

Stähling, Reinhard

Qualitätsentwicklung statt Vergleichsarbeiten. Zu einem unfruchtbaren Verhältnis von Forschung und Schule

Die Deutsche Schule 97 (2005) 2, S. 211-221



Quellenangabe/ Reference:

Stähling, Reinhard: Qualitätsentwicklung statt Vergleichsarbeiten. Zu einem unfruchtbaren Verhältnis von Forschung und Schule - In: Die Deutsche Schule 97 (2005) 2, S. 211-221 - URN: urn:nbn:de:0111-pedocs-273742 - DOI: 10.25656/01:27374

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-273742>

<https://doi.org/10.25656/01:27374>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Reinhard Stähling

Qualitätsentwicklung statt Vergleichsarbeiten

Zu einem unfruchtbaren Verhältnis von Forschung und Schule

In Folge des deutschen „PISA-Desasters“ ist die Suche nach dem Verursacher des schlechten Abschneidens 15-jähriger Schüler in vollem Gange. Die Öffentlichkeit möchte eine nachvollziehbare Erklärung von der Politik. Versäumnisse im Schulsystem wurden merkwürdig lange zum Tabuthema erklärt. Dieses Tabu pflegen diejenigen, die auf die enormen Kosten eines Umbaus unseres Erziehungswesens verweisen und ihn von Beginn an als unrealistisch darstellen. Es erscheint folgerichtig, dass man „in Zeiten knapper Kassen“ zunächst einmal den kurzfristig kostengünstigeren, systemimmanenten Weg wählen will. Niemand hat etwas gegen eine Bestandsaufnahme einzuwenden. So wird die Forschung beauftragt, herauszufinden, wer in den Schulen für „schwache Schülerleistungen“ verantwortlich ist. Unter dem Diktat „knapper Kassen“ holt man eine alte „Wunderwaffe“ hervor, die neben einer Schnelldiagnose auch noch in der Lage ist, eine kostengünstige Selbstheilungsempfehlung für den kranken Patienten Schule anzubieten: die Vergleichsarbeiten. Der Lernstand der Schüler wird flächendeckend getestet und miteinander verglichen. Alles weitere ergibt sich daraus.

In vielen Gesprächen mit kompetenten Lehrerkolleginnen und -kollegen spüre ich eine ausgeprägte Ablehnung gegenüber Forschern, die trotz ihrer offenkundigen Unerfahrenheit mit „Schulalltag“ unbescheiden Ratschläge für die Qualitätsentwicklung von Unterricht erteilen. Wenn dann auch noch mit der machtvollen Unterstützung der Schulpolitik „wissenschaftlich fundierte“ Anforderungen als Standards für den Berufsalltag verkauft werden, fangen viele Kolleginnen und Kollegen an, passiven Widerstand zu leisten. Sie wehren sich auf vielfältige, humorvolle und kreative Art offen oder unsichtbar. Ältere „Schulmeister“ sprechen davon, dass sie „diese Modewelle“ auch noch überstehen werden. Das Vortäuschen abverlangter Tätigkeiten ist dafür eine beliebte Strategie, die jeder Praktiker kennt. Es ist zugleich Kennzeichen und Auswuchs einer fehlgeleiteten Beziehung zwischen Schulpraxis und pädagogischer Forschung. Eine gegenseitige Befruchtung findet in solch einem Kontext selten statt und die Visionen von „kooperativer Professionalisierung“ oder „forschenden Lehrern“ liegen in weiter Ferne.

Viele Pädagogen fragen sich: Auf Grundlage welcher Daten macht sich Forschung ein Bild über die Kompetenzen, Tätigkeiten und Motive eines „guten“ oder eines „schlechten“ Lehrers? Welche Daten werden erhoben, um festzustellen, wie äußere Faktoren das Auftreten „guten“ Unterrichts bei berufserfahrenen Lehrern begünstigen? Wird bei der Erhebung von Schuldaten die *Möglichkeit von Artefakten (auf Täuschung beruhende Sachverhalte)* ausreichend reflektiert? Im Folgenden soll beleuchtet werden, auf welcher Datenbasis das Projekt „Vergleichsarbeiten“ die „Wirklichkeit“ abzubilden und zu verändern versucht.

1. Hohe Erwartungen an Vergleichsarbeiten

Unter der Überschrift „Wichtige Studien der Bildungsforschung“ führt Andreas Helmke (2003) in seinem Buch „Unterrichtsqualität“ neben TIMSS (1995/99), PISA (1998-2007), PIRLS (2003) und IGLU (2003) u.a. das „Projekt Vergleichsarbeiten“ (2002) auf, das im Auftrag des Landes Rheinland-Pfalz entwickelt wurde. Diese Lernstandserhebungen unter dem Namen „VERA“ (VERgleichsArbeiten) sollen in Zukunft jährlich in mehreren Bundesländern flächendeckend durchgeführt werden. Auch wenn die auf Systemvergleich abzielenden Schulleistungsstudien wie PISA mit den eher auf Qualitätsentwicklung ausgerichteten Vergleichsarbeiten nicht zu vergleichen sind, führt Helmke diese Projekte der Bildungsforschung nebeneinander auf. Er nennt zwei Hauptziele der Vergleichsarbeiten:

- Schülerleistungsmessung und Bestandsaufnahme in den Jahrgängen 4 und 9
- Verbesserung des Lehren und Lernens.

