent den Sekundarschulabschluss erreicht; von denen, die in Kanada auf Stufe 1 oder darunter lagen, haben 62 Prozent den High-School-Abschluss gemacht. Selbst wenn man seine Zweifel an den Anforderungen in manchen Examina hat: Was bedeuten solche Befunde für die diagnostische Funktion von Lernstandserhebungen und zentralen Abschlussprüfungen?

*Wie verlässlich sind die in einzelnen Tests gewählten Indikatoren?*

In internationalen Lesestudien landen die USA in den letzten knapp 20 Jahren je nach Altersgruppe und Instrument im Vergleich mit denselben Ländern auf höheren, etwa gleichen oder niedrigeren Rangplätzen. Im US-internen National Assessment of Educational Progress (NAEP) sind die Leseleistungen über die letzten 30 Jahre hinweg fast auf den Punkt genau konstant geblieben. Was bedeutet diese Differenz für die Reichweite von Testergebnissen - insbesondere als Grundlage für Qualitätsurteile über Schulsysteme - und Schulen?

*Wie weit sind die Rangvergleiche zwischen den Bundesländern vergleichbar?*

In Baden-Württemberg und Bayern nahmen knapp über 85 Prozent (freiwillig) an PISA teil, in anderen Bundesländern bis zu 95 Prozent (verpflichtend). In den östlichen Bundesländern ist der Anteil der MigrantInnen und der SonderschülerInnen wesentlich höher als in den westlichen - und auch zwischen diesen gibt es erhebliche Unterschiede in den Anteilen und deren Zusammensetzung aus verschiedenen leistungsfähigen Teilgruppen. Was bedeuten diese Differenzen für die Repräsentativität der Ergebnisse innerhalb der einzelnen Bundesländer - und für die Vergleiche zwischen ihnen?

*Wie bedeutsam ist das bildungspolitische Steuerungs-System tatsächlich?*

Bei PISA sind die deutschen 15-Jährigen im (unteren) Mittelfeld gelandet. Bei IGLU finden sich die deutschen ViertklässlerInnen im oberen Leistungsdrittel. Beide Schulformen werden und wurden nach demselben Verfahren gesteuert und kontrolliert. Kann dieses dann für die unterschiedlichen Ergebnisse verantwortlich sein?

*Hat das sog. Input-System wirklich versagt?*

Die leichten Zuwächse bei den Wiederholungen von PISA und IGLU werden als Erfolg zunehmender Output-Steuerung durch die Orientierung auf nachweisbare Lerneffekte gewertet. Bei IGLU haben sich die deutschen GrundschülerInnen allerdings von 2001 bis 2006 nur um zwei Standardpunkte pro Jahr verbessert. Von der Leseuntersuchung IEA 1991 bis IGLU 2001 haben die Leseleistungen im internationalen Vergleich pro Jahr dagegen um doppelt so viele Standardpunkte zugenommen. Ist also die Output-Steuerung wirklich erfolgreicher - selbst an ihren eigenen Kriterien und mit ihren eigenen Methoden gemessen?

*Sind die Ergebnisse von Lernstandserhebungen vergleichbar?*

Wer Zugang zur Hinterbühne des Schulalltags hat, bekommt viel mit über die unterschiedliche Vorbereitung auf Tests (langfristig über Schwerpunkt- und Stildifferenzen des Unterrichts, aber auch unmittelbar durch inhaltliches oder methodisches „teaching to the test“) und über Unterschiede in der Durchführung und Auswertung (von individuellen Hilfen bis zu bewussten Fälschungen). Was bedeutet das für die Aussagekraft der Ergebnisse - und was sagt es über die Funktionalität des Kontrollkontexts aus, in dem die Tests eingesetzt werden, um Personen und Institutionen zu beurteilen?

Auch diese Diskussion über die Funktion von Tests und den Status ihrer Ergebnisse schließe ich mit einigen Thesen:

- Menschliches Verhalten ist immer mehrdeutig, und seine Deutung theoriegebunden und damit personenabhängig - auch bei Kreuzen in einem Test (vgl. etwa die Aufgabenanalysen zu VERA von Bartnitzky 2008).

