



Husfeldt, Vera

Large-Scale-Assessments. Ihr möglicher Beitrag zur Qualitätsentwicklung von Schule und Unterricht

Die Deutsche Schule 96 (2004) 4, S. 500-513



Quellenangabe/ Reference:

Husfeldt, Vera: Large-Scale-Assessments. Ihr möglicher Beitrag zur Qualitätsentwicklung von Schule und Unterricht - In: Die Deutsche Schule 96 (2004) 4, S. 500-513 - URN: urn:nbn:de:0111-pedocs-274358 - DOI: 10.25656/01:27435

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-274358 https://doi.org/10.25656/01:27435

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument is 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 sesest Dokument micht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitlich utzer.

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.

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: Digitalisiert

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung E-Mail: pedocs@dipf.de Internet: www.pedocs.de



Vera Husfeldt

Large-Scale-Assessments

Ihr möglicher Beitrag zur Qualitätsentwicklung von Schule und Unterricht

PISA ist seit der Veröffentlichung der ersten Ergebnisse des "Programme for international student assessment" ein Thema, das wie nur wenige in der breiten Öffentlichkeit diskutiert wird. Zumindest in Deutschland fühlt sich fast jeder und jede von diesen Ergebnissen betroffen. Diese Studie hat uns mit Hilfe von modernen Messverfahren vor Augen geführt, dass die Leistungen deutscher Schülerinnen und Schüler im internationalen Vergleich nur durchschnittlich sind. PISA ist eine von vielen groß angelegten Schulleistungsstudien wie z.B. auch TIMSS, die seit der Mitte der 1990er Jahre ein öffentliches Interesse auf sich ziehen, indem sie Schülerleistungen in unterschiedlichen Bereichen erheben und Vergleiche zwischen unterschiedlichen Systemen zulassen

Aber wozu brauchen wir eigentlich diese groß angelegten Schulleistungsstudien (Large-Scale-Assessments)? Das ist angesichts der steigenden Ausgaben für die Durchführung und die Teilnahme an diesen Studien eine berechtigte Frage. Führen groß angelegte Schulleistungsstudien wie PIRLS/IGLU im Primarschulbereich (Bos et al. 2003; 2004) und PISA im Bereich der Sekundarstufe I (z.B. Deutsches PISA-Konsortium 2001; 2002) wirklich zu neuen Erkenntnissen oder teilen sie uns nur schön statistisch verpackt mit, was wir immer schon wussten, führen uns Defizite vor Augen, die wir längst kennen und die wir sowieso nicht ändern können, vor denen schon Generationen vor uns kapituliert haben? Sind groß angelegte Schulleistungsstudien wirklich in der Lage, die Qualität unseres Bildungssystems zu beeinflussen, oder dienen diese Formen der Evaluation nur den Bildungspolitikern als Steinbruch zur argumentativen Begründung ihres politischen Handelns oder Nicht-Handelns? Tragen sie überhaupt etwas zur Lösung der Probleme oder zur Aufhebung von Defiziten bei?

Gewiss: Der erste Weg zur Besserung ist bekanntlich die Einsicht, jedoch fehlen den Akteuren in Bildungspolitik und Schule meist konkretere Informationen darüber, wie die Defizite im Bildungsbereich abgebaut werden können. Um diese Lücke zu schließen und die gewonnenen Informationen für konkretes politisches und pädagogisches Handeln nutzbar zu machen, bedarf es weiterer Überlegungen, Untersuchungen und der engen Zusammenarbeit der Akteure in Schule, Politik und Forschung. Leistungsmessungen bilden im Rahmen ihrer Anlage und Konzeption eine Voraussetzung zur zielgerichteten Optimierung von Lehr- und Lernprozessen, sie können aber über eine Richtungsweisung hinaus keine konkreten Handlungsanweisungen für ein überlegtes pädagogisches oder politisches Handeln liefern. Im Folgenden sollen vor die-

sem Hintergrund die Funktionen, Möglichkeiten und Perspektiven, aber auch die Grenzen von Leistungsmessungen im Rahmen von Large-Scale-Assessments näher beleuchtet und bewertet werden.

1. Funktionen von Large-Scale-Assessments

Large-Scale-Assessments haben, wie andere Formen der Evaluation auch, mehrere Funktionen und Zicle. Zum einen sollen sie Informationen über Bildungssysteme oder Teilsysteme liefern, zum anderen sollen sie eine umfassende Rechenschaftslegung ermöglichen und schließlich die Grundlage für eine nachhaltige Qualitätsentwicklung bilden (vgl. z.B. Lange 1999; Helmke 2003).

1.1 Informations funktion

Folgt man dem "Input-Prozess-Output-Modell" (z.B. Ditton 2000), so fokussiert die Informationsfunktion auf die Einschätzung der Qualität der eingesetzten Mittel, der Qualität der Lehr- und Lernprozesse und schließlich auf eine Bestandsaufnahme der erreichten Kompetenzen am Output des Bildungssystems. Large-Scale-Assessments bedienen sich dazu des Vergleichs mit einer größeren Anzahl gleicher Institutionen. Auf diese Weise ist es beispielsweise möglich, Aussagen über erwartete Ergebnisse aus Schulen oder Klassen mit vergleichbaren Rahmenbedingungen zu treffen. Large-Scale-Assessments geben somit die Möglichkeit, die in Einzelschulen oder anderen Zusammenhängen vorgefundenen Gegebenheiten damit abzugleichen, was unter den gegebenen Rahmenbedingungen möglich wäre. Sie sind also in der Lage, die Einzelbewertung eines Bildungssystems oder eines Teilsystems in ein größeres, z.B. bundesländerweites, nationales oder internationales Bezugssystem zu bringen und damit deutlicher Qualitäten und Defizite aufzuzeigen. Ein Beispiel dafür ist die jüngste Diskussion um die Bremer IGLU-Bundeslandergebnisse im Kontext der Ergebnisse aus anderen Bundesländern (vgl. Bos et al. 2004).

