

# **Schulqualitätsdiskussion in Deutschland – ihre Entwicklung im Überblick<sup>1</sup>**

## **Inhalt:**

- 1 Entstehung der Schulqualitätsdiskussion in Deutschland**
  - 1.1 Der deutsche Bildungsnotstand**
  - 1.2 Kritik an der verwalteten Schule**
  - 1.3 Mangelnde Effizienz der Schulverwaltung**
  - 1.4 Die Kontroverse um die Wirksamkeit von Schule**
- 2 Die einzelne Schule als pädagogische Gestaltungsebene**
  - 2.1 Der „Arbeitskreis Qualität von Schule“**
  - 2.2 Konsequenzen für die Bildungsplanung und Schulentwicklung**
- 3 Unterricht als Kernaufgabe der Schule**
  - 3.1 Entwicklung von Unterrichtsqualität**
  - 3.2 Die systematische Erfassung von Lernleistungen**
- 4 Die Wiederentdeckung der Systemebene**
- 5 Bilanz und Perspektiven**
  - 5.1 Zur Programmatik der Schulqualitätsdiskussion**
  - 5.2 Erträge der Schulqualitätsdiskussion: Erfolge und Ernüchterungen**
- 6 Literatur**

**Wiesbaden, 30. Januar 2007**

---

<sup>1</sup> Erscheint in: Jürgen van Buer & Cornelia Wagner: Qualität von Schule – Entwicklungen zwischen erweiterter Selbstständigkeit, definierten Bildungsstandards und strikter Ergebniskontrolle. Ein kritisches Handbuch. Frankfurt/M. et al.: Lang Verlag 2007.

# 1 Entstehung der Schulqualitätsdiskussion in Deutschland

## 1.1 Der deutsche Bildungsnotstand

Die öffentliche Diskussion über die Leistungsfähigkeit des Schulwesens in der Bundesrepublik Deutschland lässt sich bis in die 1960er Jahre zurückverfolgen. Einen ganz entscheidenden Anstoß stellten dabei die von *Georg Picht* 1964 verfassten Thesen zur „deutschen Bildungskatastrophe“ dar. *Picht* (1965) spricht von „unbefriedigender Qualität“ und vom „unterentwickelten Stand“ des bundesrepublikanischen Bildungswesens, an anderer Stelle vom „erschreckenden Qualitätsverlust“ als Folge von „Irrwegen der deutschen Politik“ in den zurückliegenden 15 Jahren. Das Erziehungs- und Bildungswesen sei „bei weitem nicht mehr in der Lage (...) den durchschnittlichen Bildungsstand unseres Volkes auf einem Niveau zu halten, das den Standards des 20. Jahrhunderts entspricht“.

*Pichts* Thesen erhielten besonderes Gewicht durch den so genannten ‚Sputnikschock‘, der vor allem in den Vereinigten Staaten von Amerika entscheidende Schubkräfte freigesetzt hatte, und erzeugten ein Meinungsklima, das der Reform des Bildungswesens die erste gesellschaftliche Priorität beimaß. Dabei spielte das Argument eine wichtige Rolle, dass Investitionen im Bildungswesen die Wettbewerbsfähigkeit erhöhen und sich damit langfristig wirtschaftlich auszahlen würden. So schrieb *Picht* (1965, 31): „Der Ruin der Schule müßte – das ist unausweichlich – zu einem Ruin der Wirtschaft führen. Notstand des Bildungswesens heißt Notstand der ganzen Gesellschaft (...). Wir brauchen deshalb sofort ein Notstandsprogramm.“

Die im Gefolge der Diskussion über die ‚Bildungskatastrophe‘ einsetzende Schulreform der 1970er Jahre zog ein breites Spektrum an Bemühungen um Veränderungen des Schulsystems nach sich (vgl. dazu die vorliegenden Bilanzierungen von *Fend* 1990; *Friedeburg* 1989; *Hentig* 1990; *Klafki* 1982, *Klemm et al.* 1986; *Tillmann* 1987; *Zedler* 1990 u. a.). Dabei kam es nicht nur zu einer Reform der Schule in ihren zentralen Funktionen (Qualifikations-, Integrations- und Selektionsfunktion), sondern auch zu einer Reform des Steuerungssystems von Schule, die sich an der Kritik der verwalteten Schule entzündete.

## 1.2 Kritik an der verwalteten Schule

Als einer der ersten hatte *Hellmut Becker* bereits 1956 die „verwaltete Schule“ kritisiert. *Beckers* Anliegen war es, „von den oft in Regeln erstarrten Institutionen“ „Handlungsmächtigkeit“ des Individuums durch Beteiligung an den Institutionen zurückzugewinnen. „Eine solche demokratische Partizipation kann durch Bildung ermöglicht werden.“ Dieser Kerngedanke wird von *Becker* an anderer Stelle präzisiert: „Die Fähigkeit der Bürger, die komplexen Prozesse einer verwissenschaftlichten und politischen Gestaltung ihrer eigenen Existenzgrundlagen zu verstehen und an ihnen sinnvoll zu partizipieren, ist nur dann gewährleistet, wenn das Bildungswesen Raum und Gelegenheit gibt, sie zu erwerben und zu üben“ (*Becker & Hager* 1992, 16).

Erst mit der beginnenden Schulreform gerieten die Probleme der ‚verwalteten Schule‘ stärker in den Blick. Wichtige Anstöße gaben vor allem die Aufsätze von *Thomas Ellwein* (1964) und von *Horst Rumpf* über „die Misere der höheren Schule“ (1966), über „die administrative Verstörung der Schule“ (1968) und über den „Unterrichtsbeamten“ (1969). Die Diskussion war jedoch noch maßgeblich auf eine ‚Beseitigung von Anachronismen‘ und ‚Funktionsmängeln‘ ausgerichtet.

Erste psychoanalytische und organisationssoziologische Betrachtungsweisen der Institution Schule (*Fürstenau* 1969) und Analysen zur Lehrerrolle und Arbeitsstruktur in Schulen (*Wellendorf* 1969) lieferten entscheidende Belege für die mangelnde pädagogische Qualität einer bürokratisch verfassten Schule.

Mit der Kritik an der verwalteten Schule korrespondierte das Axiom der pädagogischen Autonomie: „Der pädagogische Prozeß verlangt die organisatorische Selbständigkeit der Bildungseinrichtungen“ (*Richter* 1975, 349). Die Gründe dafür seien in „der Tradierung der herrschenden Kultur und ihrer Legitimierung durch pädagogische Autorität“ zu sehen (*Hopf et al.* 1980, 410). Demgegenüber erlebten Lehrerinnen und Lehrer die Realität als „Terror der Schulverwaltung“ (*Rumpf* 1966). *Rumpf* bezog diese Kritik auf die „Zermürbung der Lehrer durch nichtpädagogische Aufgaben“ und auf unsinnige Verwaltungsvorschriften, die keinen Zusammenhang mit der pädagogischen Qualität des Unterrichts aufwiesen: „Man kann als Lehrer an einem deutschen Gymnasium so gut oder so schlecht unterrichten, wie man will, darum schert sich niemand, wenn man nur die vorgeschriebenen Prüfungshürden genommen hat (...)“ (ebd., 27).

In der Folgezeit ließ sich die Kritik am ‚Bürokratiemodell‘ als die leitende Interpretationsfolie bei der Beschreibung der Schule als Institution ausmachen. Sie durchzog alle organisationssoziologischen Analysen und bestimmte auch das Alltagsverständnis der neuen Lehrergeneration nach 1968.

Im Zuge der genannten Kritik wurde mit einer Schulverwaltungsreform begonnen, die je nach Bundesland unterschiedlich weit voranschritt. Das am weitesten gehende Reformkonzept wurde 1973 vom *Deutschen Bildungsrat* vorgelegt, der eine verstärkte Selbständigkeit der einzelnen Schule mit einer entsprechenden Verlagerung von Entscheidungskompetenzen forderte:

„Die verstärkte Selbständigkeit trägt der Tatsache Rechnung, daß die komplexen Vor-

gänge des Unterrichts nicht bis ins einzelne zentral bestimmt werden können. Die Partizipation der Beteiligten trägt der Tatsache Rechnung, daß eine Institution nicht unabhängig von den in ihr tätigen Menschen wirksam entscheiden und handeln kann“ (*Deutscher Bildungsrat 1973, 17*).