Dasselbe (äußere) Verhalten kann Ausdruck ganz unterschiedlicher Intentionen, Kognitionen, Gefühle sein. Umgekehrt können sich dieselbe Absicht, dasselbe Wissen und Können, dieselben Gefühle unterschiedlich äußern - beispielsweise, weil die Personen aus verschiedenen Kulturen mit jeweils anderen Konventionen kommen. Auch die Kreuze in den neuerdings wieder so beliebten Multiple-Choice-Tests sind mehrdeutig. Deutsche SchülerInnen bearbeiten standardisierte Tests anders als

SchülerInnen aus den USA oder England. Zumindest galt das noch bis vor kurzem, und möglicherweise sind die gefeierten Punktzuwächse allein aus wachsender Vertrautheit mit Testformaten zu erklären, die jetzt auch als Aufgabenmuster im Unterricht geübt werden. Aber auch zeitlich kann es kulturelle Differenzen geben, wie die Beziehung zwischen Fernsehkonsum und Leseleistung zeigt: Ist sie zu Beginn des Medienwandels positiv (weil sich nur Familien aus den oberen sozialen Schichten das Gerät leisten können), so wird sie später häufig negativ (weil Fernsehen in manchen Familien aus den unteren Sozialschichten soziale Lücken füllt, statt die Nutzung des reicheren Informationsangebots zu fördern). Daten und ihre Beziehungen sind also immer deutungsbedürftig. Die Qualität dieser Deutungen ist abhängig von der Sensibilität und Klugheit der BeobachterInnen und Interpreten.

- Empirische Daten sprechen nicht für sich - sie sind nur kontextbezogen interpretierbar.

Im Herbst 2007, nach der Veröffentlichung der dritten PISA-Studie, gab es eine öffentliche Kontroverse zwischen Andreas Schleicher (OECD/Paris = internationales PISA-Konsortium) und Manfred Prenzel (IPN/Kiel = nationales, deutsches PISA-Konsortium) über die Frage, ob sich die deutschen SchülerInnen im naturwissenschaftlichen Bereich verbessert hätten oder nicht. Spätestens dieser Disput sollte selbst sozialwissenschaftlichen Laien deutlich gemacht haben: Zahlen sind interpretationsbedürftig. Und noch davor - für den Leser der Studie meist nicht einmal mehr sichtbar: Wie Fragen/Aufgaben formuliert, wie bestimmte Antworten/Lösungen gewertet, welche statistischen Verfahren angewandt, wie Kategorien gebildet, wie die gewonnenen Stichproben und Kennwerte gedeutet werden - eine Fülle von Annahmen und Einschätzungen gehen bereits in die Erzeugung und dann in die Interpretation von sog. „harten“ Daten ein. Und selbst wenn verschiedene ForscherInnen zu denselben Ergebnissen kommen, bedeutet das möglicherweise nur, dass eine Subjektivität durch ihre Standardisierung als Verfahren über andere privilegiert wurde.

- Fokussierung und Kontext, d.h. Präzision und Bedeutung sind konkurrierende Ansprüche an Sozialforschung, die sich nicht gleichzeitig, sprich: durch ein Instrument erfüllen lassen.

Wer ein Gemälde in seine Elemente zerlegt und diese nach Farbe und Form sortiert (vgl. die Beispiele „aufgeräumter Kunst“ von Ursus Wehrli 2002 und 2004), wird vielleicht Einsichten in deren Struktur und die *Maltechnik* gewinnen, aber nicht den *künstlerischen Wert* des Werks beurteilen können. Auch Schülerleistungen und die Qualität von Schule kann man als „Variablenbündel“ nicht verstehen. Evaluation steht hier - wie die Humanwissenschaften generell - vor einer Art „Zoom-Dilemma“: Entweder fokussiert man wenige Variablen, die sich bei hinreichender Kontrolle präzise erfassen lassen, oder man weitert den Blick auf deren Kontext und gewinnt damit ein Bild von ihrer Bedeutung für die Beteiligten - zahlt dann aber den Preis einer jeden Kommunikation zwischen Menschen: Subjektivität.

Meine kurze Skizze soll verdeutlichen: Auch Evaluation bedarf der Evaluation -

- nach der Funktion von Instrumenten

- nach dem Status ihrer Ergebnisse
- nach ihrer Wirkung in unterschiedlichen Kontexten.