1.2 Rechenschaftslegung

Generell dienen Evaluationen und damit auch Large-Scale-Assessments nicht nur der Information, sondern sind auch ein Mittel zur Rechenschaftslegung. Das Bildungssystem und jede einzelne Schule, jeder einzelne Lehrer hat einen Bildungsauftrag zu erfüllen, der durch gesellschaftliche Erwartungen, Erwartungen des abnehmenden Arbeitsmarktes, Erwartungen der Steuer- und Rentenzahler, der Eltern und der Schüler selbst determiniert wird. Die jüngsten Diskussionen um die Bildungsstandards (vgl. Klieme et al. 2003) haben die Dringlichkeit dieser Erwartungen und damit die Relevanz des Bildungsauftrags verdeutlicht. Bildungsstandards "legen fest, über welche Kompetenzen ein Schüler, eine Schülerin verfügen muss, wenn wichtige Ziele der Schule als erreicht gelten sollen. Systematisch geordnet werden diese Anforderungen in Kompetenzmodellen, die Aspekte, Abstufungen und Entwicklungsverläufe von Kompetenzen darstellen" (Klieme et al. 2003, 14f.). Nur, welchen Sinn haben die besten Standards, wenn wir nicht wissen, ob sie erreicht werden, oder ob sie erreicht werden können und was zum Erreichen dieser Standards fehlt? An dieser Stelle müssen Evaluationen, sowohl interne als auch externe, die Funktion der Rechenschaftslegung erfüllen. Large-Scale-Assessments können in diesem Zusammenhang wertvolle vergleichende Informationen liefern und dabei zusätzlich eine große Anzahl von beeinflussenden Faktoren berücksichtigen. Mit Hilfe von Large-Scale-Assessments wird es also möglich, den Output des Bildungssystems zu erfassen und einzuschätzen, ob wesentliche, in den Standards festgelegte, Kompetenzen erreicht werden. Bildungsinvestitionen und Bildungserfolge können in einen direkten Zusammenhang gestellt und gemeinsam zur Beurteilung der Effizienz eines Bildungssystems oder Teilsystems herangezogen werden.

Damit werden die bis in den Anfang der 1990er Jahre vorherrschenden Normendebatten in der Bildungspolitik immer häufiger durch empirisch fundierte Argumentationen abgelöst, die sich zumindest in spezifischen untersuchten Bereichen deutlich auf die gemessenen Erfolge bzw. Misserfolge des Systems stützen. Noch bis zur Veröffentlichung der niederschmetternden Ergebnisse der TIMS-Studie im Jahr 1996 (z.B. Beaton et al. 1996a; 1996b) spielten Evaluationen als Instrument zur Beurteilung des Outputs von Bildungssystemen neben der Konzentration auf den Input nur eine nachgeordnete Rolle. Heute werden dagegen Daten aus Large-Scale-Assessments beispielsweise in die OECD-Berichte "Education at a glance" aufgenommen (z.B. OECD 2003).

1.3 Qualitätsentwicklung

Der Prozess des Lehrens und Lernens, der eine Verbindung zwischen dem Input und dem Output des Bildungssystems herstellt, beeinflusst durch seine Oualität in entscheidender Weise das Verhältnis zwischen den Bildungsinvestitionen und den Bildungserfolgen. Untersuchungen, die z.B. aufzeigen, dass die Größe der Klassen bis zu einem bestimmten Maß keinen Einfluss auf die Lernentwicklung der Schülerinnen und Schüler hat (vgl. z.B. Ingenkamp/Petillon/Weiß 1985; von Saldern 1993; Schrader et al. 2001), oder Untersuchungen, die den Zusammenhang zwischen dem Bildungsetat und dem Lernerfolg in unterschiedlichen Ländern nicht bestätigen können (OECD 2003), zeigen eindrücklich, dass z.B. das finanzielle Volumen, das für die Bildung zur Verfügung steht, nicht allein über die Erfolge oder Misserfolge in der Bildung entscheidet. Die Qualität der Schulen und des Unterrichts spielen eine entscheidende Rolle bei der Herausbildung des Wirkungsgrades eines Bildungssystems oder eines Teilsystems. Evaluationen können Qualitäten und Defizite im Bildungssystem aufzeigen und dadurch die Basis gewinnen, die für eine nachhaltige Qualitätsentwicklung genutzt werden kann. Dies kann je nach Art der Evaluation auf allen Ebenen geschehen, sowohl auf Einzelschulebene, auf der Ebene der Bundesländer, auf nationaler oder internationaler Ebene. Die Qualität des Lehr- und Lernprozesses allerdings kann in entscheidender Weise nur in der Einzelschule gesteuert werden. Large-Scale-Assessments sind sehr gut dafür geeignet, Qualitäten und Defizite unter Berücksichtigung der beeinflussenden Faktoren aufzuzeigen, in das Lehr- und Lerngeschehen der Einzelschule reichen sie jedoch bisher in den meisten Fällen nicht hinein. Es können zwar, wie im Anschluss an die Studie QuaSUM in Brandenburg (Lehmann et al. 2000) geschehen, durch Rückmeldungen an die Schulen einige nützliche Informationen geliefert werden, doch die Impulse zur Veränderung und zur Qualitätsentwicklung müssen aus der Schule selbst kommen (vgl. auch Peek 2004; Nilshon 2004).

2. Gegenstand von Large-Scale-Assessments

Der Focus von Large-Scale-Assessments liegt in aller Regel auf der Messung von Leistungen der Schülerinnen und Schüler. Nur am Rande werden in einigen dieser in der Regel querschnittlich angelegten Studien Aspekte der Qualität der Institutionen oder der pädagogischen Praxis (z.B. Leitung der Schule, Lehrerkooperationen, differenziertes oder integriertes Lernen) erhoben.

Schülerleistungen lassen sich z.B. nach einem Modell von Klippert (1983) in die Bereiche kognitive Lerndimension und psychosoziale Lerndimension einteilen, die ihrerseits wiederum in zwei Kompetenzbereiche unterteilt werden können. Für die kognitive Dimension sind dies die Bereiche Sachkompetenz und methodische Kompetenz, und für die psychosoziale Dimension sind es die Bereiche soziale Kompetenz und affektive Kompetenz.

In der überwiegenden Zahl der Large-Scale-Assessments werden Leistungen im Bereich der Sachkompetenz gemessen. Unabhängig von der individuellen Lerngeschichte des einzelnen Schülers oder der einzelnen Schülerin werden die Lernstände in unterschiedlichen Gebieten festgestellt. Sehr häufig entsprechen diese Gebiete den in der Schule vermittelten Fächern Mathematik. Naturwissenschaften, Lesen, Rechtschreiben, Grammatik, Englisch, Politik usw. Diese Schülerleistungen sind in dem jeweiligen curricularen Kontext zu betrachten. Robitaille et al. (1993) stellen das Curriculum dreistufig als intendiertes, implementiertes und erreichtes Curriculum dar. Das intendierte Curriculum wird in Deutschland in Form von Lehrplänen, Stundentafeln der Länder, zugelassenen Schulbüchern und seit neuestem auch in Form von Bildungsstandards festgeschrieben. Das implementierte Curriculum beschreibt den in der jeweiligen Klasse tatsächlich behandelten Stoff und den durch die jeweiligen Lehrer intendierten Erwerb von Kompetenzen. Das erreichte Curriculum schließlich beschreibt die tatsächlich erreichten Kompetenzen der Schülerinnen und Schüler.