Vergleichsarbeiten sind keine komplette externe Evaluation schulischer Arbeitsergebnisse, sondern sollen den *Ausgangspunkt von Selbstevaluation und Entwicklungsprozessen* durch die jeweilige Schule bilden. Grundlage des Datenvergleichs stellen die Schulen eines Landes dar, die unter ähnlichen Voraussetzungen arbeiten (erweiterte Bezugsgruppe). Mit Zentralstichproben wird die Durchführung überprüft und die Leistungsfähigkeit im Sinne eines Bildungsmonitorings erfasst. Ankeritems sorgen für Vergleichsmöglichkeiten im Zeitverlauf. In mindestens acht Bereichen erwartet Helmke einen Nutzen durch die Anwendung des Instruments „Vergleichsarbeit“. Zu untersuchen ist, ob und in welchem Maße die von Helmke genannten Ziele sowohl der Bestandsaufnahme als auch der Qualitätsentwicklung erreicht werden können. Wir folgen in der Darstellung den Hinweisen der Forschergruppe (vgl. www.uni-landau.de; Helmke 2003) und ergänzen sie um weitere kritische Aspekte (vgl. Brügelmann 2004 und 2005):

(1.) *Messung und „internationale Verortung“*: Die Daten sind teils vergleichbar, da Items von TIMS- und IGLU-Studien verwendet wurden. Allerdings ist die „Bedingungs- und Durchführungskontrolle geringer, da die Durchführung in der Regel von Lehrkräften und nicht von geschulten Testleitern vorgenommen wird“ (Helmke 2003). Leistungsmessungen in vielen sprachlichen Testbereichen (z.B. Textproduktion) sind nicht objektivierbar (vgl. Helmke 2003). Lehrer müssen selbst die Testergebnisse einschätzen. – Die Durchführung entspricht nicht den durch die begrenzte Konzentrationsfähigkeit bei Grundschulern notwendigen differenzierenden Rahmenbedingungen. Die Testzeiten für Mathematik (60 Minuten) und Deutsch (90 Minuten) reichen teilweise nicht aus und führen bei langsam lernenden Schülern zu Abbruch des Tests mit möglicherweise entmutigender Wirkung. Beide Tests finden innerhalb einer Woche statt. – Außerdem ist die Aufgabenfülle im Grundschulbereich unüblich und didaktisch fragwürdig (Textproduktion, Diktat u.a. innerhalb einer Testeinheit unter Zeitdruck). – Lehrer kontrollieren die Durchführung und könnten aus verschiedenen Gründen die Ergebnisse so manipulieren, dass sie den Erwartungen entsprechen. – Um regelmäßig die Leistungen des Schulsystems zu erfassen, wäre kein jährliches Testsetting erforderlich. Die Beschränkung auf repräsentative Stichproben im Zeitraum von vier bis sechs Jahren würde für diesen Zweck reichen (vgl. Brügelmann 2005).

(2.) „*Standortbestimmung*“: Die Daten sind vergleichbar mit dem Durchschnitt aller getesteten Schulen oder mit Referenzschulen, deren Einzugsgebiete ähnlich sind. Der Test ist jedoch nicht geeignet für Schüler mit sonderpädagogischem Förderbedarf im Bereich Lernen. Die Validität ist eingeschränkt, weil die Schule etwa die Hälfte der Items (Aufgaben) selbst auswählen kann. Somit ergeben sich unterschiedlich schwierige Tests (vgl. Helmke 2003). – Auch die Einschätzung des sozioökonomischen Rahmens einer Schule hat keine objektiven Daten zur Grundlage.

(3.) „*Innerschulischer Vergleich*“: Gesamtleistung und Streuung werden sichtbar, Ergebnisse lassen sich schulintern diskutieren. Ein Ranking „auf Klassenebene“ ist „ausdrücklich erwünscht“ (Helmke 2003).

(4.) „*Implementierung neuer Curricula und Bildungsstandards*“: Durch die Anwendung bundesweit diskutierter Bildungsstandards wird die Auseinandersetzung mit dieser Neuerung vorbereitet. – Wichtige Aspekte der Leistungserziehung, die in neue Richtlinien und Lehrpläne aufgenommen wurden, stehen allerdings im Widerspruch zu Testsettings. So sind unter den vorgeschriebenen VERA-Testbedingungen z.B. bei der Textproduktion weder Wörterbücher zugelassen, noch werden Brainstorming, Textplanung und -überarbeitung in Gesprächen ermöglicht, obwohl diese fachdidaktisch geforderten Methoden unbestritten sind. Verbesserte Verfahren der Erbringung von Leistung und Leistungsdokumentation wie Portfolios oder Lerntagebücher (vgl. Bartnitzky/Speck-Hamdan 2004) bleiben unberücksichtigt. Die umstrittene Einführung von Standards bekommt eine wissenschaftliche Schein-Legitimation (vgl. Schlömerkemper 2004).

(5.) „*Fachdidaktische Impulse*“: Die Häufung bestimmter Fehlermuster regen didaktische Diskussionen an und können zur „Verbesserung der Fehlerkultur“ (Helmke) führen. – Es ist jedoch zu bedenken, dass ein Scheitern von leistungsschwachen Schülern an den fachdidaktisch legitimierten Ansprüchen möglicherweise eher vertuscht wird und zur Verschleierung der Ergebnisse als zur offenen Diskussion führen kann (vgl. Brügelmann 2004).

(6.) „*Pädagogische Intervention*“, „*Verbesserung von Unterrichtsqualität*“: Durch die jährliche Wiederholung der VERA sind Vergleiche über die Zeit denkbar. Ergebnisse können Basis für Schülerförderung sein. Fortschritte lassen sich durch die empfohlene Wiederholung der Aufgaben zu einem späteren Zeitpunkt überprüfen.

