

# Nicht PISA, sondern die Fixierung auf PISA und der Umgang mit den Daten ist das Problem<sup>1</sup>

Die Autorität von PISA als Evaluationsinstanz im Schulwesen wird zunehmend in Zweifel gezogen. Die grundsätzliche Kritik von außen am forschungsmethodischem Design und den statistischen Auswertungsmethoden hat sich deutlich verschärft.<sup>2</sup> Gleichzeitig machen interne Differenzen zwischen dem internationalen PISA-Team in Paris und der deutschen Leitung deutlich, wie auslegungsfähig und interpretationsbedürftig die scheinbar objektiven Daten tatsächlich sind. So stellt sich die Frage, was es für Aussagen über »die« Naturwissenschaften bedeutet, dass die Veränderung eines Teils des Tests von 2003 auf 2006 so große Veränderungen in den Ranglisten bewirkt. PISA ist eben nur ein Blick auf Schule – und offensichtlich kann sich selbst dieser stark verändern, wenn man statt der Kieler Brille die OECD-Gläser aus Paris aufsetzt, von anderen Sehhilfen ganz zu schweigen ...

Und es stellen sich weitere Fragen: Die PISA-Stagnation beim Lesen führt MANFRED PRENZEL darauf zurück, dass in diesem Bereich Veränderungen nicht so schnell möglich seien wie in Mathematik und den Naturwissenschaften. Aber wie konnten sich dann die Leseleistungen in der Grundschule bei IGLU von 2001 bis 2006 deutlich stärker verbessern als in der Sekundarstufe?

Andererseits sieht PRENZEL die Ursache für Verbesserungen der PISA-Werte in den Maßnahmen, die als Folge des PISA-Schocks eingeleitet wurden. Nur wie erklärt er dann, dass sich die Leseleistungen der Grundschulen von 1991 bis 2001, also in den Jahren vor PISA, im Vergleich zu 2001 bis 2006 überproportional stark verbessert haben? Das heißt zumindest für die Grundschule: Mit Einführung von Bildungsstandards, Kompetenztests und all den anderen output-orientierten Kontrollmaßnahmen hat sich das Tempo der Verbesserung selbst in dem kleinen Ausschnitt der Testwerte halbiert – von negativen Nebenwirkungen

ganz abgesehen: Fixierung auch der individuellen Diagnose auf punktuelle Testwerte statt begleitender Lernbeobachtung; zunehmende Orientierung des Unterrichts in den Hauptfächern an den Testinhalten und zunehmendes Training von Testformaten; Vernachlässigung des musisch-ästhetischen Bereichs. Ähnliche Entwicklungen ließen sich schon vorher in testgläubigen Systemen wie England und den USA beobachten, die von 2001 bis 2006 an Punkten verloren haben. Und könnte es nicht sein, dass Schweden auch deshalb »schwächelt« (PRENZEL), weil dort die Testkontrolle ebenfalls in den letzten Jahren verschärft worden ist?

## PISA als Projekt, Programm und Paradigma

Man kann das Potenzial und die Schwächen von PISA aus verschiedenen Blickwinkeln diskutieren.

Als konkretes Projekt untersucht PISA alle paar Jahre die Fachleistungen am Ende der Pflichtschulzeit. Analog zu TIMSS, IGLU und anderen internationalen Leistungsstudien hat PISA damit den begrenzten Anspruch, einen kleinen Ausschnitt der Wirkungen von Schule über verschiedene Länder hinweg zu vergleichen. Mit diesem Fokus hat PISA sich als hilfreich erwiesen, um Problemen öffentliche Aufmerksamkeit zu verschaffen, die zwar in Fachkreisen und im Unterrichtsalltag schon lange bekannt sind, bildungspolitisch aber verdrängt worden sind.

## Allerdings: PISA allein

- sagt uns nicht, wie gut oder schlecht unser Schulsystem ist – dazu wird ein zu kleiner Ausschnitt der Schulqualität mit zu begrenzten Mitteln erfasst;
- kann nicht erklären, wo die Ursachen für Schwächen liegen, denn Korrelationen, vor allem in einem Querschnitt, erlauben keine Kausalschlüsse;
- bietet keine Anweisungen für deren Überwindung, denn es ist nicht als experimenteller Vergleich von alterna-

tiven Handlungsmöglichkeiten angelegt.