Large-Scale-Assessments können sich eng am implementierten Curriculum orientieren, wie das beispielsweise in der QuaSUM-Studie in Brandenburg (Lehmann et al. 2000) praktiziert wurde. In diesem Fall wurde gemessen, inwiefern die erreichten Fähigkeiten der Schülerinnen und Schüler den im Unterricht angestrebten Zielen entsprechen. Die Einführung von Bildungsstandards rückt den Focus ein wenig mehr in die Richtung des intendierten Curriculums. Aus den Bildungsstandards können konkrete Kompetenzmodelle abgeleitet werden, die Teildimensionen und Niveaustufen einzelner Kompetenzen für jeden Lernbereich abbilden sollen. Es geht hier nicht mehr darum, die einzelnen Lehrinhalte festzuschreiben, sondern um eine umfassendere Beschreibung der für das Zurechtkommen in der modernen Gesellschaft fundamentalen Kompetenzen. Diesen Kompetenzbegriff haben vor der Einführung der Bildungsstandards bereits eine Reihe von Large-Scale-Assessments, deren bekanntestes Beispiel vermutlich die PISA-Studie ist, aufgegriffen.

Im Sinne dieser Studien ist der Kompetenzbegriff neben der fachlichen Kompetenz eng mit der methodischen Kompetenz verknüpft, von der die Organisation einer zu lösenden Aufgabe in Form eines Problemlöseprozesses abhängt. Es wird hierbei davon ausgegangen, dass Personen auf der Grundlage von spezifischem Hintergrundwissen eine spezielle Situation verstehen und mit die-

ser Erkenntnis Lösungsschritte entwickeln, erproben und bewerten und ggf. variieren und neu erproben, bis die Bewertung zufriedenstellend ausfällt. Da Probleme stark situations- und kontextspezifisch sind, kann nur bedingt von einer generellen Problemlösefähigkeit gesprochen werden (Peek 1999). Sie geht situationsspezifisch gemeinsam mit der fachlichen Kompetenz in den beispielsweise in PISA verwendeten Kompetenzbegriff mit ein.

Die psychosoziale Lerndimension wird in Large-Scale-Assessments in der Regel durch Selbsteinschätzungen über soziale und affektive Merkmale erhoben. Gerade zur Erfassung von sozialen Kompetenzen bedient man sich dabei häufig bestimmter Stimuli in Form von Situationsbeschreibungen, auf die die Testpersonen dann reagieren. Schülerinnen und Schüler versuchen sich dabei in die beschriebene Situation hineinzuversetzen und auf dieser Basis ihre erwartete Handlung anzugeben. In der Studie Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU) (Max-Planck-Institut für Bildungsforschung 1996) wurde so beispielsweise prosoziales Verhalten gemessen. Da es sich bei diesen Messungen um Simulationen handelt, könnten die Antworten der Teilnehmenden durch Effekte der sozialen Erwünschtheit o.Ä. beeinflusst werden. Diese Effekte sind jedoch in diesem Fall, wie Untersuchungen zur Konstruktvalidität zeigen, vernachlässigbar (vgl. Baumert und Köller 1998).

Affektive Merkmale wie Schulzufriedenheit oder politische Einstellungen werden in den meisten Fällen über Einstellungsskalen gemessen. Aussagen wie z.B. "Unser Land sollte stolz sein, auf das, was es erreicht hat" oder "Die Nationalhymne meines Landes ist mir wichtig", die im Rahmen der IEA-Civic-Education-Study (Torney-Purta et al. 2001; Amadeo et al. 2002) zur Messung von Einstellungen gegenüber der eigenen Nation verwendet wurden, werden von den Testteilnehmern mit der Auswahl einer von vier Antwortalternativen von "stimme überhaupt nicht zu" bis "stimme sehr zu" bewertet. Mit Methoden moderner Skalierungsverfahren können mehrere solcher Items zu Einstellungsskalen zusammengefasst werden.

Insgesamt konzentrieren sich Large-Scale-Assessments jedoch am stärksten auf die Messung kognitiver Lerndimensionen. Psychosoziale Lerndimensionen werden in den meisten Studien lediglich am Rande betrachtet und vielfach nur deshalb, weil sie als Erklärung für die kognitiven Lerndimensionen herangezogen werden sollen, nicht etwa aus eigentlichem Interesse an sozialen und affektiven Kompetenzen.

3. Internationale, nationale und bundesländerweite Large-Scale-Assessments

Large-Scale-Assessments können auf unterschiedlichen Ebenen unterschiedliche Ziele verfolgen und dabei unterschiedliche Wirkungen auslösen. International vergleichende Untersuchungen dienen dazu, die Qualität von nationalen Bildungssystemen in einen größeren Bezugsrahmen zu setzen. Generelle Einschätzungen über bildungspolitische Entscheidungen können zwar daraus abgeleitet werden, doch eine konkrete Ursachenforschung ist in der Regel auf der Grundlage von Daten aus groß angelegten Schulleistungsstudien nicht möglich. Bis in den Lehr- und Lernprozess der Einzelschule reichen internationale Vergleiche in den meisten Fällen nicht, aber selbst die Variabilität unseres föderalen Bildungssystems kann nur schwer damit abgebildet werden, es sei denn,

die Stichproben fallen repräsentativ für jedes Bundesland aus, wie es erstmals mit der Erweiterung der PISA-Studie (PISA/E) realisiert wurde. Mit der PISA/E-Studie und mit der IGLU-Bundesländerauswertung können die Bildungssysteme der einzelnen Bundesländer im bundesdeutschen und im internationalen Kontext beurteilt werden (Deutsches PISA-Konsortium 2002; Bos et al. 2004).

Neben den internationalen Untersuchungen gibt es eine Reihe von Studien wie z.B. LAU (Lehmann und Peek 1997; Lehmann, Gänsfuß und Peek 1999; Lehmann et al. 2001), MARKUS (Helmke und Jäger 2002) und QuaSUM (Lehmann et al. 2000), die beschränkt auf einzelne Bundesländer einen genaueren Vergleich von Schulformen, Einzugsbereichen usw. ermöglichen und dabei den einzelnen Schulen oder sogar den einzelnen Klassen Rückmeldungen darüber geben können, wo sie im Vergleich zu anderen Schulen oder Klassen stehen. Noch stärker nähern sich Untersuchungen an Einzelschulen an, die im Anschluss an groß angelegte Schulleistungsstudien konkretere Analysen in einzelnen Schulen durchführen und die Ergebnisse direkt für den Schulentwicklungsprozess zur Verfügung stellen. Dies sind keine Large-Scale-Assessments, sind aber in ihrer Konzeption an diese angelehnt. Beispiele für solche Untersuchungen sind die Studie von Köller und Trautwein (2001) an fünf hessischen Gesamtschulen und die Untersuchung an der Laborschule Bielefeld (Trautwein et al. 2002).