## Evaluation ist mehr als Messen

Standards und Tests sind zwar nützliche Instrumente, aber es droht eine Verarmung der Evaluationskultur, wenn diese sich auf die Messung quantitativer Indikatoren konzentriert (wie in der Produktkontrolle und bei der Warentestung). Notwendig ist eine Erweiterung des Repertoires:

- Nutzung der Person als methodisches Instrument (analog zur „Kunstkritik“); und wegen der Risiken subjektiver Wahrnehmungen und Urteile:
- Entwicklung von Formen sozialer Kontrolle statt technischer (Schein-)Präzision, die - wie im Gerichtsverfahren - Mehrperspektivität und Transparenz durch entsprechende Verfahrensregeln sichern.

Notwendig ist ein nach Stufen differenziertes System der Rechenschaftspflichten und Evaluationsformen (vgl. den konkreten Vorschlag in Bartnitzky u.a. 1999 und die Übersicht über alternative Modelle in den OECD-Ländern bei Brügelmann 1980). Insofern ist es wichtig, schon die zentralen System-Monitorings (über punktuelle Leistungstests hinaus) zu ergänzen durch andere Indikatoren für die Qualität des Bildungswesens, wie es „Bildung auf einen Blick“ der OECD (2007) und innerhalb von Deutschland das Konsortium Bildungsberichterstattung (2006) versuchen. Das reicht aber nicht aus.

Schon für die Entwicklung einzelner Schulen ist eine umfassendere und kontextsensible Evaluation notwendig. Und die individuelle Leistungsbewertung ist zu einer „dialogischen Lernbeobachtung“ anhand förderorientierter Aufgaben zu entwickeln, die einen Blick unter die Oberfläche des Messbaren erlauben. PISA&Co müssen deshalb inhaltlich und methodisch durch komplementäre Formen der Evaluation ergänzt werden. Die Evaluation vor Ort ist als eigenständige Aufgaben mit besonderen Anforderungen und Möglichkeiten zu fördern. Evaluation des Unterrichts und des Lernens von SchülerInnen kann nicht an externe Methodensachverständige abgetreten werden, sondern muss als Kompetenz vor Ort kultiviert werden - unterstützt, nicht ersetzt durch ExpertInnen von außerhalb. Statt immer komplexere Instrumente zu entwickeln, brauchen wir robuste, unaufwändige, also alltagstaugliche Verfahren wie die Screening-Tests zum Lesen von Metze (vgl. [www.lesetest1-4.de/html/listo\\_aufgaben.html](http://www.lesetest1-4.de/html/listo_aufgaben.html) ).

Fazit: Lehrer müssen f/Forscher werden!

Im Folgenden (weitgehend übernommen aus Brügelmann 2008). stelle ich in aller Kürze zwei Beispiele zu diesem Ansatz vor:

- den reformpädagogischen Schulverbund „Blick über den Zaun“ als Beispiel für eine kooperative Fremdevaluation auf Schulebene (vgl. Harder 2004; Groeben u.a. (2005) und [www.blickueberdenzaun.de](http://www.blickueberdenzaun.de) ) sowie
- das Konzept „Pädagogische Leistungskultur“ des Grundschulverbands als Beispiel für eine dialogische Form der Lernbeobachtung und Leistungsbeurteilung (vgl. Bartnitzky u.a. 2005; 2006; 2007).

### **Schulevaluation durch „kritische Freunde“**

Interne Evaluation hat den Vorteil intimer Situationskenntnis und fehlender Bedrohung. Andererseits: Der Fremdblick von außen ist wichtig, um scheinbare Selbstverständlichkeiten in Frage zu stellen, und Distanz ist nötig, um sich den erkannten Schwächen zu stellen. Diese Einsichten stehen hinter den Lernstandserhebungen und den Schulinspektionen. Sie machen externe Evaluationen stark. Aber die Frage ist, in welcher Rolle die Externen kommen: als ExpertInnen, gar als Autoritäten, mit der Macht, Sanktionen zu verhängen oder zu bewirken - oder als PartnerInnen?

Externe Evaluation ist notwendig. Aber sie wird produktiver, wenn sie von Kontrolle abgekoppelt wird. Die Schulaufsicht sollte deshalb durch ein kollegiales Peer-Review ergänzt werden, das den Unterrichtsbetrieb nicht rechtlich kontrolliert, sondern fachliche Rückmeldung gibt. Die Schule sollten den „Fremdblick“ durch eine interne Bestandsaufnahme vorbereiten: Was sind unsere Ziele, wo liegen unsere Stärken, welche Probleme haben wir? Und sie sollte die Berichte der externen BeobachterInnen öffentlich kommentieren: Welche Einschätzungen teilen wir, was wollen wir unternehmen, um unsere Arbeit weiter zu verbessern? Ohne Sanktionsbefugnisse kann die Inspektion nur durch ihre Kompetenz und durch den sozialen Druck, den die Berichterstattung bedeutet, wirken.