(7.) „*Erfassung und Training diagnostischer Kompetenzen*“: Anhand der Aufgaben können Vorhersagen der Lehrer zu Leistungen einzelner Schüler und der gesamten Klasse überprüft werden. – Diagnostische Kompetenzen im Unterrichtsalltag sind jedoch vor allem durch systematische Beobachtung einzelner Schüler und Auswertungen im Team zu fördern (vgl. Bartnitzky/Speck-Hamdan 2004).

(8.) „*Beratung der Eltern*“: Die Ergebnisse ergänzen die Informationen, die einer Grundschulempfehlung zu Grunde liegen. „Insbesondere der Vergleich auf Landesebene stellt einzigartige Informationen bereit, da er einen Vergleich über Grundschulen hinweg erlaubt.“ (Helmke 2003) – Die bisherigen Lehrer-Beobachtungen und Testformen im Unterricht reichen allerdings für die Beratung der Eltern in der Regel aus. Weitere diagnostische Daten können hilf-

reich sein, sind aber nicht erforderlich. Möglicherweise führen solche Vergleiche zu Misstrauen von Eltern gegenüber den Lehrern.

2. Täuschungsmöglichkeiten bei der Datenerhebung

Im Gegensatz zu den Durchführungsbedingungen der Vergleichsarbeiten, wurden bei PISA wesentlich höhere Standards der Testsicherheit und des Datenschutzes beachtet und beispielsweise folgende Vorsichtsmaßnahmen getroffen (vgl. Deutsches PISA-Konsortium 2001, S. 53 ff):

- die Anonymität der erhobenen Informationen wurde u.a. durch Vernichtung von Namenslisten sichergestellt, so dass nicht möglich war, die erhobenen Daten einzelnen Personen zuzuordnen
- Testleiter waren in der Regel Studierende, die vorab gezielt geschult wurden
- Die Pakete der Testhefte wurden erst kurz vor Beginn der Testsitzung von Testleiter und Schulkoordinator gemeinsam geöffnet
- PISA-Material wurde streng vertraulich behandelt.

Die Datenerhebungsmethoden der Vergleichsarbeiten ermöglichen dagegen Täuschungen im nicht kalkulierbaren Umfang. Falls Schulen dieses Evaluationsverfahren nicht zur „lustbetonten Auseinandersetzung“ mit eigenen Fehlern nutzen, sondern die Fehlerzahl als *Kriterium gesellschaftlicher Selektion* fürchten (vgl. Stähling 2004) und die wissenschaftlichen Durchführungsbedingungen eigenmächtig abändern, muss die Validität der Vergleichsarbeiten angezweifelt werden:

- Testaufgaben konnten wochenlang vorher geübt werden.
- Testzeiten wurden von niemandem kontrolliert und konnten überschritten werden.
- Lehrkräfte konnten Schülern während des Tests bei der Bewältigung von Schwierigkeiten helfen.

Über ähnliche Manipulationen im Grundschulbereich informiert Hans Brügelmann (2005). Besonders bemerkenswert ist, dass zur validen Testung bei vielen Aufgabenstellungen Gespräche, Nachfragen und Klärungen sogar notwendig, aber eigentlich unzulässig sind (vgl. für Deutsch: Bartnitzky 2005; für Mathematik: Selter 2005):

- So werden uneindeutige Aufgaben bzw. „Fallen“ gestellt, jedoch eindeutige Antworten erwartet, die eigentlich erst nach Vorbesprechung und Begriffsklärung mit der Lehrkraft „korrekt“ (im Sinne der Aufgabenkonstrukteure) gegeben werden können.
- Einige Aufgabentypen können nur von solchen Kindern selbstständig gelöst werden, die einen entsprechenden kulturellen Hintergrund oder ein entsprechendes Alltagswissen haben, um z.B. deutsche Redewendungen oder Wetterkarten zu verstehen. Lehrerhilfe liegt bei diesen Aufgaben nahe.
- Zu Lesetexten werden Fragen gestellt, die nur bei einem bestimmten, (teilweise für Fachdidaktiker nicht nachvollziehbaren) Vorverständnis „richtig“ beantwortet werden können. Lehrerunterstützung bietet sich somit an oder ist sogar Voraussetzung zur korrekten Bearbeitung der Aufgabe.

Berichte aus verschiedenen angelsächsischen Ländern bestätigen auch einen nennenswerten Anteil an Täuschungsversuchen durch Lehrer bei vergleichenden Leistungstests (vgl. Brügelmann 2004). Damit wären die Testergebnisse weder für eine schulinterne Bestandsaufnahme noch als Grundlage der Schul-

entwicklung geeignet. Auch die Vergleichsdaten des Landes wären unbrauchbar.

Mit Blick auf die große Stichprobe (nahezu alle mit Ausnahme von bestimmten Sonderschülern oder ausländischen Schülern mit kurzem Aufenthalt) wird die Forschergruppe vermutlich „Täuschungsversuche“ als tolerierbare „Ausreißerwerte“ behandeln, weil angenommen wird, dass Täuschungen nur die Ausnahme sind bzw. in allen Fällen etwa die gleiche Rolle spielen. Infolgedessen können solche Extremwerte vernachlässigt werden und haben als Ausreißer keinen Einfluss auf den zu ermittelnden Mittelwert. Ob allerdings „gefälschte“ Testwerte (Testaufgaben lange vorab geübt, Testzeit überschritten, Hilfen durch Lehrer u.a.) als Ausreißer zu identifizieren sind, hängt davon ab, wie häufig solche Täuschungen vorkommen. Sind sie nicht die Ausnahme, sondern die Regel, dann wäre ein gefälschter Testwert nicht als „Ausreißer“ zu erkennen.