Die spezifische Bedeutung der Erhebungen – sowohl der bundesländerweiten, der nationalen und der internationalen Erhebungen – liegt im Vergleich und im Abgleich einzelner Segmente des Bildungssystems. Die pädagogischen Schlüsse, die daraus also gezogen werden können, müssen in diesem Licht betrachtet werden. Bundesländerweite Studien mit klassenbezogenen Rückmeldungen sind sicherlich besser dazu geeignet, Schulentwicklungsprozesse in den Einzelschulen in Gang zu setzen, als internationale Studien, die nur einen kleinen Anteil der Gesamtschülerschaft in ihre Stichproben aufnehmen können. Sie sind eher dazu geeignet, die nationalen Ergebnisse international zu verorten und spezifische Probleme oder Qualitäten des Gesamtsystems aufzuzeigen.

4. Möglichkeiten und Grenzen von Large-Scale-Assessments

Acht Jahre nach dem "TIMSS-Schock" und 40 Jahre nach der ersten internationalen Schulleistungsstudie der IEA kann hinsichtlich der Wirkungen von Large-Scale-Assessments eine umfassende Bilanz gezogen werden. Eine oberflächliche Abkehr von den in den 1960er und 1970er Jahren vorherrschenden Normendebatten bei der Umsetzung von Reformvorhaben ist offensichtlich. Zunehmend werden Ergebnisse der unterschiedlichen Schulleistungsstudien zur Feststellung und Erklärung von Defiziten im Bildungssystem herangezogen und kontrovers diskutiert. Der Output des Bildungssystems ist in Form der Leistungsergebnisse der Schülerinnen und Schüler in die Diskussion geraten und spielt heute neben den Inputfaktoren oder im Zusammenhang mit den Inputfaktoren eine wichtige Rolle bei der Begründung von bildungspolitischen Maßnahmen. Bildung ist durch die aufschreckende Wirkung der Vergleichsmessungen insgesamt wesentlich stärker in das Blickfeld öffentlichen Interesses gerückt: es wird auf politischer Ebene mehr darüber nachgedacht,

was veränderungsbedürstig oder veränderungswürdig ist. Die Annahme, das deutsche Bildungssystem nehme im internationalen Vergleich eine Spitzenposition ein, ist einem kritischeren Blick gewichen, der vielfach ein öffentliches Nachdenken über Bildung in Deutschland in Gang gesetzt hat. Dieser selbstkritische Blick und die vermehrten Bildungsdiskussionen sind sicherlich positive Momente, doch werden sie allzu häufig in politischen Diskussionen vor allem im Wahlkampf für die Rechtfertigung beliebiger eigener Standpunkte instrumentalisiert (vgl. auch Herrlitz 2004). Hartmut von Hentig gibt hierzu eine Auswahl aus Vorschlägen zu bildungspolitischen Maßnahmen wieder, die sich auf die Ergebnisse von PISA beziehen, aber sich aus ihnen oder aus den Ergebnissen anderer Studien nicht oder kaum begründen lassen:

"eine wundersame Fülle sich hestig widersprechender einzelner Maßnahmen, fast alle bekannt und prêt-à-porter, fast keines einer weiteren Erkundung und Erprobung bedürstig. [...]: Die Verlängerung der Unterrichtszeit / Die Beendigung des auf eine Million bezifferten Stundenausfalls / 'Inhalte statt Strukturen' / Einführung eines sesten Kanons / Die Entscheidungsbesugnis der Direktoren vermehren / Nachrüsten mit 'interaktivem equipment' / Früheinschulung / Sprachförderung im Kindergarten / Engere Verknüpfung von Kindergärten, Vorschulen und Grundschulen / Keine Aufnahme von Ausländerkindern, die die deutsche Sprache 'nicht beherrschen' [...] / Ganztagsschule 'flächendeckend' / 'Mut zu strenger Benotung' / 'Weniger Staat, mehr Familie' / 'Das Berussbild des Lehrers korrigieren' / Deutliche Maßstäbe für die erwartete Leistung (benchmarking) / Eine größere 'Abnehmerorientierung' ('Schulen und Schüler, die sich als Kunden auf dem Bildungsmarkt begegnen') / 'Mehr Wettbewerb unter den Bundesländern'... " (2003, 229)

Die genannten Verbesserungsvorschläge, die sich teilweise auf den Prozess des Unterrichtens und Lernens in der Schule beziehen, können nur in den seltensten Fällen aus den vorliegenden Untersuchungen abgeleitet werden. Bisher gibt es nur ganz zaghafte, kleinere Initiativen, die dem Lehr- und Lernprozess in der Einzelschule Beachtung schenken oder didaktische Formen im Unterricht genauer empirisch betrachten. Die empirischen Studien, die im Anschluss an PISA oder TIMSS in Einzelschulen durchgeführt wurden (vgl. Köller und Trautwein 2001; Trautwein et al. 2002), nehmen zwar speziell einzelne Schulen etwas genauer in den Blick, können aber darüber hinaus wenig zum Prozess des Lernens und Unterrichtens aussagen. Die Bildungspolitik reagiert in vielen Fällen mit öffentlichkeitswirksamen Maßnahmen, deren Wirkung im Unklaren bleibt. So veranlasst die Landesregierung Niedersachsens kurz nach der Aufnahme ihres Regierungsgeschäftes die Einstellung von 2500 neuen Lehrern, in der Hoffnung, damit könnte die Unterrichtsversorgung nicht nur quantitativ, sondern auch qualitativ optimiert werden (vgl. Busemann 2003). Eine Überprüfung dieser Hoffnung ist aber nicht vorgesehen.

Der Fokus der Large-Scale-Assessments richtet sich neben der Betrachtung von Outputfaktoren außerdem im Wesentlichen auf die Untersuchung von Fachkompetenzen und hier wiederum auf drei spezielle Fachkompetenzen: Lesekompetenz, mathematische Kompetenz und naturwissenschaftliche Kompetenz. Hartmut von Hentig stellt in diesem Zusammenhang zu Recht fest, dass PISA "durch sein Verfahren auf die Erwartung an einzelne, vornehmlich kognitive Leistungen der herkömmlichen Schule beschränkt" bleibt (2003, 221).