Ein gelungenes Beispiel für das Austarieren von externer Sicht und notwendiger Vertrautheit ist die Initiative „Blick über den Zaun“. Dieser Verbund von inzwischen über 80 reformpädagogisch orientierten Schulen besteht seit 1989. In ihm schließen sich jeweils acht bis zehn Schulen (bewusst aus verschiedenen Schulformen) zu einem Arbeitskreis zusammen. Gemeinsam sind den Schulen die Standards des „Blick über den Zaun“, die sich vor allem auf die Qualität der pädagogischen *Prozesse* beziehen (vgl. von der Groeben u.a. 2005). Zweimal im Jahr wird eine Schule von jeweils zwei VertreterInnen der anderen Schulen des gleichen Arbeitskreises besucht, die in der Regel zwei Tage in der Schule mitleben (vgl. zum Ablauf im Einzelnen die Handreichung von Backhaus u.a. 2009). Die gastgebende Schule kann einen Beobachtungsauftrag formulieren, aber die Gäste sind frei, das zu beobachten, was ihnen wichtig erscheint. Sie nehmen am Unterricht teil, sie unterhalten sich mit KollegInnen, mit SchülerInnen und VertreterInnen der

Eltern. In einer Schlussrunde spiegeln die BesucherInnen einzeln, d.h. aus ihrer je individuellen und damit unterschiedlichen Perspektive ihre Eindrücke dem Kollegium zurück.

Es sind acht Prinzipien, an denen sich die Organisation dieser Form kollegialer Fremdevaluation orientiert:

- **Freiwilligkeit:** jede Schule entscheidet selbst, *ob* sie am Verfahren des Peer-Review teilnimmt.
- **Grundkonsens:** Mit den „Standards“ und dem „Leitbild“ ist ein pädagogischer Anspruch bestimmt, an dessen Umsetzung zu arbeiten sich alle Schulen verpflichten. Zudem geben sie einen gemeinsamen Rahmen für die Evaluation der Arbeit in den Schulbesuchen. Die Standards gelten als Richtlinien, die aber von keiner Schule immer ganz erfüllt werden können.
- **Autonomie:** Jeder Arbeitskreis verständigt sich darauf, *wie* er im Rahmen des verabschiedeten gemeinsamen Programms seine Arbeit gestalten und organisieren will.
- **Gleichrangigkeit:** Bei den Besuchen begegnen KollegInnen einander *auf Augenhöhe* und in Kenntnis dessen, dass alle vor denselben Problemen stehen und den Alltag aus eigener Erfahrung kennen.
- **Vielfalt:** In jedem Arbeitskreis begegnen sich Schulen unterschiedlicher Schulformen, pädagogischer Traditionen und Regionen - dies eröffnet verschiedene „fremde Blicke“ und verringert den Konkurrenzdruck.
- **Kontinuität:** Über die mehrjährige Zusammenarbeit wachsen Vertrauen und Verstehen und damit die *Glaubwürdigkeit von* und die *Offenheit für* Kritik.
- **Sensibilität:** Gesprächsführung und Feedback müssen von dem Ziel bestimmt sein, die besuchte Schule zu verstehen und sie so zu beraten, dass sie ihre *eigenen* Ziele besser verwirklichen kann.
- **Nachhaltigkeit:** Wenn es gelingt, Ergebnisse der Besuche über geeignete Arbeitsformen in den Alltag der Schulen zu transportieren, profitieren beide Seiten - GastgeberInnen wie BesucherInnen.