Solche Täuschungsmöglichkeiten sind deswegen wahrscheinlich, weil die Daten nicht anonymisiert werden. Sie dienen der Elterninformation und sollen auch als Basis für Förderungen genutzt werden. Die Eltern erhalten eine Rückmeldung über die Ergebnisse ihres Kindes und können die Testarbeiten einsehen. Zusätzlich bekommen sie Schaubilder mit den Ergebnissen der eigenen Klasse und der Schule, die Vergleiche ermöglichen. In der aus Eltern- und Lehrervertretern zusammengesetzten Schulkonferenz wird über das VERA-Ergebnis der Schule berichtet und der Standort im schulübergreifenden Vergleich dargestellt. Betroffene Schüler und Lehrer können bei schlechten Schülerleistungen mit Konsequenzen rechnen. Schüler und Eltern werden – aufgrund ihrer Schulerfahrungen – erwarten, dass die Testwerte Einfluss auf das Zeugnis haben (schriftlichen Empfehlungen zur weiterführenden Schule bzw. Zeugnisnoten), obgleich offiziell die Daten zu diesem Zweck nicht verwendet werden sollten.

Schulbuchverlage haben den Trend der „Output-Steuerung“ entdeckt und bieten „optimale *Vorbereitungen auf Vergleichsarbeiten*“ auf dem Markt an. Diese auf Vergleichsarbeiten eingestellten Übungsmaterialien finden laut Aussage von Verlagen reißenden Absatz in Schulen. Auch Tageszeitungen steigern ihre Auflagenzahl damit, dass sie dem „Informationsbedürfnis der Eltern“ entgegenkommen und Testaufgaben der Vergleichsarbeiten mehrere Tage vor der Testdurchführung veröffentlichen. Ehrgeizige oder besorgte Eltern lassen solchen Zeitungsmeldungen einige Übungsstunden zu Hause folgen. Dass solche Geschäftsinteressen das eigentliche Anliegen der Vergleichsarbeiten nicht nur konterkarieren, sondern sogar durch öffentliche Vorwegnahme der Testaufgaben die Validität der Leistungsmessung weiter absenken, scheint man in einer Wettbewerbsgesellschaft nicht verhindern zu können.

Zu fragen ist: Wieso wurden diese Manipulationsmöglichkeiten nicht vorab ausgeschlossen? Hat die Forschungsgruppe um Andreas Helmke bei einer so weit angelegten Studie pragmatische bzw. ökonomische Beschränkungen (kein Geld für Testleiter?) berücksichtigen müssen? Die Aussagekraft der Lernstandserhebung ist damit erheblich eingeschränkt. Eine gültige Information zum Lernstand der Schüler eines Bundeslandes kann man durch VERA nicht erhalten. Wenn sich Erziehungswissenschaftler bei der Datenerhebung im wettbewerbsorientierten deutschen Schulsystem bei Lehrern auf das „Zurückstellen eigener Interessen und eigener Eitelkeiten“ (vgl. Husfeld 2004) verließen, würden sie das Selektionsprinzip unseres Bildungswesens in ihren Analysen

vernachlässigen. Um das erste Hauptziel der VERA, die Leistungsmessung oder Bestandsaufnahme zu verfolgen, sind andere Erhebungsmethoden notwendig (z.B. Gewährleistung der Anonymität der zu erhebenden Informationen, Kürzung der Testzeit in der Grundschule, Reduzierung der Aufgabenfülle, Schulung externer Testleiter und Codierer wie bei PISA u.a.). Eine Untersuchung, die im „Feld“ erhobene Schülertestdaten in Beziehung setzen will mit Unterrichtsqualität, wird heuristisches Vorverständnis, eine suchende Erwartungshaltung und vor allem „pädagogischen Takt“ mitzubringen haben.

3. Wirkung auf die Qualitätsentwicklung

Auch um das zweite Hauptziel der Vergleichsarbeiten, die Verbesserung des Lehrens und Lernens zu erreichen, müssen die Datenerhebungs- und Auswertungsmethoden andere sein. Für die Beantwortung der Frage, wie Lehrer ihr „Lehren“ und Schüler ihr Lernen verbessern können, gibt es in der Erziehungswissenschaft sehr unterschiedliche Ansätze, die eine Vielzahl von qualitativen Erhebungsinstrumenten zur Verfügung stellen (biografische Forschungsmethoden, Felderhebungen, Videostudien u.a.). Wenn es um den praktischen Alltag einzelner Schulklassen geht, deren Lehrer die Qualität des Unterrichts steigern wollen, sind Beratungsmethoden von großer Bedeutung (Teamsupervision, Video-Feedback, Unterrichtshospitationen u.a.). Multiprofessionelle Kooperativen in den Schulen versuchen, der Komplexität und Widersprüchlichkeit pädagogischer Alltagsprozesse gerecht zu werden.

Öffentlich präsentierte Leistungsvergleiche (Ranking) von Parallelklassen, die ausdrücklich durch die „Vergleichs-Arbeiten“ angestrebt werden, scheinen besonders dann contraindiziert, wenn in Schulen zuvor keine vertrauensvolle Kooperationsbasis aufgebaut worden ist. Dann nämlich wird ein Vergleich nicht zur (meist wenig kultivierten) Zusammenarbeit ermutigen, sondern Dialoge abbrechen lassen und das häufig zitierte „*Einzelkämpfertum*“ in den Schulen verstärken. Schulforschung wird an den Rahmenbedingungen einer Schule in der Wettbewerbsgesellschaft nichts ändern, muss sich jedoch bei ihrer Verantwortung für die Qualitätsentwicklung von Schule fragen lassen, ob ihre Methoden dem Innovationsauftrag zuwider laufen.