Der Mangel von PISA, so seine Kritik, liege "in der Verengung des Blicks auf ein kleines, gut zu bezeichnendes, öffentlich hoch bewertetes Spektrum von

drei Kompetenzen. [...] und erzeugt so den Eindruck, es stünde jeweils die Schule insgesamt auf dem Prüfstand" (ebd., 222). Es lässt sich vermuten, dass die Schulen ihrerseits in vielen Fällen auf die vermehrten Messungen im Bildungsbereich reagieren, indem sie versuchen, messbare Kompetenzen und damit vor allem Kompetenzen in den drei Bereichen Lesen, Mathematik und Naturwissenschaften gezielter zu vermitteln. Hartmut von Hentig befürchtet bereits, dass "man die drei formalen Kompetenzen für die eigentliche und insofern ausreichende "Basis" halten könne – für das Kerncurriculum" (ebd., 225).

Dies ist eine ernst zu nehmende Befürchtung, deren Ursache jedoch nicht maßgeblich den Large-Scale-Assessments zuzuschreiben ist, sondern eher den Personen und Institutionen, die die Ergebnisse aus diesen Studien interpretieren und für sich nutzen. Large-Scale-Assessments erheben weder den Anspruch, Schule so zu verändern, dass sie vollständig in ein messbares Schema passt, noch den Anspruch, pädagogische Konzepte zu ersetzen. Bildungsforscher sind sich darüber im Klaren, dass diese Untersuchungen hinsichtlich der Vielfalt der schulischen Anforderungen unvollständig und in viele Richtungen erweiterbar sind. Neuere Studien wie PISA 2003 (http://www.ipn.unikiel.de/projekte/pisa/) oder Studien im Rahmen des DFG-Schwerpunktprogramms Bildungsqualität in Schulen (BIOUA) (http://www.ipn.unikiel.de/projekte/biqua/biqua.htm) beziehen in stärkerem Maße Prozessvariablen in ihre Erhebungen mit ein, so dass zu hoffen ist, dass in Zukunft präzisere Informationen darüber zur Verfügung stehen, wie der Prozess des Lehrens und Lernens in den Schulen effektiver gestaltet werden kann. Bisherige Studien klammern jedoch den Prozess des Lehrens und Lernens noch weitgehend aus und beziehen sich darüber hinaus mehrheitlich auf die genannten drei formalen Kompetenzen. Sollen also die Möglichkeiten und Grenzen von Large-Scale-Assessments beschrieben werden, so ist dabei zu beachten, dass diese Beschreibung sich in den meisten Fällen auf den Bereich messbarer Kompetenzen am Output des Bildungssystems beschränken muss. Dies berücksichtigend kann nun betrachtet werden, inwiefern die oben genannten Funktionen der Evaluation im Rahmen von Large-Scale-Assessments wahrgenommen wer-

4.1 Informationsfunktion

Hinsichtlich der gewonnenen Informationen aus Large-Scale-Assessments lässt sich sagen, dass diese wenig aussagekräftig für das konkrete Handeln der einzelnen Lehrer und der einzelnen Schule sind. Groß angelegte Schulleistungsstudien geben trotz des schwierigen Vergleichs unterschiedlicher Bildungssysteme dezidierte Auskünfte über mittlere Leistungsunterschiede teilweise auch differenziert nach bestimmten Gruppierungsvariablen wie z.B. Geschlecht. Es lässt sich dadurch die Qualität nationaler Bildungssysteme in einem größeren Kontext einordnen. Vielfach wird dies in Form von Rankingtabellen eindrucksvoll illustriert. Auch wenn es sich dabei nur um grobe Informationen über die mittlere Leistung in ausgewählten Bereichen handelt, ergeben sich für die Bildungspolitik daraus wichtige Hinweise. Diese Hinweise müssen jedoch in einem System, in dem föderale Länder die Bildungshoheit innehaben, hinsichtlich einer Verallgemeinerung vorsichtig interpretiert werden. In der Regel lassen die internationalen Studien nur eine recht vage Einschätzung der Qualitäten eines bundeslandspezifischen Bildungssystems zu, wenn

man denn auf der Grundlage weniger gemessener Kompetenzen in Kombination mit einigen Einflussfaktoren überhaupt auf die Qualität eines Bildungssystems schließen mag. Die Möglichkeit, die Ergebnisse der Bundesländer in ein internationales Bezugssystem zu bringen und damit im internationalen Vergleich den unterschiedlichen Systemen in Deutschland gerecht zu werden, ist erstmals auf der Grundlage der Erweiterungsstudie PISA/E (Deutsches PISA-Konsortium 2002) realisiert worden. Alle anderen internationalen Studien außer IGLU (Bos et al. 2004) - können zwar die Leistungen im Gesamtsystem gut verorten, greifen jedoch bei der Beurteilung der mittleren Schülerleistungen in einzelnen föderalen Bildungssystemen zu kurz. Detaillierte Informationen für kleinere Untereinheiten dieser Systeme werden durch Large-Scale-Assessments demzufolge auch nicht zur Verfügung gestellt. Jede Untersuchung kann natürlicherweise jeweils nur über den gewählten Untersuchungsgegenstand aussagen. Bei internationalen Studien sind dies die Bildungssysteme der einzelnen Nationen, bei nationalen Studien (wie der PISA/E) sind es die föderalen Bildungssysteme der Bundesrepublik Deutschland, nicht aber die Einzelschulen oder gar der Unterricht eines bestimmten Lehrers in einer bestimmten Klasse. Möchte man nähere Angaben über kleinere Untergruppierungen in den einzelnen föderalen Bildungssystemen bekommen, so muss die Studie von vornherein darauf ausgelegt sein. Large-Scale-Assessments können zumindest mit den Möglichkeiten, die uns heute zur Verfügung stehen, noch nicht alle Fragen gleichzeitig in einer einzigen Untersuchung beantworten. Bundesländerspezifische Untersuchungen wie z.B. LAU, MARKUS oder QuaSUM waren in der Lage, genauere Aussagen über einzelne Schulformen, Einzugsgebiete, ja sogar über einzelne Schulen und Klassen zu geben. Über klassenbezogene Rückmeldungen konnten die Lehrkräfte sogar die Leistungsverteilungen ihrer Klassen mit den unter gleichen Bedingungen zu erwartenden Verteilungen vergleichen. Akzeptiert man die oben ausgeführten Grenzen von Large-Scale-Assessments, so kann man sagen, dass innerhalb dieser Grenzen die Informationsvermittlung auf den unterschiedlichen Ebenen je nach Zielpopulation ausgesprochen gut gelungen ist. Darüber darf allerdings nicht vergessen werden, dass außerhalb dieser Einschränkungen noch ein gro-Ber Forschungsbedarf besteht. PISA 2003 scheint hier mit dem Ansatz der Erhebung von unterschiedlichen Prozessvariablen einen Schritt weiter zu gehen.