Die Erträge dieses kollegialen Austauschs sind vielfältig (vgl. Seydel 2007, 5-7):

„Auf die Zaungäste haben diese Besuche in der Regel drei wichtige Wirkungen:

1. Die oft geradezu verwirrende Konfrontation mit einer anderen, z. T. sehr fremden, Schulkultur klärt den Blick auf die eigene Schule.
2. Die Übernahme von neu in der besuchten Schule Gesehenem geschieht selten direkt, sondern zeitverzögert und mit einer Reihe von Transformationen, manchmal mit einem ‚sleeper effect‘, wenn erst im Nachhinein klar wird: ‚Das hatte ich ja dort und dort gesehen.‘ Nach der Rückkehr des Grenzgängers ist die Neugier der daheimgebliebenen Kollegen auf das, was er gesehen hat, nur von kurzer Dauer. Mit einiger zeitlicher Verzögerung kommt dann aber bei passender Gelegenheit die Frage: ‚Du warst doch in der Bodenseeschule - wie haben die denn die Organisationsprobleme des Epochenunterrichts gelöst?‘ etc. Der Bericht über eine andere Schule bekommt eine ganz andere Qualität, wenn es im Kollegium jemanden gibt, der ihre Schwelle überschritten hat. Er hat nicht nur über deren Zaun geblickt, sondern mit dem Nachbarn selbst gesprochen.
3. Mindestens genauso wichtig - wenn nicht sogar wichtiger - im Vergleich zum ‚sachlichen‘ Transfereffekt ist der motivationale Aspekt der Ermutigung und Rückenwärme: ‚Meine Schule ist im Vergleich zu der anderen gar nicht so schlecht.‘ ‚Meine Arbeit wird von den anderen wahrgenommen



und wertgeschätzt.' ,Andere haben auch ungelöste pädagogische Probleme und sind trotzdem nicht verzagt.' ...

Das Gastgeschenk, das die Besucher als Dank zurücklassen, ist von ganz besonderer Art.

Das Kollegium bekommt am Ende des Besuches einen ungewöhnlichen Spiegel über das Gesamtbild der Schule, über ihre Stärken, Schwächen und Entwicklungspotentiale. Die unterschiedliche Herkunft der Besucher - die Differenz z.B. zwischen einem antiautoritär geprägten Glockseelehrer und einem durch gemeinsame Formen und Werte geleiteten Montessorilehrer - ergibt differenzierte Blickwinkel und Einfärbungen der zurückgemeldeten Bilder. Die Fragen, die diese ‚kritischen Freunde‘ stellen, die Beobachtungen, die sie mitteilen, die Anregungen, die sie vorsichtig formulieren, tragen die Chance einer ganz anderen Wirkung in sich als der Besuch des Schulrates oder gar Inspektors. Sie sind Angebote auf Augenhöhe. Und weil es - auf Grund der Unterschiedlichkeit der Herkünfte der Besucher - nie ein ‚konsistentes‘ Bild ergibt, bleibt die Deutungshoheit bei der besuchten Schule. Die Differenz der Bilder fordert heraus. Oft waren noch Jahre nach einem Besuch die Provokationen der Zaungäste im Schulentwicklungsprozess präsent.“

Als Rahmen für die Gespräche mit einzelnen KollegInnen können Fragen wie die folgenden hilfreich sein, wenn man den Unterricht in einer Klasse in den Blick nimmt:

- Was sind Ihre wichtigsten Ziele und Prinzipien?

Nachfrage: „Wie stehen Sie zu folgenden Vorgaben der Richtlinien/Lehrpläne, zu folgenden Positionen der schulpädagogischen und (fach-)didaktischen Diskussion?“

- Wo steht Ihre Lerngruppe, wo stehen einzelne Kinder in den zentralen Entwicklungsdimensionen?

Nachfrage: „Ist Ihnen aufgefallen, dass Marc...?“

- Wo sehen Sie die Stärken und die Schwächen Ihrer Arbeit, also Ihrer Versuche, die eigenen Ansprüche und die vorgegebenen Anforderungen umzusetzen?

Nachfrage: „Mir ist bei der Beobachtung Ihres Unterrichts aufgefallen, dass...“

- Welche Umstände erschweren es Ihnen, Ihre Ansprüche im Alltag umzusetzen?

Nachfrage: „Könnte es auch daran liegen, dass ...?“

- Was haben Sie sich als nächste Schritte zur Entwicklung Ihrer Arbeit vorgenommen?

Nachfrage: „Haben Sie auch an folgende Möglichkeiten gedacht: ....?“

- Welche Unterstützung/welche Rahmenbedingungen wären nötig bzw. hilfreich?