Dabei war der Gedanke leitend, dass in einer komplexer werdenden Gesellschaft die unterschiedlichen Alltagssituationen nicht mehr zentral verwaltet werden können (ebd., A 4): „Die Kontrolle des verschiedenartig und unübersichtlich gewordenen Schulalltags durch einen Schulrat, der nur gelegentlich anwesend sein kann und die besonderen pädagogischen Probleme in den einzelnen Lerngruppen nicht kennt, ist weitgehend eine Fiktion.“

Dabei werden explizit die „Grenzen zentraler Steuerung von Schule und Unterricht“ betont (ebd., A 9). Daraus wird gefolgert: „Die staatliche Verwaltung sollte sich deshalb (...) auf den Erlaß von Rahmenvorgaben beschränken und deren Einhaltung kontrollieren. Die Bewältigung der je besonderen Probleme und Schwierigkeiten in der Schule, die Erarbeitung von Lösungsmöglichkeiten und -alternativen sowie deren Kontrolle ist dagegen eine Aufgabe, die nur in eigener Verantwortung der Schule, in Prozessen der Kooperation und Selbstkontrolle im Medium schulischer Öffentlichkeit geleistet werden kann“ (*Deutscher Bildungsrat 1973, A 11*).

Diese Ideen waren der Bildungsverwaltung zum damaligen Zeitpunkt zu radikal und wurden deshalb von keinem Bundesland bei der zaghaft beginnenden Bildungsverwaltungsreform berücksichtigt. Da eine Bildungsreform ohne eine leistungsfähige Schulverwaltung nicht gelingen kann, breitete sich angesichts einer nach wie vor vorherrschenden retardierten Bildungsverwaltung in der Folgezeit eine „Resignation über die Schwierigkeiten der Bildungsreform“ aus (*Bessoth 1974, 77*).

### **1.3 Mangelnde Effizienz der Schulverwaltung**

Spätestens im Zuge der Kritik an der verwalteten Schule stand die Leistungsfähigkeit der Bildungsverwaltung zur Diskussion. Denn die Kritik an der Bildungsverwaltung beinhaltet im Kern die Frage nach ihrer Effektivität. Eine hierarchisch organisierte Verwaltung arbeitet „nach dem Prinzip von Befehl und Gehorsam“: „Die modernen Planungs- und Managementtechniken, die die Effektivität der Verwaltung steigern, haben im Bildungswesen nur zögernd Eingang gefunden. Von einer Ablösung der bürokratischen durch eine planende Verwaltung kann jedoch noch keine Rede sein“ (*Richter 1975, 348*). So kritisiert beispielsweise *Caspar Kuhlmann* (1972) in seiner Analyse der Zusammenhänge von Bildungsreform und Bildungsverwaltung die Ausrichtung der Schulorganisation auf die Belange einfacher Verwaltung. *Richard Bessoth* kommt in seiner Evaluation der Schulverwaltung zu einem klaren Urteil: „Konkret gesprochen: Die Schulen in den Ländern werden schlecht verwaltet, im Bildungsbereich besteht ein – mehr oder weniger intensiv wahrgenommenes – Mißmanagement“ (1974, 63). Des Weiteren schreibt er: „Daß die Ausgangssituation so unbefriedigend beurteilt werden muß, gilt für alle Verwaltungsbereiche im Schulwesen. Ihre Organisationsstrukturen und die angewandten Verwaltungstechniken entstammen im wesentlichen dem 19. Jahrhundert und sind in vielerlei Hinsicht überaltert und

überholt“ (ebd., 63). In seiner Analyse von Management-Defiziten (1977) kommt er anhand seiner Leistungsindikatoren zu einer differenzierten Beschreibung von Verursachungsfaktoren. Seine Bilanz fällt ziemlich vernichtend für die Schulaufsicht aus, die sich „hart am Rande der Funktionsunfähigkeit“ bewege (ebd., 127) und gar nicht stattfinde (ebd., 102). Eine Untersuchung beim Managementpersonal in der Bildungsverwaltung zeige dabei, „daß es sich nicht um die Summe individueller Inkompetenz handelt, sondern um eine Fülle von Systemfehlern“ (*Bessoth* 1978, 49).

Angesichts des ausgeprägten Beharrungsvermögens und der systembedingten Lernresistenz der etablierten Schulbürokratie plädiert der Autor für einen radikalen Strukturwandel nach dem Muster privatwirtschaftlicher Management-Konzepte. Nur durch eine Transformation „von einem bürokratischen System zu einem offenen System“ lasse sich eine „Revitalisierung“ der Bildungsverwaltung erreichen (*Bessoth* 1995).

Dass die Funktionsweise der Schulaufsicht auch von anderen Experten kritisch beurteilt wurde, zeigen weitere Untersuchungen (*Allmann* 1994 und *Koetz et al.* 1994). Die Wirksamkeit der Schulaufsicht wurde nach wie vor durch eine Flut von Gesetzen und vor allem durch die Verwaltungsvorschriften eingegrenzt: „Für die Schulleitung und die LehrerInnen stellt sich das Problem, ihre Freiheiten bei der Vielzahl der Regelungen zu erkennen und sich danach zu richten, zumal die SchulleiterInnen PädagogInnen und keine JuristInnen sind. Für die Schulaufsicht liegt das Problem im Vollzug dieser unübersichtlichen Regelungen“ (*Koetz et al.* 1994, 121).

*Werner Biewer* kommt in seiner Untersuchung zur Schulaufsicht zu dem Ergebnis, „daß der Einfluß der Schulverwaltung auf die Ausprägungen der Verwaltungsvorgaben bei den SchulleiterInnen und LehrerInnen trotz zentralisierter Steuerung und Kontrolle gering ist“ (*Biewer* 1993, 179). *Rosenbusch* (1992, 53) stellt in seiner Untersuchung fest, dass die Schulaufsicht „in der bisherigen Form durch die LehrerInnen nicht als angemessen eingeschätzt, sondern als überholt bzw. völlig überholt angesehen (wird) (...) LehrerInnen sehen in der Institution des Schulrates eher eine Verwaltungsinstanz, mit der keine pädagogisch kollegiale Verbindung mehr besteht (...). Zwischen Schulaufsichtsbeamten und LehrerInnen findet kaum ein Transfer praktischen oder theoretischen Wissens statt.“ Und an anderer Stelle schreibt *Rosenbusch* (1992, 54): „Beratungsgespräche mit dem Schulrat im bisherigen Modus werden generell als nicht effektiv eingeschätzt. Fast 75 % der LehrerInnen sehen kaum eine Wirksamkeit gegeben.“

Diese und weitere Untersuchungen (*Burkhard & Rolff* 1994; *Gampe* 1994; *Lengen* 1989) machen vor allem zweierlei deutlich:

- den hohen Anteil an verwaltender Tätigkeit bei Schulaufsichtsbeamten sowie
- ihr geringes Ausmaß an beratender Tätigkeit.

Diese Tätigkeitsbeschreibung steht in geradezu diametralem Gegensatz zum Wunschbild von Schulaufsichtspersonal. Schulräte wollen beraten (vgl. *Bruckmann* 1986), nur die Verhältnisse am Arbeitsplatz sind nicht so. Auch Lehrerinnen und Lehrer wünschen sich Beratung, „wenn sie in einem möglichst herrschaftsfreien Raum erfolgt, d. h. auf keinen Fall im Zusammenhang mit der Beurteilung“ (*Rosenbusch* 1992, 55).