4.2 Funktion der Rechenschaftslegung

Was die Funktion der Rechenschaftslegung angeht, so ist sie stark an die Informationsvermittlung gekoppelt. Zur Rechenschaftslegung gehört neben der Sicherstellung von Informationen aber noch die Begründung eingesetzter Strategien und die konsequente Planung weiterer Handlungsschritte. Wie bereits oben dargestellt, werden eine Reihe von politischen Entscheidungen mit den Ergebnissen aus Schulleistungsstudien begründet. Vielfach sind diese Begründungen aber aus den Ergebnissen der Studien nicht ableitbar. Vielmehr werden unterschiedliche, sich einander widersprechende Konsequenzen aus denselben Studien gezogen, wie die Diskussionen um die Gesamtschulen eindrucksvoll belegen. Sowohl Befürworter als auch Gegner der Gesamtschulen nutzen die Ergebnisse aus PISA für ihre Argumentationen (z.B. Bundesfachausschuss Bildungspolitik der CDU Deutschlands 2002; AG Bildungspolitik der PDS 2002).

Neben der Rechenschaftslegung gegenüber den Steuerzahlern und Wählern sollen Evaluationen und damit auch die Large-Scale-Assessments auch Rechenschaft gegenüber den unmittelbar betroffenen Eltern und Schülern ablegen. Allerdings hängt es von der Bereitschaft des einzelnen Lehrers oder der einzelnen Lehrerin ab, ob, wenn vorhanden, die klassen- oder schulbezogenen Rückmeldungen, die ja eine Einordnung der Klasse oder der Schule in den größeren Rahmen der Zielpopulation ermöglichen, den Schülern und Eltern bekannt werden. Auch die Entscheidung, inwiefern Schüler oder Eltern in den weiteren Entwicklungsprozess, oder in die konsequente Planung des zukünftigen Unterrichts, eingebunden werden, liegt in der Regel allein in der Hand der Lehrkräfte.

Auch die Stellung, die die staatliche Bildungsverwaltung einnimmt, kommt zumindest im Schulbereich praktisch einer Monopolstellung für das Angebot der Dienstleistung "Bildung" gleich (Baumert und Köller 1998). Daraus ergibt sich auf der einen Seite eine starke Verpflichtung zur optimalen Gestaltung des Angebots sowie zur stetigen wirksamen und öffentlichen Überprüfung der Ziele und der Zielerreichung (Peek 1999). Andererseits ergibt sich zwangsläufig auch ein geringerer Einfluss der betroffenen Eltern und Schüler daraus. Eine echte Rechenschaftslegung im Rahmen von Large-Scale-Assessments ist daher bisher eigentlich nur im Bereich international vergleichender oder bundesländervergleichender Studien möglich. Das bedeutet aber, dass sie an den betroffenen Eltern und Schülern vorbeigeht. Darüber hinaus beschränkt sie sich weitgehend auf den engen Bereich der gemessenen Outputfaktoren.

4.3 Funktion der Qualitätsentwicklung

Den unterschiedlichen Zielpopulationen von Large-Scale-Assessments muss auch bei der Beurteilung der Qualitätsentwicklung als Konsequenz der Ergebnisse Rechnung getragen werden. Vorausgesetzt wird hierbei, dass das Bewusstsein über die eigenen Qualitäten und Defizite einen Prozess in Gang setzt, im Rahmen dessen versucht wird, die Qualitäten auszubauen und die Defizite zu verringern. Diese Qualitätsentwicklung kann einerseits auf der Ebene der Bildungspolitik stattfinden und andererseits konkreter in der Einzelschule. Dass die politischen Konsequenzen aus Large-Scale-Assessments teilweise wenig begründet scheinen, ist bereits weiter oben ausgeführt worden. Auch die Entwicklung in den Einzelschulen kommt bisher nur sehr schleppend in Gang. Dafür sind vor allem drei Probleme verantwortlich: ein Kompetenzproblem, ein Kommunikationsproblem und ein Macht- oder Konkurrenzproblem.

Moderne Datenverarbeitungsmethoden haben es ermöglicht, sehr präzise und sichere Aussagen auf der Basis der erhobenen Daten zu treffen. Neuere Testskalierungsmethoden können den Schülern unabhängig von der Schülerstichprobe und unabhängig vom Testinstrument (wenn ein Kernbereich von Aufgaben in allen Testinstrumenten gleich ist) einen Leistungswert zuordnen. Auf diese Weise ermöglichen sie es, unterschiedliche Schülerpopulationen, die jeweils unterschiedliche Tests bearbeitet haben, miteinander zu vergleichen. Selbst die präzise Erhebung von Lernentwicklungen ist dadurch möglich geworden. Faktoren, die die Leistung beeinflussen, können innerhalb eines Modells auf unterschiedlichen Systemebenen beurteilt werden. Doch je komple-

xer und damit präziser die Verfahren werden, desto schwieriger wird es für Personen außerhalb des Umfelds der empirischen Bildungswissenschaften, diese nachzuvollziehen oder nach eigenen Fragestellungen zu erweitern. Dieses Problem tritt sowohl auf der Ebene der Bildungspolitik als auch auf der Ebene der Einzelschulen auf. Die Distanz zwischen den Vertretern der Bildungsforschung und der Bildungspolitik auf der einen Seite und zwischen den Vertretern der Bildungsforschung und der Einzelschule auf der anderen Seite wird durch den Fortschritt im Bereich empirischer Untersuchungen immer größer.