Nachfrage: „Würde es Ihnen helfen, wenn...?“

Beratung statt Kontrolle bedeutet Konfrontation mit einer anderen Sicht, ohne dass diese sich als Norm versteht. Die Grundidee: Die Diskussion über die *Kriterien* für „guten Unterricht“ ist zu trennen von der Frage nach der Qualität seiner *tatsächlichen Umsetzung*. Es macht wenig Sinn, den tatsächlichen Unterricht mit Ansprüchen zu

bewerten, die die Bewerteten gar nicht teilen. Diese normativen Fragen sind vorweg zu klären - beispielsweise in einer Diskussion über das Schulprogramm, seine Stärken und Schwächen. Dabei kann zugleich Verständigung über die Kriterien erzielt werden, mit deren Hilfe der beobachtete Unterricht sinnvoll zu beurteilen ist. Testleistungen der SchülerInnen können eine Informationsquelle sein, um Schwächen auf die Spur zu kommen. Aber ertragreicher ist das in langjähriger Erfahrung fundierte Urteil der KollegInnen.

Auch dieses Verfahren hat seine Schwierigkeiten und das Instrumentarium ist entwicklungsfähig. Entscheidend ist der Ansatz: Evaluation „von außen“, aber nicht „von oben“, seien es die wissenschaftlichen ExpertInnen der Lernstandserhebungen, seien es die VertreterInnen der Verwaltung bei der Schulinspektion. Begegnung auf Augenhöhe ist die Grundlage für Offenheit und damit für die Bereitschaft, sich den eigenen Schwächen zu stellen und ernsthaft an ihnen zu arbeiten.

Diese Anforderungen gelten auch für die Leistungsbewertung auf der Schülerebene.

### **„Dialogische Lernbeobachtung“ in einer pädagogischen Leistungskultur**

Der Grundschulverband hat bereits vor der Veröffentlichung von PISA ein Konzept für ein umfassendes Evaluationssystem vorgeschlagen. In ihm werden nicht nur die Rechenschaftspflichten von Politik und Verwaltung, sondern auch die Evaluationsaufgaben der einzelnen Schule, der Lehrpersonen und der SchülerInnen konkret entfaltet. Besondere Bedeutung wird der begleitenden Lernbeobachtung beigemessen. Statt nur abzuprüfen, ob Ziele erreicht sind, sollen Lernfortschritte und Schwierigkeiten im Lernprozess erfasst und diagnostisch gedeutet werden.

Dass die üblichen Klassenarbeiten und ihre Benotung diese Funktion nicht erfüllen können, ist seit vielen Jahrzehnten bekannt; standardisierte Tests, die als Alternative angeboten werden, haben wiederum ihre eigenen Probleme (vgl. die aktuelle Zusammenfassung in: Arbeitsgruppe Primarstufe 2006). Der Grundschulverband hat deshalb sein Konzept „Pädagogische Leistungskultur“ entwickelt und darin folgende Kriterien für Aufgaben zur Lernbeobachtung formuliert. Sie sollten...

- der Lehrperson Informationen erbringen
  - nicht nur über aktuelle Einzelleistungen,
  - sondern auch über die Strategien („Tiefenstrukturen“)
  - und über deren Entwicklung („Lerngeschichte“);
- für die SchülerInnen auch inhaltlich eine produktive Lernsituation darstellen; vor allem aber
- dialogisch angelegt sein als wechselseitige Verständigung über Lernziele, Bewertungskriterien und tatsächliche Leistungen und damit

- die Fähigkeit der Kinder zur Selbsteinschätzung eigener Arbeiten entwickeln. Das diagnostische Repertoire von LehrerInnen kann durch heuristisch eingesetzte Tests, durch Beobachtungsbögen und verschiedene Dokumentationsformen (wie Portfolios) erweitert und differenziert werden (vgl. die vielfältigen Vorschläge für die verschiedenen Lernbereiche und Jahrgangsstufen in Bartnitzky u. a. 2005-2007). Das Konzept einer „pädagogischen Leistungskultur“ fordert daneben aber verschiedene „Institutionen“ im Unterrichtsalltag, die den SchülerInnen helfen, ihre eigene Arbeit kritisch-konstruktiv zu überprüfen und an den Arbeiten anderer ihre Maßstäbe zu schärfen. Nur drei Beispiele aus dem Grundschulbereich:
  - Schreibkonferenzen, in denen nach bestimmten Regeln Entwürfe vorgestellt, kommentiert und dann mit Hilfe Anderer überarbeitet werden (vgl. Spitta 1998 und ergänzende Hilfen für den Sprachunterricht bei Brinkmann/Brügelmann 2005);
  - Rechendiskussionen in der Klasse, z.B. zum „harten Brocken des Tages“, so dass die Kinder Schwierigkeiten, Hypothesen und Lösungsstrategien austauschen und damit voneinander lernen können (eingeführt hat diese Idee Erichson 2004 im Rechtschreibunterricht, vgl. für den Mathematikunterricht Küppers 2005; Sundermann/Selter 2005);
  - im Sachunterricht Metagespräche über Stärken und Schwächen von Präsentationen vor der Klasse, über Arbeitsergebnisse von Gruppen oder Einzelnen bis hin zu deren Bewertung durch das Plenum nach vereinbarten Kriterien (vgl. die Beispiele für den Sachunterricht bei Schönknecht/Klenk 2005, S.22ff).