Die konstatierte Diskrepanz hat mit der „Überfrachtung des Aufgabenprofils“ zu tun, wie *Schratz* (1996) in seiner Untersuchung über „Die Rolle der Schulaufsicht in der autonomen Schulentwicklung“ anschaulich herausarbeitet („Bin für alles zuständig“). Insgesamt dominiere ein Bild des „Verwaltungsfachmanns“ (*Burkhard & Rolff* 1994, 240). Auch die in den 1990er Jahren einsetzenden systematischen Bemühungen um eine Kooperation zwischen Schulaufsicht und Schule fallen nach *Rosenbusch & Schlemmer* (1999) enttäuschend aus. So sei es nicht gelungen, „das häufig zitierte ‚strukturell gestörte Verhältnis‘ zwischen Lehrerschaft und Schulaufsicht aufzulösen oder gar endgültig zu beseitigen. Selbst in Bremen, der entschiedensten Neuregelung, gelingt es nur schwer, die historisch gewachsenen Vorbehalte abzubauen, obwohl, wie gesagt, die Mitglieder der Schulinspektion weder Dienstvorgesetzte sind noch als solche auftreten“ (*Rosenbusch & Schlemmer* 1999, 60).

#### **1.4 Die Kontroverse um die Wirksamkeit von Schule**

Mit der Schulreform der 1970er Jahre wurden zwar große finanzielle Anstrengungen zur Verbesserung des Schulsystems in inhaltlicher, personeller und sachlicher Hinsicht unternommen, dennoch spielte die Frage nach der Effektivität dieser Anstrengungen so gut wie keine Rolle. Bereits die Anstrengungen selbst, der Anstieg höherer Bildungsabschlüsse, positive Indikatoren für Chancengerechtigkeit sowie veränderte Studiengänge, eine reformierte Lehrerbildung, neue Lernangebote, Curricula und pädagogische Konzepte wurden als Beleg für die Leistungswirksamkeit angesehen. Die Frage, in welchem Verhältnis die investierten Mittel zum Ertrag der Bildungsanstrengungen stehen, wurde nicht gestellt. Die in jener Zeit aufkommende Schulbegleitforschung sowie die zahlreichen wissenschaftlichen Begleitungen der Innovationsprojekte, vor allem der Modellversuche der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK), richteten ihr Augenmerk primär auf die Entwicklungsberatung und nur sekundär auf Probleme bei der Realisierung und Möglichkeiten der Verbesserung der Maßnahmen. Nicht von ungefähr beklagten die Reformbeteiligten – bemerkenswerterweise Schulaufsichts-, Schulleitungs- und Lehrpersonen gleichermaßen – die mangelnde wissenschaftliche Aufarbeitung der Versuchserfahrungen (vgl. *Steffens* 1979).

Eine bekannte Ausnahme gab es allerdings: Die Wirksamkeit und Effizienz von Schule begann dort zu interessieren, wo die Grundstruktur unseres Schulsystems zur Debatte stand, nämlich die Dreigliederung der allgemeinbildenden Schulen in Hauptschule, Realschule und Gymnasium. Mit der Einrichtung von Gesamtschulen kam die Angst vor „Leistungseinbußen“ des deutschen Schulsystems auf. In diesem Zusammenhang und im Interesse einer Legitimierung der jeweiligen Positionen gaben nacheinander zuerst Befürworter des Gesamtschulsystems (vor allem die Länder Hessen, Nordrhein-Westfalen und Niedersachsen) und später auch Opponenten (insbesondere die Länder Rheinland-Pfalz und Baden-Württemberg) größere Evaluationsstudien in Auftrag. In langwieriger Arbeit hat sich auch der Innovationsausschuss der Bund-Länder-Konferenz mit den Erfahrungen des Schulsystemvergleichs befasst und die Ergebnisse in einem ca. 800-seitigen Bericht niedergelegt (*Bund-Länder-Kommission* 1982). In Abhängigkeit von parteipolitischen Positionen gab es dabei divergierende

Bilanzierungen in den Evaluationsbereichen Kosten und fachliche Leistungen. Während die Politiker sich also teilweise uneins waren, konnte in der Forschung ein weitgehend einmütiges Ergebnis konstatiert werden, was *Helmut Fend* in dem lapidaren Satz zum Ausdruck brachte, dass die Gesamtschulfrage definitiv beantwortet werden konnte (1982).

Von der Ausnahme der Gesamtschulforschung abgesehen kann allerdings zuge-spitzt festgehalten werden, dass die Frage nach der Wirksamkeit von Schule in der Schulreform der 1970er Jahre kaum gestellt und noch weniger debattiert wurde.

Ganz anders verlief demgegenüber die Diskussion über die Effektivität des Schulsystems in den anglo-amerikanischen Ländern, allen voran den USA. Obwohl Untersuchungen zur Leistungsfähigkeit von Schule bereits in den 1950er und frühen 1960er Jahren durchgeführt wurden, erhielt die Effektivitätsfrage erst mit den Ergebnissen aus der größten jemals durchgeführten Schuluntersuchung, nämlich dem berühmten Coleman-Report aus dem Jahr 1966, einen breiten Stellenwert in der Bildungsreformdebatte der USA. Die Untersuchung wurde vom U.S. Office of Education in Auftrag gegeben und sollte die vermutete Chancenungleichheit im staatlichen Schulwesen offenlegen. Im Ergebnisreport wurde die niederschmetternde Bilanz gezogen, dass Schule nur einen kleinen Einfluss auf den Lernerfolg von Schülerinnen und Schülern habe und dass „Background“-Bedingungen (Elternhaus, Intelligenz u. ä.) von weit größerem Gewicht seien. Das Fazit lautete: „Schools do not make a difference“ (vgl. auch *Jencks* 1973).

*Weiß* (1985, 1063) merkt dazu an: „Wohl kaum eine empirische Untersuchung im Bildungsbereich hat eine so intensive wissenschaftliche Debatte ausgelöst wie der Coleman-Report. Das Ergebnis dieser Debatte stellt einen eindrucksvollen Beleg für die Wichtigkeit der Kontrollfunktion der ‚scientific community‘ dar.“ Diese Debatte wurde begleitet von vielen Studien der gleichen Machart: „In diesen Studien stand vor allem das Ziel im Vordergrund, nach (statistischer) Ausschaltung des Einflusses außerschulischer Faktoren, insbesondere aus dem familialen Bereich, die Effektivität bildungspolitischer Gestaltungsparameter, vor allem der schulischen Ressourcen, offenzulegen. Auslöser für solche Untersuchungen zur Ressourcenwirksamkeit war weniger ein wissenschaftliches, auf Theorieprüfung gerichtetes Interesse, sondern vor allem ein wachsender Informationsbedarf für Zwecke der Politikberatung, im Wesentlichen verursacht durch die in der zweiten Hälfte der 60er Jahre einsetzende ‚Accountability-Bewegung‘“ (vgl. *Richter* 1975). Die für das Bildungswesen Verantwortlichen sahen sich zunehmend mit der Forderung konfrontiert, über die Konsequenzen ihres Handelns Rechenschaft abzulegen und einen Effektivitätsnachweis über die Verwendung der ihnen zur Verfügung gestellten Mittel zu erbringen“ (*Weiss* 1985, 1061).

Obwohl in diesen Studien ein breites Spektrum an Bemühungen um Schule und Unterricht erfasst wurde, zeigte sich insgesamt ein ernüchterndes Bild:

- 1) Über die Wirksamkeit von Schule gibt es erst wenig abgesichertes Wissen.
- 2) Aus dem, was man bisher weiß, ist zu entnehmen, dass Schule nicht viel bewirkt.
- 3) „More research is needed.“

Diese Folgerungen bilden den roten Faden der von *Averch et al.* (1974) präsentierten Befunde zu einer Vielzahl empirischer Studien. Die Arbeit hat den Titel „How Effective is Schooling?“

So gesehen musste damals konstatiert werden, dass trotz der Vielzahl von Untersuchungen die empirische Schulforschung erst relativ wenig abgesicherte Aussagen über die Wirkung von Schule und über innerschulische Wirkungszusammenhänge geliefert hatte. Die Probleme sind in erster Linie in den methodischen Schwierigkeiten einer adäquaten Erfassung schulischer Praxis zu sehen. Darauf verweisen auch *Averch et al.* (ebd., 171). Die Kritik gilt u. a. dem Sachverhalt, dass das ganze Geschehen im Klassenzimmer noch zu wenig untersucht wurde und dass die verwendeten Maße zu grob waren, um das zu erfassen, was wirklich in der Schule passiert. Das ist auch der entscheidende Ausgangspunkt einer neuen Forschungsrichtung, „die aus der Einsicht heraus entstanden ist, daß von einer – auch methodisch verbesserten – Schuleffekt-Forschung, die sich auf das Input-Output-Paradigma stützt, kein nennenswerter Erkenntniszuwachs zu erwarten ist. Als Hauptgrund dafür wird das Fehlen von binnenschulischen Prozeß-Variablen angeführt“ (*Weiss* 1985, 1084).