Die angestrebten öffentlichen Leistungsvergleiche (Schulklassen, Schulen) verunsichern die Lehrkräfte und führen in der Regel nicht zu produktiven Auseinandersetzungen, da Eltern in diesem Zusammenhang mehr als fordernde, statt als ermutigende Gesprächspartner erlebt werden. Die unter Eltern üblichen, mehr oder weniger heimlichen Vergleiche von Lehrkräften bzw. Schulen und deren pädagogischen Profilen erhalten durch Vergleichsarbeiten neue, den Output betreffende Argumente. Zu den vielen Spekulationen über die Wirksamkeit unterschiedlicher Unterrichtsformen bzw. Vermittlungsmethoden bei einzelnen Lehrern kommen dann noch die schein-objektiven „Outputdaten“ einer zweifelhaften Schülerleistungsmessung. Welches pädagogische Grundverständnis spiegelt sich darin wider, dass mit den VERA-Daten der inner-schulische Vergleich „ausdrücklich gewünscht“ (vgl. Helmke 2003) und der Wettbewerb zwischen Lehrkräften angestoßen wird?

Nach Veröffentlichung der Ergebnisse von Vergleichsarbeiten ist zu erwarten, dass viele Lehrer die Belastungen durch dann einsetzende Elternkritik als zunehmend unerträglich und entmutigend empfinden. Dies gilt besonders dort, wo die Schulleitung und Schulaufsicht sich nicht schützend vor sie stellt und sozi-

ale Unterstützung ausbleibt. Dann ist möglich, dass der Vergleich das Gegenteil von dem bewirkt, was eigentlich (zu Recht) beabsichtigt ist. Es kann zu einem Aufschaukelungsprozess (vgl. Stähling 1998, S. 27ff) kommen: Die *Veröffentlichung „schlechter“ Schülerleistungen* führt im ersten Schritt zur Beunruhigung der Eltern. Es folgt Kritik an der Arbeit des verantwortlichen Lehrers. Der Stress des bereits durch die erkannten Defizite belasteten Lehrers wird dadurch gesteigert. Eine mögliche Stressreaktion des betroffenen Lehrers ist die Verringerung der Effizienz seines Handelns. Seine professionelle Tätigkeit ist beeinträchtigt (vgl. Stähling 2000). Wenn er an dieser Stelle keine Unterstützung erhält, ist damit zu rechnen, dass seine Schüler in Folge der geringeren Lehrereffizienz und der Verunsicherung durch Elternkritik weniger lernen. Beim nächsten Test zeigen sich wiederholt „schlechte“ Schülerleistungen. Die vorhandenen Unsicherheiten vergrößern sich bei allen beteiligten Personen. Die Gefahr eines Burnouts beim betroffenen Lehrer ist groß. In dieser eventuell einsetzenden Isolierung könnte sich ein Lehrer durch Schüler, Eltern und Kollegen in seinem Wohlbefinden bedroht fühlen. Die Bereitschaft, in solch festgefahrener Situation neue methodische Wege zu erproben, ist sehr gering. Möglicherweise wird der Druck auf leistungsschwache Schüler erhöht und die „Schuld“ für das Defizit ihnen selbst oder deren Eltern zugeschoben. Aufschaukelungsprozesse dieser Art enden nicht selten mit dem für das deutsche Schulsystem bezeichnenden Satz „Die schwachen Schüler gehören hier nicht hin!“ Notwendige Fördermaßnahmen werden nach solchen Etikettierungen meist nicht umgehend eingeleitet, sondern (verspätet) *an andere Institutionen delegiert*: Nachhilfe, Sonderschule, Beratungsstelle u.a., die einen immensen Kostenfaktor darstellen.

Der *heimliche Lehrplan* solcher Aufschaukelungsprozesse verfolgt die verschärfte Selektion und damit den Abstieg des schwachen und randständigen Schülers (vgl. Stähling 2005). Die Vergleichsarbeiten verfehlen in diesen Fällen also ihre eigentlichen offiziellen Ziele, die Qualitätssicherung des Unterrichts und die Förderung besonders von „leistungsschwachen“ Schülern. Jüngere internationale Studien zeigen, dass der Umgang von Lehrerinnen und Lehrern mit extern generierten Daten (zentrale Tests und Prüfungen) nicht – wie beabsichtigt – eine „Evaluationskultur“ oder „Fehlerkultur“ (Helmke 2003), sondern eine *negative Konkurrenzorientierung* zur Folge haben (vgl. Brügelmann 2004). Lehrkräfte nutzen die Resultate von vergleichenden Leistungstests nur dann effektiv, wenn das Schulklima, die interne Kommunikation und das Professionswissen grundsätzlich innovationsförderlich sind (vgl. van Ackeren 2004). Entscheidend ist in diesem Zusammenhang der Umgang mit Fehlern. Fehler gehören zu jedem Lernprozess. Obwohl Fehler unausweichlich sind und als Voraussetzung für jeden Lernvorgang gelten, werden sie in unserem Erziehungswesen als Selektionsinstrument missbraucht. Fehler nützen dann nicht in erster Linie der Information und lustbetonten Auseinandersetzung, sondern sie gelten als Maßstab für „schwache Leistung“, als sackgassenartige „Fehl-Leistung“. Die Fehlerzahl dient als Auswahlkriterium, als scheinbar verlässliches, messbares Merkmal der Selektion. Nur durch Vermeidung von Fehlern und Fehl-Leistungen meint der Schüler einem Versagen zu entkommen. Es entsteht der Eindruck, Kinder gingen zur Schule, um zu lernen, weniger Fehler zu machen (vgl. Stähling 2004a).