Integriert in Aufgaben für alle Bereiche des Anfangsunterrichts im Lesen und Schreiben hat Erika Brinkmann (2008) Hilfen zur Selbst- und Fremdeinschätzung für die Kinder in den Materialien für ihre „ABC-Lernlandschaft“ vorgelegt. Sie sind unmittelbar auf die Bildungsstandards der KMK bezogen, entgehen aber den oben genannten Fällen standardisierter Tests.

Der Lesedidaktiker Schmalohr (1997, S.42f) hat in seiner Arbeit mit jugendlichen und erwachsenen Analphabeten drei einfache Fragen genutzt, um sie zum Nachdenken über ihre Probleme zu bringen und mit ihnen in ein Gespräch über sinnvolle Lernwege zu kommen:

1. „Wie lese ich, wo habe ich Schwierigkeiten?“
2. „Woran könnte das liegen?“
3. „Was kann ich tun?“

Hier zeigt sich derselbe Geist wie in der Peer-Evaluation auf Schulebene: Beratung statt Kontrolle. Ein solches Verständnis von Evaluation respektiert und nutzt die Kompetenz der Betroffenen, ihre Probleme selbst zu erkennen, Ursachen für diese zu finden und Ideen für ihre Überwindung zu entwickeln - ohne sie allein zu lassen. Evaluation wird zur Dienstleistung statt zur Kontrollinstanz, sei es durch ExpertInnen oder Verwaltung (vgl.

mein Plädoyer bereits in der ersten Runde der „empirischen Wende“ der Bildungsforschung Brügelmann 1976).

## Literatur

Arbeitsgruppe Primarstufe (2006): Sind Noten nützlich und nötig? Zifferzensuren und ihre Alternativen im empirischen Vergleich. Eine wissenschaftliche Expertise des Grundschulverbandes, erstellt von der Arbeitsgruppe Primarstufe an der Universität Siegen. Frankfurt: Grundschulverband e.V.

Backhaus (Red.) (2009): Schulen lernen von Schulen. Beispiele und Portraits aus dem Schulverbund 'Blick über den Zaun'. Reformpädagogische Arbeitsstelle 'Blick über den Zaun'. Siegen: Universität.

Backhaus, A., u.a. (2009): "Blick über den Zaun": Schulen lernen von Schulen. Vorschläge zur Planung und organisatorischen Ausgestaltung von Peer-Reviews durch kritische Freunde. Reformpädagogische Arbeitsstelle 'Blick über den Zaun'. Siegen: Universität.

Bartnitzky, H. (2008): VERA Deutsch 2008: Wie Kindern durch Vergleichsarbeiten Unrecht geschieht. Grundschule aktuell, 103, 4-8.

Bartnitzky, H., u.a. (1999): Zur Qualität der Leistung - 5 Thesen zu Evaluation und Rechenschaft der Grundschularbeit. Frankfurt: Grundschulverband - Arbeitskreis Grundschule e.V. Auch in Schmitt (1999, 165-196).

Brinkmann, E. & Brügelmann, H. (2005): Pädagogische Leistungskultur - Materialien für Klasse 1 und 2: Deutsch. In Bartnitzky, H., u.a. (Hrsg.) (2005&2006&2007). Pädagogische Leistungskultur: Materialien für Klasse 1/2 und Klasse 3/4. Beiträge zur Reform der Grundschule, Bd. 119 & 121 & 123. Frankfurt: Grundschulverband, Heft 3.

Brinkmann, E., u.a. (2008): ABC-Lernlandschaft. Seelze: Lernbuch-Verlag/Friedrich.