Anschließende Untersuchungen, die von einer Skepsis gegenüber den Befunden und Schlussfolgerungen von *Coleman et al.* (1966) und *Jencks* (1973) geleitet waren, kamen zu entgegengesetzten Folgerungen über die Wirksamkeit von Schule. Diese Untersuchungen folgten anderen Fragestellungen und setzten andere Forschungsschwerpunkte als die bisherigen. Sie widersprachen dabei nicht den bekannten Befunden über die zentrale Bedeutung schulischer Background-Faktoren (wie sozio-ökonomischer Status). Sie kamen aber zu anderen Ergebnissen und Schlussfolgerungen, da sie andere Aspekte schulischer Prozesse in den Mittelpunkt ihrer Analysen rückten und sich auch anderer Untersuchungsmethoden bedienten. Die weit verbreitete Grundüberzeugung war, dass „schools can make a difference“.

Die ersten Studien dieser Machart entstanden bereits Anfang der 1970er Jahre (*Weber* 1971; *Klitgaard & Hall* 1973; *New York State Department of Education* 1974). Weitere und differenziertere Studien folgten im weiteren Verlauf der 1970er Jahre (*Madden & Lawson* 1976; *Brookover et al.* 1976; *Brookover & Lezotte* 1977; *Brookover et al.* 1977). Mit den Überblicksarbeiten von *Edmonds* (1979; 1981) erfuhren die neuen Forschungsergebnisse eine erste Breitenwirkung. Am bekanntesten wurde später der Review von *Purkey & Smith* aus dem Jahr 1983 (auf Deutsch: *Purkey & Smith* 1990). Für eine internationale Verbreitung dieses Ansatzes spielten die ab 1987 durchgeführten internationalen Kongresse zur „School Effectiveness“ eine wichtige Rolle (vgl. dazu *Creemers et al.* 1989).

## 2 Die einzelne Schule als pädagogische Gestaltungsebene

Erste Befunde in der deutschen Schulforschung stammen aus der Mitte der 1970er Jahre und entstanden im Rahmen der Konstanzer Schulforschung. Der erste Nachweis dieses Ansatzes findet sich in der Arbeit von *Helmut Fend* aus dem Jahr 1977. Weitere und differenziertere Befunde wurden von *Fend* und Mitarbeitern in der Folgezeit vorgelegt.

Die englische Schulforschung lieferte vergleichbare Ergebnisse mit der Publikation der Forschungsarbeit über zwölf Londoner Gesamtschulen von *Rutter et al.* aus dem Jahr 1979. Diese Arbeit wurde in Deutschland durch die von *Hartmut von Hentig* kommentierte deutsche Übersetzung von 1980 bekannt (*Rutter et al.* 1980) und hatte in der Bundesrepublik – neben den Untersuchungsergebnissen zur Gesamtschule – bis zu diesem Zeitpunkt den größten Einfluss auf die Diskussion zur Effektivität und Qualität von Schule.

*Rutter et al.* zeigten sehr anschaulich auf, dass Schulen sehr verschieden sein können, und nannten mögliche Bedingungen für die schulischen Unterschiede. Im Mittelpunkt standen dabei ihre Ausführungen, dass die Unterschiede der Sekundarschulen „selbst dann erhalten (bleiben), wenn in der Analyse nur diejenigen SchülerInnen berücksichtigt wurden, deren Familiensituation und Persönlichkeitsmerkmale vor dem Übergang zur Sekundarstufe weitgehend vergleichbar waren“ (ebd., 210), und dass „die festgestellten Ergebnisunterschiede in engem und systematischem Zusammenhang mit bestimmten situativen Merkmalen (standen), die für die Schule als soziale Organisation charakteristisch waren“ (ebd., 241). Weiter heißt es: „Sekundarschulen scheinen mithin, anders als bisweilen vermutet, auf die Leistungen und Verhaltensweisen ihrer SchülerInnen einen durchaus nicht unbedeutenden Einfluß auszuüben“ (ebd., 241). Damit lagen neben anderen weniger bekannt gewordenen bzw. nicht in diesen spezifischen Diskussionszusammenhang eingebrachten Resultaten Ergebnisse zur Schulforschung vor, die bei der – insbesondere im anglo-amerikanischen Raum – geführten Kontroverse über die Wirksamkeit von Schule die Position jener zu stützen scheinen, die trotz der ernüchternden Befunde zur kompensatorischen Erziehung und den geringen Fördermöglichkeiten von Schülerinnen und Schülern weiterhin die These verfochten, dass „schools can make a difference“.

Die Ergebnisse ergänzen den Diskussionsstand zur Gesamtschulforschung, die dem Schulsystemeffekt eine relativ geringe Bedeutung zuspricht (für die Schulleistungen liegt der Prozentsatz aufgeklärter Varianz bei einem Prozent und darunter), wohingegen der Faktor Einzelschule (die Schulsituation) einen deutlich wichtigeren Stellenwert einnimmt (6-8% aufgeklärte Varianz), der fast an die Bedeutung der Schülerintelligenz (mit 10-11%) heranreicht und andere Background-Faktoren wie die Sozial-schicht (mit 2-3%) übertrifft (vgl. *Fend et al.* 1980, 686).

Mit der Betrachtung einzelner Schulen im Vergleich zu anderen, also mit der Behandlung von Schulen als ‚Fällen‘, hatte die Konstanzer Schulforschung (unter Leitung von *Helmut Fend*) in entscheidender Weise die Frage nach der Schulqualität ins Blickfeld gerückt. Maßgeblich dafür waren insbesondere die Erfahrungen mit ganz unterschiedlichen ‚Schulgestalten‘ trotz einheitlicher Rahmenbedingungen und ähnlicher Schülerschaft, wie etwa beim ehemaligen Wetzlarer Flächenversuch in Hessen mit zwölf Integrierten Gesamtschulen.

Damit war die Einsicht verbunden, dass der Schule als pädagogischem Handlungsfeld eine entscheidende Bedeutung zukommt. Wie sich eine einzelne Schule gestaltet, wie sie arbeitet, wie in ihr gelebt und gelernt wird, wie Lehrerinnen und Lehrer miteinander und mit den Schülern umgehen, was sie von den Schülern verlangen – das alles wird zwar von Gesetzen und Erlassen vorstrukturiert, ihr typisches Gepräge und

ihre besondere Gestalt bekommt die Schule jedoch erst durch die in der einzelnen Schule handelnden Personen. Schulen sind ‚Individualitäten‘ mit eigenem Profil und unterschiedlicher Qualität.

Diese Umorientierung auf dem Gebiet der Schulentwicklung verlangt im Kern, die einzelne Schule als „pädagogische Handlungseinheit“ (*Fend 1987*) zu betrachten und zu behandeln. Schulen werden demnach als „soziale Individualitäten“ mit ihrem je eigenen Profil begriffen (vgl. *Steffens & Bargel 1987a*). Damit wird zugleich die innere Gestaltung und Führung der Schule als Ganzes ins Zentrum der Aufmerksamkeit gerückt: deren Zielgerichtetheit und Strukturierung, das Ausmaß der Zusammenarbeit in den Kollegien, die Art der Lehr- und Lernprozesse bis hin zu den Formen des Umgangs mit- oder gegeneinander im Alltag der Schulen.