Dass sich in Schulen, die Rechenschaft über ihre Leistungen durch zentrale Prüfungen ablegen müssen, die Unterrichtsinhalte und Methoden in der Pra-

xis an den bevorstehenden Prüfungen orientieren („teaching to the test“), belegen Erfahrungen aus Frankreich (vgl. van Ackeren 2004) und den USA (vgl. Brügelmann 2004). Durch solche Standards wird nicht Bildung initiiert, sondern gerade die „bildungsferne“ Furcht vor Fehlern verbreitet! Solche heimlichen Lehrpläne haben allerdings – als negative Auswirkungen des Konzepts „Bildungsstandards“ – mit Bildung nichts zu tun (vgl. ausführlich Schlömerkemper 2004).

4. Fazit

Insgesamt erscheint zweifelhaft, ob es überhaupt irgendeine Erhebungsmethode gibt, die zugleich die zwei genannten Ziele der Vergleichsarbeiten (Leistungsmessung und Leistungsverbesserung bei Lehrern und Schülern) valide bearbeiten kann. Müssten nicht eigentlich die Fragen der Leistungsmessung und der Leistungsverbesserung methodisch getrennt voneinander behandelt werden? Wenn also Vergleichsarbeiten zugleich für wissenschaftliche als auch für Zwecke der Unterrichts- und Schulentwicklung nur sehr eingeschränkt nutzbar sind, stellt sich die Frage, weshalb sämtliche Schüler und Lehrer der Jahrgänge 4 und 9 damit konfrontiert werden müssen, zumal immense Lehrerressourcen und Finanzen aufgeboten werden, die dann bei der Förderung der „schwachen Schüler“ fehlen. Für eine systemorientierte Frage wären Stichproben ausreichend, während für die diagnostisch-pädagogische Arbeit alle Daten genutzt werden können, soweit sie in einem validen Verfahren erhoben werden.

Allein der Korrekturaufwand wird von Brügelmann (2005) auf mindestens einen Halbtage pro Fach und Klasse für die Auswertung und einen Tag für die Eingabe am PC für die zentrale Auswertung geschätzt. Die Ministerin des Landes Nordrhein-Westfalen Ute Schäfer erlaubte folglich den Schulleitern bei Bedarf, die durch VERA belasteten Lehrkräfte für maximal einen Tag freizustellen, wobei aus organisatorischen Gründen Unterrichtsausfall als Möglichkeit angeboten wurde (Brief vom 28.9.04).

Neben den forschungsmethodischen Bedenken gegen ein Instrumentarium wie Vergleichsarbeiten gilt es, den Wert von Schulleistungsmessungen (auch im PISA-Forschungskonzept) grundsätzlich zu hinterfragen. Es handelt sich hierbei um *Black-Box-Messungen*, die keinerlei Aufschluss darüber geben können, durch welche Lernprozesse die gemessenen Schülerleistungen zustande kommen. Horst Rumpf (2004) verweist außerdem auf die offenkundige Tatsache, dass lebenspraktische Kompetenzen der Schüler, deren Erfassung durch PISA beabsichtigt ist, nicht in einem schriftlichen Testverfahren ermittelt werden können: „Es gibt keine Möglichkeit zu überprüfen, ob ein getesteter Schüler die dabei gezeigten Kompetenzen oder Schwächen auch in realen Lebenssituationen in der entsprechenden Ausprägung zeigt. Man denkt unwillkürlich an Fahrprüfungen, die nur auf dem Papier stattfänden. Verfängt sich das Verfahren nicht in tragikomischer Weise in dem Schulwissen, über das es hinauszukommen beabsichtigt – im Interesse der Lebensdienlichkeitsprüfung?“ (S. 89)

Folgt man diesem Gedanken, dann haben wir es in der Unterrichtsforschung mit der Gefahr doppelter Täuschung zu tun:

- Es ist damit zu rechnen, dass Lehrer auf externe Testverfahren mit abweichenden Durchführungsbedingungen (z.B. mehr Testzeit, Vorübungen u.a.) reagieren und damit die Ergebnisse unbrauchbar machen.

- Die Testwerte der Schulleistungsmessungen bilden nur eingeschränkt ab, ob die Leistungen der Schüler lebenspraktischen Wert haben bzw. sie in der Lage sind, damit im Alltag zu bestehen.

5. Gemeinsame Verantwortung von Schule und Forschung

Es bleibt offen und undeutlich, was Gegenstand der Schulforschung ist: Die in der Selbstdarstellung „vorgeführten“ Profile von Schulen oder die – schwer zu erfassende – Realität ohne Hospitanten, Beurteiler oder wissenschaftliche Beobachter. Die Gefahr einer *erziehungswissenschaftlichen Fehldeutung von Unterrichtsalltag* besteht darin, dass die Erforschung von Schule, Unterricht und Lehrerhandeln nicht den Gegenstand ihrer Untersuchung im Auge hat, sondern ein Artefakt. Dieser nicht als solcher herausgestellte Teil von Schulwirklichkeit, nämlich das, was Schulaufsicht, Schulleiter und Lehrer (aus nicht genügend erforschten Gründen) den (häufig ahnungslosen!) Hospitanten und Forschern zeigen, ist nicht immer geeignet, Datenmaterial für die Unterrichtsforschung zu liefern. Die Gültigkeit solcher Datenerhebungen als Basis für die Veränderung von Unterricht ist zu bezweifeln.

Das ungelöste methodische Problem ist die Frage, wie der Teil von Schulalltag, der nicht „vorgeführt“ wird, erfasst werden kann. Ein erster Schritt ist die Erkenntnis, dass zur Weiterentwicklung von Unterricht das verfügbare Wissen und Können *empirischer Methodik und Feldforschung* auf diesen Gegenstand Anwendung finden bzw. die *Forschungsmethodik weiterentwickelt* werden muss. Es fehlen fundierte, repräsentative und gültige Studien. Aufwändige Gruppenuntersuchungen sind aber erst dann sinnvoll, wenn eine begründete Entscheidung vorliegt, welche Variablen zur Untersuchung zugelassen werden sollen.