Diese Distanz ist zum Teil eine Folge und zum Teil eine Ursache des Kommunikationsproblems zwischen den unterschiedlichen Handlungsebenen. In jüngster Zeit gibt es jedoch verstärkte Bemühungen über die Arbeit von Landesinstituten als Schnittstellen zwischen Politik und Wissenschaft die Kommunikation zwischen den beiden Bereichen zu verbessern, Verständnis zu schaffen und sich einander anzunähern. Ähnliche Bemühungen gibt es auch auf Seiten der Schulen und Universitäten, die über Zentren zur Lehrerbildung oder Zentren zur Schul- und Unterrichtsforschung versuchen, im Dialog mehr Verständnis füreinander zu schaffen und gemeinsam Qualitätsentwicklung in Gang zu bringen. Im Göttinger Zentrum für empirische Unterrichts- und Schulforschung (ZeUS) ist beispielsweise eine Verbindung zwischen der Schulpädagogik, der pädagogischen Psychologie und unterschiedlichen Fachdidaktiken auf der einen Seite und den Schulen auf der anderen Seite so geschlossen worden, dass sich eine gegenseitige Anregung für Schul- und Unterrichtsforschung, Schulentwicklung und Unterrichtsgestaltung daraus entwickeln kann. Das Macht- oder Konkurrenzproblem hingegen verhindert in vielen Fällen die notwendige Zusammenarbeit innerhalb der einzelnen Handlungseinheiten. Die Bildungspolitik ist häufig an bestimmte Vorstellungen der jeweils regierenden Parteien gebunden. In manchen Fällen bildungspolitische Reformvorhaben nur als Ausdruck der Abgrenzung zur Opposition und nicht als wirkliche, konsequent durchdachte Bemühung zur Verbesserung des Bildungssystems zu verstehen. Erneuerung kann aber andererseits auch heißen, althergebrachte Traditionen zu hinterfragen und evtl. etwas zu verändern, für dessen Einführung es zu einer anderen Zeit wohl überlegte Argumente gab. Dass dies von politischen Gegnern als Eingeständnis eigener Fehler interpretiert werden könnte, dürste die bildungspolitischen Entscheidungen zusätzlich beeinflussen. Ähnliche Konkurrenzprobleme zeigen sich auch im Bereich der Einzelschule, Evaluation, vor allem interne Evaluation, heißt Offenlegung eigener Stärken, aber auch eigener Schwächen. Ein Engagement für Qualitätsentwicklung bedeutet auch, dass die Qualität noch gesteigert werden kann und dass die eigene Arbeit verbesserungswürdig ist. Das Streben nach einer Kooperation kann in einem Schulsystem, in dem jeder Lehrer und jede Lehrerin für sich selbst kämpft, ebenso als Eingeständnis mangelnder Kompetenz ausgelegt werden. Das Zurückstellen eigener Interessen und eigener Eitelkeiten ist also ein erster wichtiger Schritt sowohl im Zusammenhang mit politischen Entscheidungen als auch bei der Initiative und Durchführung konkreter Projekte zur Schul- und Unterrichtsentwicklung in der Einzelschule. Die Zusammenarbeit von Vertretern aus Schulen, Bildungsforschung und Bildungspolitik in Netzwerken zur Schulentwicklung wie z.B. SQIB (Schulqualität in Berlin) oder QUISS (Qualitätsentwicklung von Schule und Unterricht durch interne und externe Evaluation) sind Beispiele für Initiativen in denen dies gelungen ist.

5. Perspektiven

Large-Scale-Assessments sind im Zusammenhang der Schul- und Unterrichtsentwicklung ein wichtiges Instrument. Sie können Anregungen für schulinterne Diskussionen um die Qualität der Schule und des Unterrichts liefern und damit auch Anstöße für tiefer gehende interne Evaluationen geben. Allerdings sind sie weit davon entfernt, die Lösung für alle Probleme des Schulalltags zu sein. In der Form, die sie bisher durch ihre Entwicklung in 40 Jahren angenommen haben, sind sie in der Lage, umfassende und präzise Informationen über bestimmte Outputfaktoren in einem Kernbereich von messbaren Kompetenzen zu geben. Darüber hinaus sind noch zahlreiche andere Kompetenzbereiche zu bedenken, deren Ausbildung genauso wie z.B. die Ausbildung der Lesekompetenzen zum Auftrag der Schule gehört. Large-Scale-Assessments haben weiterhin bisher wenig über den Prozess des Lernens und Unterrichtens aussagen können. Auch hier besteht noch erheblicher Forschungsbedarf. Klassen- und schulbezogene Rückmeldungen im Zusammenhang mit Fortbildungs- und Entwicklungsprojekten sind in diesem Zusammenhang erste Versuche, sich von Seiten der Bildungsforschung über die direkte Bereitstellung von Informationen Zugang zum Prozess des Lehrens und Lernens zu verschaffen. Bedenkt man, welch kleiner Ausschnitt aus dem Aufgabenspektrum der Schule dabei berücksichtigt wird, wird deutlich, dass wir erst am Anfang eines Prozesses stehen, für den die Erforschung, Beurteilung und Optimierung einer Vielzahl von schul- und unterrichtsbezogener Faktoren große Herausforderungen sind. Die Zusammenarbeit der Akteure in den unterschiedlichen Handlungsebenen in der Bildungspolitik, in der Bildungsforschung und in den Einzelschulen ist Voraussetzung für die Bewältigung dieser Aufgabe. Large-Scale-Assessments haben es vor allem seit der Mitte der 90er Jahre geschafft. Probleme aufzudecken und eine gemeinsame öffentliche Diskussion über Schul- und Unterrichtsentwicklung in Gang zu bringen. Aber sie stellen erst den Anfang dar, auf dem aufbauend tatsächlich die Chance besteht, etwas zu bewirken.

Literatur

AG Bildungspolitik der PDS. 2002: Zukunftswerkstatt Schule, 3/2002

Amadeo, Jo Ann; Judith Torney-Purta, Rainer H. Lehmann, Vera Husfeldt und Roumiana Nikolova. 2002: Civic Knowledge and Engagement among Upper Secondary Students. Citizenship and Education in Sixteen Countries. Amsterdam: IEA

Baumert, Jürgen; Olaf Köller. 1998: Nationale und internationale Schulleistungsstudien. Was können sie leisten, wo sind ihre Grenzen? Pädagogik 6/1998

Beaton, Albert E.; Michael O. Martin, Ina V. S. Mullis, Eugenio J. Gonzales, Teresa A. Smith und Dana L. Kelly 1996a: Science Achievement in the Middle School Years: IEA's TIMSS. Chestnut Hill, MA: Boston College

Beaton, Albert E.; Ina V. S. Mullis, Michael O. Martin, Eugenio J. Gonzales, Dana L. Kelly und Teresa A. Smith 1996b: Mathematics Achievement in the Middle School Years: IEA's TIMSS. Chestnut Hill, MA: Boston College

Bos, Wilfried; Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther und Renate Valtin 2003: IGLU: Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Münster: Waxmann

Bos, Wilfried; Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther 2004: IGLU: Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich. Münster: Waxmann.