In dieser Perspektive wird der Ausgestaltungsspielraum betont, der jeder Schule zu eigen ist und der in Kenntnis der damaligen Schulforschungsergebnisse als beträchtlich angesehen wurde: Die spezifische Qualität einer Schule ist als Resultat kollegialer Problemlösungsprozesse zu begreifen, wobei die Art und Weise der gewählten Lösungen von den wahrgenommenen Rahmenbedingungen und Problemkonstellationen, vom pädagogischen Anspruch und von der Einstellung und den Fähigkeiten der Lehrerinnen und Lehrer sowie von den Aushandlungsprozessen im Kollegium und der Verbindlichkeit ihrer Einhaltung beeinflusst wird (vgl. u. a. *Steffens & Bargel 1993*). Schulqualität, so die Argumentation, entsteht auf der Handlungsebene der einzelnen Schule. Mit dieser Akzentuierung wurde nicht die Notwendigkeit von Bemühungen um die ‚äußere‘ Schulreform verkannt, sie wurde jedoch als Reaktion auf die Erfahrungen verstanden, dass selbst die bestgemeinten und -fundierte Reformvorhaben in der Schulpraxis in ihr Gegenteil verkehrt werden können, wenn die innerschulischen Entwicklungsprozesse nicht mitbedacht werden. Äußere und innere Reform müssen demnach Hand in Hand gehen (vgl. dazu *Roeder 1992*).

Die Unterschiedlichkeit von Schulen trotz einheitlicher Rahmenbedingungen und die begrenzten Möglichkeiten „makroorganisatorischer Vorstrukturierungen“ schulischen Geschehens hatte bereits *Fend (1977)* anschaulich aufgezeigt (s. a. *Fend 1980*). Seine Analysen können als der eigentliche Ausgangspunkt der dann in den 1980er Jahren mit dem „Arbeitskreis Qualität von Schule“ in Deutschland einsetzenden Schulqualitätsdiskussion und Schulqualitätsentwicklung begriffen werden (vgl. dazu *Steffens & Bargel 1987b*).

## **2.1 Der „Arbeitskreis Qualität von Schule“**

Die vorangehend skizzierte neue Sichtweise von Schule hat in der deutschen Schulforschung und Schulentwicklung eine breitere Rezeption erfahren. Sowohl Initiator als auch Diskussionsforum war der „Arbeitskreis Qualität von Schule“, der 1985 eingerichtet wurde, um eine kontinuierliche Diskussion über die Qualität von Schule und Unterricht anzuregen und zu bündeln, in deren Mittelpunkt die Frage steht: „Was ist eine gute Schule, und wie ist sie zu verwirklichen?“ Hier kamen im jährlichen Turnus, auf Einladung des Hessischen Instituts für Bildungsplanung und Schulentwicklung (HIBS) bzw. (später) des Hessischen Landesinstituts für Pädagogik (HeLP) in Wies-

baden sowie der Forschungsgruppe Gesellschaft und Region e. V. in Konstanz, Experten aus Wissenschaft und Praxis zusammen, um pädagogische Vorstellungen, empirische Befunde und praktische Erfahrungen zu schulischen Qualitätsmerkmalen systematisch zusammenzutragen und wesentliche Aspekte für das Gelingen von Schule und Unterricht herauszuarbeiten.

Mit der Verlagerung des Blickwinkels hin zur einzelnen Schule als „pädagogischer Handlungseinheit“ (Fend 1987) schienen unvermittelt verschiedene Forschungs- und Interventionsansätze, die sich bisher in einem unverbundenen Nebeneinander bewegten, (sich gegenseitig ergänzend) ineinander aufzugehen:

- 1) Die Befunde aus der bundesrepublikanischen Gesamtschulforschung (Fend 1982) fanden ihre Entsprechungen in neueren Studien über Wirkungen von Schulbedingungen auf Schülerinnen und Schülern (namentlich Rutter et al. 1979). Bei Gegenüberstellungen von Schulen mit hohem und niedrigem Leistungserfolg ihrer Schülerinnen und Schüler, wie sie in amerikanischen Studien seit Ende der 1970er Jahre vorgenommen wurden (vgl. z. B. Edmonds 1979), orientierten sich die Analysen an den schulspezifischen Bedingungen. Fallstudien bestätigten und ergänzten die gewonnenen Ergebnisse (vgl. z. B. Lightfoot 1983). Studien über schulische Veränderungsprozesse ließen erkennen, dass der Erfolg von Reformprogrammen im Wesentlichen von den konkreten Bedingungen einer Schule und eines Kollegiums abhängt (vgl. u. a. Fullan 1982; Roeder 1987). Evaluationsstudien über Curriculumprogramme betrachteten die erzielten Veränderungen im Hinblick auf förderliche und hemmende innerschulische Gegebenheiten (vgl. Purkey & Smith 1983, 436ff.). Die Schulklimaforschung, bis dato der einzige Ansatz, der die einzelne Schule als Ganzes – als Umwelt von Lehr- und Lernprozessen – zum Forschungsgegenstand hatte, erhielt durch die vorliegende Betrachtungsweise neuen Auftrieb und damit die Gelegenheit zur eigenen Weiterentwicklung, wenn nicht gar Fortsetzung im Schulqualitätsansatz.
- 2) In der Lehrerforschung hinsichtlich Einstellungen und Bereitschaften zur Schulreform zeigten sich analoge Erkenntnisse, die bislang jedoch noch nicht hinreichend aufgegriffen wurden: Nicht der einzelne Lehrer bzw. die einzelne Lehrerin ist der entscheidende Träger bzw. die entscheidende Trägerin von Curriculumreform, sondern der Erfolg konstituiert sich erst im gemeinsamen Lehrerhandeln, z. B. bei kooperativer Unterrichtsplanung (vgl. Haller & Wolf 1973; Heipcke 1977; Scheffer 1977; Wolf & Klütsche 1977). Ähnliches lässt sich auch für die Herausforderungen von Schulverdrossenheits- und „Burn-out“-Phänomenen feststellen (vgl. Hoffmann 1988): dass nämlich die Kooperation im Kollegium ein Stützsystem für innerschulische Adaptationsprozesse ist. Auch die neuere Lehr- und Lernforschung untersuchte die Möglichkeiten des Lehrers bzw. der Lehrerin, die Unterrichtsqualität zu steigern (vgl. z. B. Weinert & Helmke 1987).
- 3) Die Akzentsetzung auf Schulqualität beinhaltet auch eine normative Komponente, ohne die die Gestaltung von Schule ziellos wäre. Schulqualität basiert auf lebenswerten Lehr- und Lernbedingungen, die sich in humanen Umgangsformen, sozialer Verantwortung, einer Schulgemeinschaft, einem positiven Selbstwertgefühl der Lernenden und Lehrenden und in entsprechenden psycho-hygienischen Verhältnissen, in Vertrauen und Abwesenheit von Anonymität, einer Interessen- und Bedürfnisorientierung,

Bemühungen um Kritikfähigkeit und Selbstständigkeit, geordneten erzieherischen Verhältnissen und vor allem in einer Bearbeitung so genannter „Schlüsselprobleme“ unserer Gesellschaft ausdrücken. Hier wird ein Bogen gespannt von den normativen Gesichtspunkten einer „guten“ Schule bei *Bohnsack* (1987) über *Klafki* (1991) bis hin zu *Hentigs* Leitvorstellungen einer Humanisierung der Schule (1987, 31f.). Diese Orientierungen sind nicht losgelöst zu betrachten von den neuen Herausforderungen der Schule durch den gesellschaftlichen Wandel (vgl. z. B. *Steffens & Bargel* 1992).