Im Bereich der Lehrerforschung sind wir noch auf dem Stand der Hypothesenbildung. Bedingungsvariablen wurden in ihrem Zusammenwirken bisher noch nicht genügend beobachtet. Es gilt, die „*Tabus über dem Lehrberuf*“ (Adorno), das „*Streng-Vertrauliche*“, das „*Immer-so-durchmogeln*“ und das „*Nicht-aus-der-Schule-plaudern*“ als Phänomene des Schulsystems neu zu entdecken. Empirische Studien der Bildungsforschung gelten somit als entwicklungsbedürftig. Das Schulsystem ermöglicht aufgrund vielfältiger beruflicher Anforderungen und politischer Erwartungen dem berufserfahrenen Lehrer kaum, der Unterrichtsforschung gültige Daten zu liefern. Unterrichtsalltag und dessen „Output Schülerleistung“ lässt sich nur unklar in seinem *Wirkungszusammenhang* erfassen. Selbst wenn der „Output“ valide erhoben werden könnte (was nicht selbstverständlich ist, siehe Vergleichsarbeiten!), wäre noch nicht geklärt, wodurch der „Effekt“ entstanden ist. Nur wenn die Prozesse innerhalb der „black-box“ erfasst werden, sind Qualitätskriterien für guten Unterricht zu ermitteln.

Somit sind die Grenzen von groß angelegten und kostspieligen Schulleistungsstudien wie PISA und IGLU abgesteckt. In diesem Sinne wünscht man mit Horst Rumpf einer „überanstrengten Erziehungswissenschaft“ und deren Helfern aus der empirischen Psychologie (vgl. Bohl 2004), „sich etwas bescheidener zu geben“ (Rumpf 2004, S.219). Aus Forschungsergebnissen lassen sich keine unmittelbaren, das Lehrerhandeln leitende Schlussfolgerungen ziehen. Angesichts einer teilweise dünnen Datenbasis muss auch der Transfer der wissenschaftlichen Erkenntnis in den Schulalltag bescheiden ausfallen. Die Arbeit der Pädagogen vor Ort, die sich nicht leicht vermessen lässt, befindet

sich in einem besonderen Spannungsverhältnis zur empirischen Erziehungswissenschaft, da der Nutzen der Forschung kaum unmittelbar erfahrbar ist. Ein fruchtbares und produktives Verhältnis von Schule und Forschung kann gelingen, wenn die *Qualität des Unterrichtsalltags* möglichst ohne Tabus in *gemeinsamer Verantwortung* unter die Lupe genommen werden kann. Den Realitäten ins Auge zu sehen fällt dabei schwer, zumal „Feiertagsdidaktiken“ ein ganz anderes Bild von Schule vermitteln. Wer erlebt hat, wie erfahrene Lehrer, die als „Lehrerbildner“ für einige Jahre vom Unterricht befreit, der reinen „Feiertagsdidaktik“ und Vorführstunden-Praxis anheim fielen und anschließend bereits nach fünf Jahren – zu ihrem eigenen Schrecken – nicht mehr in der Lage waren, einen Unterrichtsvormittag in Zufriedenheit zu meistern, der wird den tiefen Graben zwischen Praxis und Wissenschaft erahnen. „Für die Praxis verdorben“, kommentiert dann – achselzuckend – der humorvolle Kollege. Forschung und Schule werden nicht produktiv kooperieren, wenn Forschung sich von gesellschaftlichen Auftraggebern verleiten lässt, den professionellen Experten in den Klassenzimmern „auf die Finger zu hauen“.

Die Qualität der pädagogischen Arbeit ist möglicherweise weniger an (externer) wissenschaftlicher Anerkennung zu messen, sondern an „schulpraktischen“ Maßstäben des Machbaren. Dabei ist die „*kooperative Professionalisierung im Lehrberuf*“ (kurz: KoProfiL, vgl. Schlömerkemper 2001) eine lohnenswerte Perspektive: Forscher, Studierende, Sozialpädagogen, Erzieher, Lehrer, Sonderpädagogen, Fachleiter und andere Experten übernehmen gemeinsam Verantwortung für schulische Arbeit. Sie arbeiten vor Ort mit den Kindern und Lernen in der Auseinandersetzung mit der „echten Praxis“, die komplexe, widersprüchliche Wirklichkeit zu bewältigen. Der Lehrer steht nicht als „Einzelkämpfer“ allein mit dem oft überfordernden Schulalltag. Das Team stützt die Arbeit, indem es Aufgaben verantwortlich plant und durchführt. Teammitglieder verstehen sich zugleich auch als Forschende und Lernende. Die praxisbezogenen Reflexionen des Teams lassen sich auf die Widersprüche pädagogischen Handelns ein. Forschungsarbeiten sind mit der Praxis verzahnt, Fragen erwachsen aus den Notwendigkeiten des Unterrichts und Forschungsergebnisse fließen zurück in den Schulalltag.