- Bundesfachausschuss Bildungspolitik der CDU (2002): Konsequenzen aus PISA: Beschluss des Bundesfachausschusses Bildungspolitik der CDU Deutschlands, 21.02.2002
- http://www.cdu.de/politik-a-z/bundesfachausschuesse/konsequenz-pisa.htm
- Busemann, Bernd 2003: Unterrichtsversorgung und Lehrereinstellung Fundament für gute Schule: Regierungserklärung des Niedersächsischen Kultusministers. Niedersächsisches Kultusministerium
- Deutsches PISA-Konsortium (Hg.) 2001: PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske und Budrich
- Deutsches PISA-Konsortium (Hg.) 2002: PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich. Opladen: Leske und Budrich
- Ditton, Hartmut. 2000: Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. In: Helmke, A., W. Hornstein, E. Terhart, (Hg.): Qualität und Qualitätssicherung im Bildungsbereich. (Zeitschrift für Pädagogik, 41. Beiheft). Weinheim und Basel: Beltz, S. 73-92
- Helmke, Andreas. 2003: Unterrichtsqualität erfassen, bewerten, verbessern. Seelze: Kallmeyersche Verlagsbuchhandlung
- Helmke, Andreas und Reinhold S. Jäger (Hg.). 2002: Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext (S. 39-70). Landau: Verlag Empirische Pädagogik
- Hentig, Hartmut von 2003: Die vermessene Bildung: Die ungewollten Fragen von TIMSS und PISA. In: Neue Sammlung, Heft 2, April/Mai/Juni 2003, S.211-233
- Herrlitz, Hans-Georg 2004: Erziehungswissenschaft und Bildungspolitik zwei getrennte Kulturen. In: Die Deutsche Schule, 96, 2004, 1, S. 6-9
- Ingenkamp, Karl-Heinz; Hanns Petillon, Manfred Weiss 1985 Klassengröße: Je kleiner desto besser? Weinheim: Beltz.
- Klieme, Eckard; Hermann Avenarius, Werner Blum, Peter Döbrich, Hans Gruber, Manfred Prenzel, Kristina Reiss, Kurt Riquarts, Jürgen Rost, Heinz-Ellmar Tenorth, H. J. Vollmer 2003: Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Frankfurt/M: Deutsches Institut für Internationale Pädagogische Forschung.
- Klippert, Heinz 1983: Leistungsbeurteilung in der Schule. In: Erziehen heute, 1, 1983, S. 15ff
- Köller, Olaf; Ulrich Trautwein 2001: Evaluation mit TIMSS-Instrumenten: Untersuchungen in der 8. Jahrgangsstufe an fünf Gesamtschulen. In: Die Deutsche Schule, 93. S. 167-185
- Lange, H. 1999: Qualitätssicherung in Schulen. In: Die Deutsche Schule 91, 2, S. 144-159
- Lehmann, Rainer H.; Rainer Peek 1997: Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996. Hamburg
- Lehmann, Rainer H.; Rüdiger Gänsfuß, Rainer Peek 1999: Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen Klassenstufe 7. Bericht über die Untersuchung im September 1998. Hamburg
- Lehmann, Rainer H.; Rainer Peek, Rüdiger Gänsfuß, Sabine Lutkat, Stefan Mücke, Ingola Barth 2000: QuaSUM Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg. (Reihe Schulforschung in Brandenburg, Heft 1). Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- Lehmann, Rainer H.; Rainer Peek, Rüdiger Gänsfuß, Vera Husfeldt 2001: Aspekte der Lernausgangslage und der Lernentwicklung Klassenstufe 9. Ergebnisse einer längsschnittlichen Untersuchung in Hamburg. Hamburg: Behörde für Bildung und Sport.
- Max-Planck-Institut für Bildungsforschung (Hg.) 1996: Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). 2. Bericht für die Schulen. Berlin

- Nilshon, Ilse. 2004: Qualitätssicherung auf der Grundlage von System-Monitoring. Sechs Fallstudien an Brandenburger Gymnasien. In: Ministerium für Jugend Bildung und Sport (Hg.) Schulrückmeldungen von Schulleistungsstudien am Beispiel des Qua-SUM-Projektes: Zwei Untersuchungen zur Wirksamkeit. Potsdam: Ministerium für Jugend Bildung und Sport
- Organisation for Economic Co-operation and Development (OECD). 2003: Bildung auf einen Blick OECD Indikatoren. Paris: OECD
- Peek, Raincr 2004: Klassenbezogene Rückmeldungen aus Schulleistungsstudien und ihre Rezeption in beteiligten Schulen im Land Brandenburg. In: Ministerium für Jugend Bildung und Sport (Hg.): Schulrückmeldungen von Schulleistungsstudien am Beispiel des QuaSUM-Projektes: Zwei Untersuchungen zur Wirksamkeit. Potsdam: Ministerium für Jugend Bildung und Sport
- Peek, Rainer 1999: Output Control in General Education. In: S. Seeber, J. van Buer (Hg.): Control of Educational Processes Economic and Educational Perspectives. Documentation of the Summer School from 6th September to 11th September 1998 in Sofia (Bulgaria). Sofia und Berlin. S. 69-84
- Robitaille, David F.; W. H. Schmidt, Senta A. Raizen, Curtis McKnight, Edward Britton, Cynthia Nicol 1993: Curriculum frameworks for mathematics and science. Vancouver
- Saldern, Mathias von 1993 Klassengröße als Forschungsgegenstand. Landau: Verlag der Universität
- Schrader, Friedrich-Wilhelm., Andreas Helmke, Ingmar. Hosenfeld, Arnd Ridder 2001: Klassengröße und Mathematikleistung. Empirische Pädagogik, 15, 4, 601-625
- Torney-Purta, Judith; Rainer H. Lehmann, Hans Oswald und Wolfram Schulz 2001: Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen. Amsterdam: IEA
- Trautwein, Ulrich; Petra Stanat, Rainer Watermann, Stefan Krauss, und Martin Brunner 2002: Was die Schule von der Polis lernen kann: Die Laborschule Bielefeld ist anders und erfolgreich. Frankfurter Rundschau 14.11.02

Vera Husfeldt, geb. 1968; Prof. Dr. phil; Studium in Hamburg, Rom und Berlin der Erziehungswissenschaft, Soziologie und Elektrotechnik; seit 2002 Juniorprofessorin für Empirische Schulforschung am Pädagogischen Seminar der Georg-August-Universität Göttingen; Arbeitsschwerpunkte: empirische Schulforschung und international vergleichende Bildungsforschung;

Anschrift: Pädagogisches Seminar, Georg-August-Universität Göttingen, Baurat-Ger-

ber-Str. 4/6, 37073 Göttingen Email: vhusfel@uni-goettingen.de