- 4) Schulentwicklungsvorhaben waren zunehmend auf die einzelne Schule ausgerichtet. Das betraf erstens vor allem systematische Ansätze schulpsychologischer Beratung von Schule (vgl. z. B. *Aurin et al.* 1977). Es galt zweitens für Ansätze der Organisationsentwicklung von Schule, die bei ihren Interventionen nicht auf einzelne Lehrerinnen und Lehrer, sondern auf ein ganzes Kollegium ausgerichtet sind (vgl. z. B. *Steuer* 1983; *Pieper* 1986; *Dalin et al.* 1990), und drittens für eine schulinterne Lehrerfortbildung, die das ganze Kollegium umspannt (vgl. z. B. *Edelhoff* 1988; *Schöning* 1988; *Wenzel et al.* 1990; *Greber et al.* 1991). Darüber hinaus sind Bemühungen um die Herstellung von Schulgemeinschaften zu erwähnen (vgl. z. B. *Lind* 1993; *Oser* 1993), die Entwicklung von Lehrerteams (vgl. u. a. *Schlömerkemper & Winkel* 1987) sowie die Kollegiumsarbeit am Schulkonzept (vgl. z. B. *Dörger* 1992; *Dörger & Steffens* 1986).

Der „Arbeitskreis Qualität von Schule“ traf sich in den Jahren 1985 bis 2001 zu insgesamt 17 dreitägigen Fachtagungen. Ein Großteil der Tagungsreferate und Tagungsergebnisse wurde in thematischen Bänden in einer gesonderten Veröffentlichungsreihe „Beiträge aus dem Arbeitskreis Qualität von Schule“ von *Tino Bargel und Ulrich Steffens* herausgegeben, zunächst vom Hessischen Institut für Bildungsplanung und Schulentwicklung (HIBS), später dann von der Nachfolgeeinrichtung Hessisches Landesinstitut für Pädagogik (HeLP). Vor allem aus den ersten vier Publikationen lässt sich auch der programmatische Akzent des Arbeitskreises ablesen, nämlich eine primär empirische Ausrichtung, Berücksichtigung quantitativer und qualitativer empirischer Verfahren, Interdisziplinarität der Ansätze (verschiedene theoretische Zugänge wurden unter einer Leitfrage gebündelt), Priorisierung der Handlungsebene Schule bei gleichzeitiger Beachtung der Unterrichts- und Systemebene, eine Auseinandersetzung mit den normativen Implikationen von Schulqualität und eine Perspektive von Schulentwicklung, bei der die Gestaltungsmöglichkeiten von Schulen im Zentrum stehen (verstärkte Selbstständigkeit der einzelnen Schule) (vgl. *Steffens & Bargel* 1987b, 1987c, 1987d und 1988).

Die Arbeiten des Arbeitskreises zogen zahlreiche Publikationen im deutschsprachigen Raum nach sich (vgl. v. a. *Tillmann* 1989; *Aurin* 1990; *Rolff* 1991) und führten zu einer breiten Diskussion in der Schulentwicklungsforschung und Schulentwicklung. Zahlreiche Fachtagungen belegen dies gleichermaßen wie Veranstaltungen und Seminare in der Lehrerfortbildung. Die Schulqualitätsdiskussion erreichte alsbald auch Österreich und die Schweiz; dort fanden ebenfalls Fachtagungen, Kongresse und Workshops statt (für die Schweiz siehe *Büeler* 1995; *Szaday* 1995 und *Szaday et al.* 1996; für Österreich siehe *Achs et al.* 1995; *Specht & Thonhauser* 1996).

## 2.2 Konsequenzen für die Bildungsplanung und Schulentwicklung

Die skizzierte neue Sichtweise von Schule als pädagogischer Gestaltungsebene hatte große Auswirkungen auf die weitere Schulentwicklung und Bildungsplanung. In einer ersten Rezeptionsphase Ende der 1980er und Anfang der 1990er Jahre fanden die Befunde zu den Gestaltungsbedingungen von Schule und Unterricht – unter Bezugnahme auf den „Arbeitskreis Qualität von Schule“ und auf anglo-amerikanische Untersuchungen (maßgeblich *Rutter et al. 1979*) – Eingang in die Lehrerfortbildung und in neuere Ansätze der Schulentwicklung (vgl. z. B. *Ermert 1986; Vierlinger 1989; DGBV 1998*). Eine maßgebliche Katalysatorfunktion hatte dabei das Institut für Schulentwicklung (IfS) in Dortmund mit seinen zahlreichen Veröffentlichungen zur Schulentwicklung (vgl. für die damalige Zeit u. a. *Philipp 1992; Holtappels 1995a; Rössner 1999*).

Trotz der zahlreichen Initiativen zur Schulentwicklung und Lehrerfortbildung fand der neue Ansatz zunächst keinen Niederschlag in konkreten Entwicklungsmaßnahmen der Bildungsverwaltung. Das änderte sich erst im Verlauf der 1990er Jahre. Vorreiterfunktionen übernahmen dabei die Bundesländer Bremen, Hessen und Nordrhein-Westfalen. Beispielsweise ergriff Hessen Anfang der 1990er Jahre eine Gesetzesinitiative, die die Rechte der Schule auf Selbstverwaltung und pädagogische Eigenverantwortung ausweitete. Dabei ging es um eine neue Balance zwischen den schulspezifischen Eigendefinitionen und dem übergreifenden Orientierungsrahmen, den der Staat sicherzustellen hat (vgl. *Frommelt & Steffens 1998, 30*).

Das hieß aber auch, wie es der Justiziar des Hessischen Kultusministeriums damals formulierte, „Abschied von der rechtlichen Fiktion der Gleichartigkeit öffentlicher Schulen zu nehmen. Sie waren es tatsächlich nie und dürfen es auch nicht sein, wenn sie auf die unterschiedlichen Anforderungen an Erziehung und Ausbildung ihres jeweiligen Umfeldes pädagogisch angemessen reagieren sollen. Geboten, aber unabdingbar ist ein Maßstab der Gleichwertigkeit“ (*Köller 1992; s. a. Köller 1991*). Mit dieser Initiative war zugleich auch eine Beendigung der Schulstrukturdebatte (integriertes versus gegliedertes Schulsystem) intendiert, die sich politisch nicht mehr durchhalten ließ.

Bereits 1992 wurde das entsprechende Hessische Schulgesetz in seiner ersten Fassung in Kraft gesetzt. Allerdings dauerte es noch mehrere Jahre (bis in die zweite Hälfte der 1990er Jahre hinein), bis entsprechende Strategien und Verfahren für die Selbstgestaltung und Eigenverantwortung der Schule entwickelt und bereitgestellt werden konnten. Dabei erwiesen sich meines Erachtens zwei Impulse als entscheidend:

Erstens waren dies angelsächsische Ansätze der Qualitätssicherung und Qualitätsentwicklung, die vor allem durch die Publikation von *Theo Liket*, „Freiheit und Verantwortung“ (*Liket 1993*), und durch entsprechend adaptierte Entwicklungsvorhaben in Nordrhein-Westfalen und Bremen in Deutschland Verbreitung fanden (vgl. dazu *Fleischer-Bickmann 1998; Maritzen 1998; Burkard & Kohlhoff 1998; Böhrs et al. 1998*). Als neue Steuerungsinstrumente wurden „schulexterne“ und „schulinterne Evaluation“ und „Schulprogramme“ propagiert, die in den Folgejahren in verschiedenen Bundeslän-

dern erprobt wurden (maßgeblich in Bremen, Hamburg, Nordrhein-Westfalen und Hessen).

Zweitens handelte es sich um Ansätze des „New Public Management“ (vgl. *Dubs* 1994a, 1994b), die sich nach *Dubs* (1996) aufgrund von drei Entwicklungslinien zu verbreiten begannen: der Diskussion um Schulqualität und Schulautonomie, der Reformen in der öffentlichen Verwaltung und der schlechten Haushaltslage der öffentlichen Hand. In diesem Zusammenhang fanden mit der Zeit Management-Konzepte wie das EFQM (1995) auch eine gewisse Verbreitung im schulischen Bereich.