Solche regelmäßigen Bestandsaufnahmen sind dennoch sinnvoll, wenn man sie als einen Baustein in einem umfassender konzipierten System-Monitoring sieht<sup>3</sup> – und wenn man die Grenzen von Design und Instrumenten nicht nur kennt, sondern in der Bewertung der Befunde und in den anschließenden Entscheidungen berücksichtigt.<sup>4</sup>

In Form von VERA und anderen flächendeckenden Lernstandserhebungen haben die Bundesländer PISA&Co allerdings zu einem *Programm* des System-Monitoring gemacht. Dieses erhebt darüber hinaus den Anspruch, die Arbeit von Schulen und Lehrerinnen, ja sogar den Leistungsstand einzelner Schülerinnen bewerten zu können.

Inzwischen ist PISAnismus mit seiner Testmentalität für viele sogar zum generellen *Paradigma* der Evaluation von Unterricht und von individuellen Lernprozessen geworden. Damit überzieht dieser Ansatz seinen Kredit und es droht eine Verarmung der Evaluationskultur.

## Alternativen

Notwendig ist ein nach Stufen differenziertes System der Rechenschaftspflichten und Evaluationsformen.<sup>5</sup> Insofern ist es als erstes wichtig, zentrale System-Monitorings (über punktuelle Leistungstests) 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.

Für die Entwicklung einzelner Schulen ist eine umfassendere und kontextsensible Evaluation notwendig, und die individuelle Leistungsbewertung ist zu einer »dialogischen Lernbeobachtung«

VON HANS  
BRÜGELMANN

HANS BRÜGELMANN,  
Autor des Buches »Schule verstehen  
und gestalten«, ist Professor für  
Grundschulpädagogik und -didaktik  
an der Universität Siegen



anhand förderorientierter Aufgaben zu entwickeln, die einen Blick unter die Oberfläche des Messbaren erlauben (»Pädagogische Leistungskultur« im Sinne des Grundschulverbands). PISA&CO müssen deshalb inhaltlich und methodisch durch komplementäre Formen der Evaluation ergänzt werden. Damit wird es notwendig, die Evaluation vor Ort als eigenständige Aufgaben mit besonderen Anforderungen und Möglichkeiten zu fördern.

Im Folgenden stelle ich in aller Kürze zwei Beispiele vor:

- den reformpädagogischen Schulverband »Blick über den Zaun« als Beispiel für eine Evaluation auf Schulebene und
- das Konzept »Pädagogische Leistungskultur« des Grundschulverbands als Beispiel für eine dialogische Form der Lernbeobachtung und Leistungsbeurteilung.

## 1) 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 – 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 Rück-

meldung gibt. Die Schule sollte 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 60 reformpädagogischen Schulen besteht seit 1989. In ihm schließen sich jeweils 7–10 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.<sup>6</sup> Zweimal im Jahr wird eine Schule von jeweils zwei VertreterInnen der anderen Schulen des gleichen Arbeitskreises besucht, die zwei bis drei Tage in der Schule mitleben. 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 Vertretern der Eltern. In einer Schlussrunde spiegeln die BesucherInnen einzeln, d. h. aus ihrer je individuellen und damit unterschiedlichen Perspektive ihre Eindrücke dem Kollegium zurück. Die Erträge dieses kollegialen Austauschs kommen beiden Seiten zugute: den Gastgebern und den »Zaungästen« (vgl. SEYDEL 2007, 5–7).

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 Schülerschulebene.

## 2) »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.<sup>7</sup> Der Grundschulverband hat deshalb sein Konzept »Pädagogische Leistungskultur« entwickelt und darin folgende Kriterien für Aufga-

ben 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.<sup>8</sup> Das Konzept einer »pädagogischen Leistungskultur« fordert daneben aber verschiedene »Institutionen« im Unterrichtsallday, 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<sup>9</sup>;
- Rechendiskussionen in der Klasse, z. B. zum »harten Brocken des Tages«<sup>10</sup>,

so dass die Kinder Schwierigkeiten, Hypothesen und Lösungsstrategien austauschen und damit voneinander lernen können;

- 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.<sup>11</sup>

Der Lesedidaktiker SCHMALOHR (1997, 42 f.) 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.