Brügelmann, H. (1976): Curriculumevaluation - eine Dienstleistung für die Unterrichtspraxis. In Seybold, H. (Hrsg). Innovation im Unterricht. Curriculumentwicklung und handlungsorientierte Forschung. Ravensburg: Otto Maier, 75-93. Download unter: [www.agprim.uni-siegen.de/printbrue/curriculumevaluation.pdf](http://www.agprim.uni-siegen.de/printbrue/curriculumevaluation.pdf)

Brügelmann, H. (1980): Experimental decision-making and responsive accountability. Expert report for the 'Basic Education Policies Project'. Paris: OECD/CERI. Download unter: [www.agprim.uni-siegen.de/printbrue/brue.oecd.evaluation.pdf](http://www.agprim.uni-siegen.de/printbrue/brue.oecd.evaluation.pdf)

Brügelmann, H. (2008a): Fieber genau zu messen ist noch keine Diagnose, Fieber erfolgreich zu senken keine Therapie. Wie Leistungstests in ihren Leistungsmöglichkeiten durch PISA & Co überfordert werden. Beitrag zum Forum „Schule ist mehr als PISA - Zur Bedeutung reformpädagogischer Ansprüche an die schulische Bildung von heute“ der ZEIT-Stiftung in Hamburg am 6./7. März 2008. Download unter: [www.agprim.uni-siegen.de/printbrue/brue.08a.pisa\\_refpaed.pdf](http://www.agprim.uni-siegen.de/printbrue/brue.08a.pisa_refpaed.pdf)

Erichson, C. (2004): Der harte Brocken des Tages. Grundschule Deutsch, 1. Jg., H.2, 14-17.

Groeben, A.v.d., u.a. (2005): Unsere Standards. Ein Diskussionsentwurf, vorgelegt von „Blick über den Zaun“ - Bündnis reformpädagogisch engagierter Schulen. Neue Sammlung, 55. Jg., H. 2, 253-297.

Harder, W. (2004): Von anderen Schulen lernen. Blick über den Zaun auf pädagogischen Erkundungsreisen. Pädagogik, 56. Jg., H. 1, 45-48. Nachdruck in Backhaus (2009).

Heymann, H. W. (1996): *Allgemeinbildung und Mathematik. Bildungstheoretische Reflexionen zum Mathematikunterricht an allgemeinbildenden Schulen.* Weinheim: Beltz.

Küppers, H. (2005): *Mathematik.* In Bartnitzky, H., u.a. (Hrsg.) (2005&2006&2007). *Pädagogische Leistungskultur: Materialien für Klasse 1/2 und Klasse 3/4. Beiträge zur Reform der Grundschule, Bd. 119 & 121 & 123.* Frankfurt: Grundschulverband, Heft 4.

Schmalohr, E. (1997): *Das Erlebnis des Lesens. Grundlagen einer erzählenden Lesepsychologie.* Stuttgart: Klett-Cotta.

Schmitt, R. (Hrsg.) (1999): *An der Schwelle zum dritten Jahrtausend. BundesGrundschulKongress 1999.* Frankfurt: Grundschulverband - Arbeitskreis Grundschule.

Schönknecht, G. & Klenk, G. (2005): *Sachunterricht.* In Bartnitzky, H., u.a. (Hrsg.) (2005&2006&2007). *Pädagogische Leistungskultur: Materialien für Klasse 1/2 und Klasse 3/4. Beiträge zur Reform der Grundschule, Bd. 119 & 121 & 123.* Frankfurt: Grundschulverband, Heft 5.

Seydel, O. (2007): „Blick über den Zaun“. *Wie Schulen voneinander lernen können.* In Schulverbund „Blick über den Zaun“ (Hrsg.). *Beobachten, bewerten, beraten. Verfahren und Werkzeuge für eine andere Evaluation.* Reformpädagogische Arbeitsstelle 'Blick über den Zaun'. Siegen: Universität.

Spitta, G. (Hrsg.) (1998): *Freies Schreiben - eigene Wege gehen.* CH-Lengwil: Libelle.

Sundermann, B. & Selter, C. (2005): *Mathematikleistungen feststellen, beurteilen und fördern. Beschreibung des Moduls 9 für das Projekt SINUS-Transfer Grundschule.* Download unter: [www.sinus-grundschule.de/](http://www.sinus-grundschule.de/)

Wehrli, U. (2002): *Kunst aufräumen.* Zürich: Kein & Aber Verlag.

Wehrli, U. (2004): *Noch mehr Kunst aufräumen.* Zürich: Kein & Aber Verlag.