Beide Impulse fielen auf einen fruchtbaren Boden der „Organisationsentwicklung“. Der US-amerikanische Ansatz wurde in den 1980er Jahren von *Richard Bessoth* in Deutschland bekannt gemacht und weiterentwickelt. Für seine maßgebliche Verbreitung und Konkretisierung an Schulen war das IfS in Dortmund mit entsprechenden Veröffentlichungen (vgl. z. B. *Rolff* 1993) und dem von *Hans-Günter Rolff* initiierten Netzwerk „Organisationsentwicklung“, das in späteren Jahren in „Schulentwicklung“ umbenannt wurde, verantwortlich.

Erst mit den neuen Strategien und Verfahren zur Qualitätssicherung und Qualitätsentwicklung von Schule erhielten Bestrebungen nach einer verstärkten Selbstständigkeit und Eigenverantwortung von Schulen ein solideres Fundament und erfuhren politisch-administrativ eine größere Zustimmung. Zahlreiche in jener Zeit erschienene Veröffentlichungen zur Schulautonomie sowie zur Qualitätssicherung und Qualitätsentwicklung sind Ausdruck dieser Entwicklung (vgl. z. B. *Avenarius et al.* 1998; *Buer et al.* 1997; *Daschner et al.* 1995; *Martini* 1997). Eine zentrale Bedeutung kam dabei der beginnenden Öffnung der deutschen Schulentwicklung und Schulentwicklungsforschung für Erfahrungen in anderen Ländern zu (vgl. z. B. *Haenisch* 1996a; *Altrichter* 1998; *Hopes* 1998).

Im Zuge dieser Umstrukturierung des Schulsystems mit der Verlagerung von Entscheidungskompetenzen von der System- auf die Schulebene („Deregulierung“) geriet auch zunehmend die Rolle der Schulaufsicht in die Diskussion. Obwohl sie bereits seit Jahren in der Kritik stand (siehe dazu Abschnitt 1.3; vgl. auch *Steffens & Bargel* 1998), hatte sich an ihrem Auftrag und ihren Funktionen bis dato nichts Wesentliches geändert. Erste größere Änderungsversuche waren im Rahmen des nordrhein-westfälischen Fortbildungssystems zu verzeichnen (vgl. dazu *Buchen et al.* 1995) und führten zu der Fortbildungsmaßnahme „Schulentwicklung und Schulaufsicht – Qualitätsentwicklung und Qualitätssicherung von Schule (QUESS)“, mit der im Schuljahr 1992/1993 begonnen wurde (vgl. *Kohlhoff* 1995). Trotz der erfolgversprechenden Erfahrungen mit QUESS (vgl. dazu *Burkhard & Kohlhoff* 1998) kam es zu keiner Institutionalisierung im Anschluss an das Modellprojekt.

Ähnlich erging es auch dem ambitionierten Vorhaben der Einführung einer Schulin-spektion in Bremen (vgl. dazu *Maritzen* 1998), mit deren Planung 1994 begonnen wurde und die bereits im ersten Jahr ihrer Erprobung (1996) abgebrochen wurde (zu den Gründen vgl. *Hans-Peter Füssel* im vorliegenden Band). Alle Bestrebungen, die Schulaufsicht weitergehend zu reformieren und sich auf neue Formen einer externen Evaluation einzulassen, waren damit erst einmal versandet. Die Schulaufsicht war

gleichzeitig bemüht – über ihre angestammten Verwaltungsaufgaben hinaus – die Ausübung ihrer Kontrollfunktion weiter zu reduzieren und zunehmend eine Beratungsfunktion für Schulen wahrzunehmen.

Obwohl es kein Pendant zur schulinternen Selbstregulierung in Form einer institutionalisierten externen Evaluation gab, fanden Ende der 1990er Jahre Maßnahmen zur Schulprogrammarbeit und schulinternen Evaluation – teilweise nur auf freiwilliger Basis – größere Verbreitung: zunächst in Hessen, Nordrhein-Westfalen und Bremen sowie in Berlin, Brandenburg und Niedersachsen, später dann auch in anderen Bundesländern. Jedoch wurde bereits damals kritisch angemerkt: „Inwieweit sich die damit verbundene Hoffnung erfüllen wird, dass alle an der schulischen Arbeit Beteiligten Schulprogramm und Evaluation als Hilfe für die Weiterentwicklung ihrer Schulen begreifen und nutzen werden, bleibt allerdings abzuwarten. Fraglich ist auch, inwieweit diese Instrumente geeignet sind, den durch die oben dargestellten Wirkungen von Modernisierungsprozessen und Bildungsexpansion auf den Schulen lastenden Problemdruck tatsächlich zu vermindern und die Akzeptanz des öffentlichen Schulwesens zu erhöhen. Vor dem Hintergrund restriktiverer Rahmenbedingungen und angesichts der Tatsache, dass die neuen Konzepte schulischer Qualitätssorge in Deutschland keineswegs unumstritten sind, können deren Erfolgchancen nicht von vornherein als günstig beurteilt werden. Es ist deshalb noch nicht genau absehbar, welche anderen deutschen Bundesländer den von Nordrhein-Westfalen, Bremen und Hessen eingeschlagenen Weg ebenfalls beschreiten und wie weit sie ihn mitgehen werden“ (Markstahler 1999, 32).

Dennoch schien der Trend zur verstärkten Selbstständigkeit der Einzelschule unaufhaltsam zu sein. Liket (1993, 81ff.) nennt vier Begründungszusammenhänge für Maßnahmen der „Qualitätsvorsorge“, die international bedeutsam sind:

- Eine offene demokratische Gesellschaft verlangt Rechenschaft über das Funktionieren von Bildungseinrichtungen;
- Entwicklungen in der Erziehungswissenschaft erlauben den Einsatz von Evaluationsmaßnahmen;
- wirtschaftliche Bedingungen führen zu der Frage nach der Effektivität von Bildungsinvestitionen;
- Profilierungen der einzelnen Schulen rufen die Frage nach ihren spezifischen Qualitäten hervor.

Tillmann (1995, 33ff.) sieht folgende vier Hauptlinien, die für die Diskussion maßgebend sind:

- den Diskurs über „Demokratie“ („Die größere Selbstständigkeit der einzelnen Schule ergibt sich aus der Forderung, daß innerhalb einer demokratischen Gesellschaft auch die einzelnen Institutionen Möglichkeiten zur demokratischen Mitgestaltung und Mitverantwortung eröffnen müssen“),
- den Diskurs über „pädagogische Entwicklung“ („Autonomie’ wird hier gefordert, damit ‚gute Pädagogik’ realisiert werden kann. Diese Argumentationslinie hat eine lange reformpädagogische Tradition (...); sie wird in sozialwissenschaftlich aufge-

klärter Weise von den Konzepten der ‚Organisationsentwicklung‘ ergänzt und weitergeführt“),

- den Diskurs über „Schule als Betrieb“ („Modernes Management bei ‚flachen‘ Hierarchien, flexibler Einsatz von Ressourcen, bessere Motivierung der Mitarbeiter, Erschließung zusätzlicher Finanzquellen (Drittmittel) werden hier als Stichworte genannt“),
- den Diskurs über „Sparpolitik und Arbeitsbelastung“ („Weil über knappe Mittel am besten direkt ‚vor Ort‘ entschieden werden sollte, ergibt sich daraus aus der Sicht der Kultusadministration ein weiteres Argument für eine höhere Selbständigkeit gerade auch bei der Verwaltung von Geldern und dem Einsatz von Lehrerstunden“).