### Fazit

Damit schließt sich der Kreis: Evaluation bedeutet Macht und Abhängigkeit – deshalb darf man sie nicht ExpertInnen überlassen, denn dieses Problem ist nicht technisch zu lösen. Aber Schule ist ein öffentlicher Raum und dies schließt ein, dass alle Beteiligten

rechenschaftspflichtig sind. Dafür brauchen sie je nach Aufgabe unterschiedliche Verfahren – und Unterstützung. Ein demokratisches Verständnis von Evaluation fordert, die Betroffenen nicht durch externe Beurteilungen zu entmündigen, sondern ihre persönliche Evaluations- und Problemlösekompetenz zu stärken.<sup>12</sup>

Dezentrale Evaluation kann bildungspolitisch orientierte Studien wie PISA nicht ersetzen. Beide Ansätze haben ihre spezifische Funktion in einem umfassenderen Rechenschaftssystem. Deshalb sind drei Punkte abschließend besonders wichtig:

- Der PISA-Stil muss auf Systemevaluation begrenzt werden und darf nicht zum Paradigma für Evaluation generell werden. Insbesondere sind der Sinn und die Form flächendeckender jährlicher Lernstandserhebungen zu überdenken.
- Die Scheinpräzision von Zahlen aus Erhebungen mit Leistungstests muss immer wieder bewusst gemacht und ihre mehrperspektivische Deutung schon bei der Veröffentlichung gesichert werden.
- Die Ressourcen für Evaluation dürfen nicht mehr so stark auf die zentrale Evaluation konzentriert werden. Wir brauchen eine massive politische, wissenschaftliche und finanzielle Unterstützung für die Entwicklung der Evaluationskompetenz vor Ort.

## Anmerkungen

- 1 Überarbeitete Fassung eines Beitrags zum GEW-Kolloquium »PISA – abschaffen oder weiterentwickeln« am 21. 11. 2007 in Berlin. Die Literaturnachweise finden sich in dem in Anm. 4 genannten Beitrag.
- 2 Vgl. die Beiträge zu HOPMANN u. a. (2007) und JAHNKE/MEYERHÖFER (2007)
- 3 Vgl. das schon breiter angelegte »Bildung auf einen Blick« der OECD (z. B. 2007).
- 4 Vgl. dazu ausführlicher: BRÜGELMANN, H. (2008): Fieber genau zu messen ist noch keine Diagnose, Fieber erfolgreich zu senken keine Therapie. Wie Leistungstests in ihren Leistungsmög-

- lichkeiten 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 [http://www.agprim.uni-siegen.de/printbrue/brue.o8a.pisa\\_refpaed.pdf](http://www.agprim.uni-siegen.de/printbrue/brue.o8a.pisa_refpaed.pdf) [27. 11. 2007]
- 5 Vgl. den konkreten Vorschlag in BARTNITZKY u. a. (1999) und die Übersicht über alternative Modelle in den OECD-Ländern bei BRÜGELMANN (1980).
- 6 Vgl. VON DER GROEBEN u. a. (2005).

- 7 Vgl. die aktuelle Zusammenfassung in: ARBEITSGRUPPE PRIMARSTUFE (2006).
- 8 Vgl. die vielfältigen Vorschläge für die verschiedenen Lernbereiche in: BARTNITZKY u. a. (2005–2007).
- 9 Vgl. SPITTA (1998) und ergänzende Hilfen für den Sprachunterricht bei BRINKMANN/BRÜGELMANN (2005).
- 10 Eingeführt von ERICHSON (2004) im Rechtschreibunterricht, vgl. für den Mathematikunterricht KÜPPERS (2005); SUNDERMANN/SELTER (2005).
- 11 Vgl. die Beispiele für den Sachunterricht bei SCHÖNKNECHT/KLENK (2005, 22 ff.)
- 12 Vgl. BRÜGELMANN (2007; 2008).