Das Wechselspiel von Theorie und Praxis entwickelt sich in diesen Teams zu einem Habitus, den Johann Friedrich Herbart 1802 als „pädagogischen Takt“ beschrieb. Die Arbeit in solchen „Kooperativen“ (Schlömerkemper 2001) erfordert sehr viel „Takt“ und Respekt der verschiedenen Berufsgruppen zueinander. Supervision wird dabei eine hilfreiche Stütze darstellen, da es bei allem pädagogischen Handeln in erster Linie um das soziale Miteinander geht. Praxisberichte über „*multiprofessionelle Teams*“ zeigen, dass „kooperative Professionalisierung“ auch die Qualität integrativer Arbeit steigert und damit einen Beitrag leisten kann zur Weiterentwicklung des integrativen Unterrichts und der „inkluisiven Pädagogik“ (vgl. Stähling 2004b). Das professionelle Selbstverständnis der Schulpraktiker wird dabei aufgewertet. Lehrer werden nicht zu „dummen Schülern“ degradiert, die man durch externe Inspektoren zu kontrollieren hat, sondern sie stehen als kompetente Praktiker auf gleicher Augenhöhe mit den Forschern und tragen gemeinsam Verantwortung mit ihnen. Beispiele „guten“ Unterrichts von „guten“ Lehrern (vgl. Riegel 2004) lassen sich empirisch untersuchen, um Merkmale herauszufiltern, die für die Weiterentwicklung und Veränderung von Bedeutung sind (vgl. Schwarz

1997). Eine Integration von Allgemeiner Didaktik und empirischer Unterrichtsforschung ist dabei anzustreben (vgl. Bohl 2004). Dann dient Forschung der Schule und den dort lernenden Menschen.

Literatur

- Adorno, Theodor W. 1965: Tabus über dem Lehrberuf. In: Kadelbach, Gerd (Hrsg.): Erziehung zur Mündigkeit. Frankfurt/M.: Suhrkamp, 1971, S. 70-87
- Bartnitzky, Horst; Angelika Speck-Hamdan (Hrsg.) 2004: Leistungen der Kinder. Wahrnehmen – würdigen – fördern. Frankfurt/M.: Grundschulverband
- Bartnitzky, Horst 2005: VERA Deutsch 2004: Ungeeignet und bildungsfern. In: Grundschule aktuell, 89, 2005, 1, S. 10-16
- Bohl, Thorsten 2004: Empirische Unterrichtsforschung und Allgemeine Didaktik. In: Die Deutsche Schule, 96, 2004, 4, S. 414-425
- Brügelmann, Hans 2004: „Output“ statt „Input“? – kritische Anmerkungen zur aktuellen Leistungsdebatte. In: Bartnitzky, Horst; Angelika Speck-Hamdan (Hrsg.) 2004: Leistungen der Kinder. Wahrnehmen – würdigen – fördern. Frankfurt/M.: Grundschulverband, S. 10-26
- Brügelmann, Hans 2005: Wahrheit durch VERA? Anmerkungen zum ersten Durchgang der landesweiten Leistungstests in sieben Bundesländern. In: Grundschule aktuell, 89, 2005, 1, S. 7-9
- Deutsches PISA-Konsortium (Hrsg.) 2001: PISA 2000. Opladen: Leske+Budrich
- Helmke, Andreas 2003: Unterrichtsqualität. Seelze: Kallmeyer
- Husfeldt, Vera 2004: Large-Scale-Assessments – Ihr möglicher Beitrag zur Qualitätsentwicklung von Schule und Unterricht. In: Die Deutsche Schule, 96, 2004, 4, S. 500-513
- Riegel, Enja 2004: Schule kann gelingen. Frankfurt: Fischer
- Rumpf, Horst 2004: Diesseits der Belehrungswut. Weinheim: Juventa
- Schlömerkemper, Jörg 2001: Kann pädagogische Theorie alltagstauglich werden? Ansätze zu einer nachhaltigen Lehrerbildung. In: Die Deutsche Schule, 93, 2001, 1, S. 17-29
- Schlömerkemper, Jörg 2004: „Standards“ dürfen „Bildung“ nicht ersetzen! In: Schlömerkemper, Jörg (Hg.): Bildung und Standards. Die Deutsche Schule. 8. Beiheft. Weinheim: Juventa, S. 5-10
- Schwarz, Bernd 1997: Qualität von Lehrern im Spiegel empirischer Forschung – Forschungstendenzen, Probleme und Befunde. In: Schwarz, Bernd; Klaus Prange (Hg.): Schlechte Lehrer/innen. Weinheim: Beltz
- Selter, Christoph 2005: VERA Mathematik 2004. VERbesserungsbedürftige Aufgaben! VERkapttes Ausleseinstrument? In: Grundschule aktuell, 89, 2005, 1, S. 17-20
- Stähling, Reinhard 1998: Beanspruchungen im Lehrerberuf. Einzelfallfeldstudie und Methodenerprobung. Münster: Waxmann
- Stähling, Reinhard 2000: Unterrichtsqualität und Disziplin. In: Grundschule, 32, 2000, 2, S. 20-22
- Stähling, Reinhard 2004a: Schulqualität oder: Lob des Fehlers. In: Grundschulverband-aktuell, 88, 2004, 4, S. 7-10
- Stähling, Reinhard 2004b: Multiprofessionelle Teams in altersgemischten Klassen. Ein Konzept für integrativen Unterricht. In: Die Deutsche Schule, 96, 2004, 1, S. 45-55
- Stähling, Reinhard 2005: Der aufhaltbare Abstieg des schwachen Schülers. In: Die Deutsche Schule, 97, 2005, 1, S. 67-77
- van Ackeren, Isabell 2004: Internationale Erfahrungen für die Entwicklung in Deutschland nutzbar machen! In: Schlömerkemper, Jörg (Hg.): Bildung und Standards. Die Deutsche Schule. 8. Beiheft. Weinheim: Juventa, S. 250-261
- <http://www.uni-landau.de/vera/>

Reinhard Stähling, geb. 1956, Dr. paed., Grundschullehrer
Anschrift: Grundschule Berg Fidel, Hogenbergstr. 160, 48153 Münster;
Email: ggs-bergfidel@gmx.de