### 3 Unterricht als Kernaufgabe der Schule

Die deutsche Schulqualitätsdiskussion war mit der Veröffentlichung der TIMSS-Ergebnisse im Jahr 1997 (TIMSS-2) in eine neue Phase getreten. Die Ergebnisse der „Third International Mathematics and Science Study“ zeigten, dass die Leistungen der deutschen Schülerinnen und Schüler im internationalen Vergleich nur mittelmäßig sind (*Baumert et al. 1997*). Die Ursachen für die zwischen den TIMSS-Teilnehmerstaaten festgestellten, teilweise dramatischen Leistungsunterschiede sahen die für den deutschen Untersuchungsteil verantwortlichen Wissenschaftlerinnen und Wissenschaftler vor allem in der die Schule tragenden Kultur, d. h. in der generellen Wertschätzung schulischen Lernens und der Bereitschaft zu Anstrengung sowie spezifischen Unterstützungsleistungen. Daneben seien die Ursachen in der Gestaltung des Fachunterrichts selbst zu suchen. Das aus deutscher Sicht düstere Bild hellte sich nicht auf, wenn man die Ergebnisse der entsprechenden Studie für die Sekundarstufe II (TIMSS-3) hinzunahm, die 1998 erschien (*Baumert et al. 1998*; s. a. 2000a und 2000b).

Trotz der an den TIMS-Studien wegen ihrer methodischen Anlage und der Art der Ergebnispräsentation geübten Kritik, auf die hier nicht eingegangen wird, werden diese Studien inzwischen mehrheitlich so wahrgenommen und verarbeitet, dass ihre Ergebnisse aus deutscher Sicht auf „erhebliche Defizite bei den erreichten Standards und zudem auch auf die mangelnde Vergleichbarkeit erworbener Zertifikate“ verweisen (*Klemm 1998, 275*).

Die Diskussion über das deutsche Schulsystem hat durch die TIMS-Studie einen entscheidenden An Schub erfahren: Die Leistungsfähigkeit des deutschen Schulsystems stand zur Debatte. Die öffentliche Diskussion zur TIMS-Studie hat sich der vergleichsweise schlechten Testergebnisse der deutschen Schülerinnen und Schüler angenommen. Der durch TIMSS in Deutschland ausgelöste Schock zeitigte mehrfache Wirkungen. In erster Linie rückte die Frage nach dem Stellenwert und der Qualität des Fachunterrichts wieder mehr ins Zentrum des schulpädagogischen Interesses. Erziehungswissenschaftler fragten sich, „ob nicht die fachlichen Leistungsziele bisher zu sehr vernachlässigt worden sind und ob nicht überhaupt die insbesondere an Grund- und Sekundarschulen entstandene neue Lernkultur – bei allem reformpädagogischen Charme – in ihrer Zeitnutzung zu wenig effektiv und in ihrem kognitiven Anforderungsniveau zu anspruchsarm und nivellierend ist“, so der langjährige Beobachter der Schulentwicklung *Rudolf Messner* (1998, 102). Diese Diskussion passt zu der Kritik seitens der Wirtschaft und der Hochschulen an den Schülerleistungen. Die Kritik lau-

tet, dass viele Schülerinnen und Schüler nicht mehr die grundlegenden Kulturtechniken beherrschten und dass von den Abiturientinnen und Abiturienten viele nicht mehr studierfähig seien. Die Wettbewerbsfähigkeit einer ganzen Nation stehe zur Disposition – so äußerte sich beispielsweise der ehemalige Bundesminister für Bildung, Forschung und Technologie.

Zudem löste der "TIMSS-Schock" in der deutschen Schuldebatte eine "empirische Wende in der Bildungspolitik und Bildungsplanung" aus. Es war der ehemalige Vorsitzende der Amtschefkommission "Qualitätssicherung" der Kultusministerkonferenz (KMK), *Hermann Lange*, der diese Formel unter Bezugnahme auf *Heinrich Roths* "empirische Wende" in der Erziehungswissenschaft prägte (*Lange* 1999). Nicht bloße Annahmen über vermeintliche Wirkungen, sondern empirische Befunde sollten nunmehr die Grundlage bildungspolitischer Entscheidungen sein.

Geht man von diesem Sachverhalt aus, so hatten die TIMS-Studien einen nachhaltigen Effekt auf die Systemsteuerung von Schule und Unterricht. Ihre bildungspolitischen Konsequenzen sind deutlich zu erkennen. So befasste sich die Konferenz der Kultusminister bereits wenige Monate nach der Veröffentlichung der Ergebnisse von TIMSS-2 auf ihrer Konstanzer Konferenz im Oktober 1997 mit diesem Thema und forderte die Entwicklung einer „neuen Kultur der Anstrengung“ an Schulen. Daneben seien folgende Handlungsfelder entscheidend: Stärkung der Wertschätzung des Lernens in Deutschland, Hebung des Stellenwerts der mathematisch-naturwissenschaftlichen Disziplinen in der Gesellschaft und Verbesserung der Qualität und Organisation des entsprechenden Fachunterrichts an Schulen. Von mindestens ebenso großer Bedeutung war der Beschluss der Konferenz, länderübergreifende Vergleichsuntersuchungen durchzuführen. Damit sollten die Qualität schulischer Bildung gesichert sowie die Gleichwertigkeit der schulischen Ausbildung, die Vergleichbarkeit der Schulabschlüsse und die Durchlässigkeit des Bildungssystems innerhalb der Bundesrepublik Deutschland gewährleistet werden. Diese länderübergreifenden Vergleichsuntersuchungen sollten nach einem weiteren Beschluss der Kultusministerkonferenz im Rahmen des TIMSS-Nachfolgeprojekts PISA realisiert werden.<sup>2</sup>

Das ist vermutlich die größte Resonanz, die in Deutschland jemals von einer empirischen pädagogischen Studie ausging – sie war größer als die der „Rutter-Studie“ (vgl. *Rutter et al.* 1980) und der Konstanzer Schulforschung (vgl. v. a. *Fend* 1982). Man kann deshalb aus heutiger Sicht die Vermutung äußern, dass sich die Geschichte der Schul- und Unterrichtsentwicklung in der Bundesrepublik Deutschland in eine Zeit vor und nach TIMSS einteilen lässt.

Auch wenn die Befunde aus der TIMS-Studie keine wesentlich neuen Erkenntnisse über den Unterricht an deutschen Schulen zu Tage gefördert hatten (vgl. z. B. bereits *Hage et al.* 1985), so leitete die anschließende Diskussion doch eine weitere nachhalti-

---

<sup>2</sup> Mit Blick auf diese KMK-Beschlüsse könnte man freilich kritisch nachfragen, ob das Maßnahmenpaket geeignet war, die handfesten Leistungsdefizite der deutschen Schülerinnen und Schüler merklich zu reduzieren: „Die sich hinter den Ergebnissen von TIMSS verbergenden strukturellen Probleme des Schulsystems (...) könnten in der komparatistischen Perspektive verloren gehen: Setzte man den zeitlichen Aufwand für Unterricht und das Nicht-Wissen bei Schülern in Relation, so ergibt sich ‚führwahr ein Skandal‘, wie Heinrich meint (1998, S. 49). Er wundert sich, dass die ‚in den Ergebnissen sich manifestierende strukturelle Irrationalität des schulischen Betriebes‘ (ebd.) in der TIMSS-Rezeption nicht hinreichend kritisiert wird“ (*Böttcher* 2002, 19).

ge Wende in der Bildungsplanung und Schulentwicklung ein, und zwar in zweifacher Hinsicht: hinsichtlich Unterrichtsentwicklung einerseits und hinsichtlich einer systematischen Erfassung von Lernleistungen andererseits.

### 3.1 Entwicklung von Unterrichtsqualität

TIMSS hat nicht nur eine breite Diskussion über die Leistungsfähigkeit des deutschen Schulsystems ausgelöst, sondern darüber hinaus den Blick für die Notwendigkeit von Unterrichtsentwicklung geschärft und damit die Unterrichtsebene als eine von mehreren wichtigen Handlungsebenen in den Mittelpunkt der weiteren Schulentwicklungsanstrengungen gerückt. TIMSS rief eine breite öffentliche Kritik am vorherrschenden lehrerzentrierten und fragengeleiteten Unterricht in Deutschland hervor, der überwiegend inhaltsbezogen, auf Erwerb und Reproduktion von Wissen ausgerichtet ist und sich damit merklich von Unterrichtsskripten z. B. aus japanischen Schulen unterscheidet, die demgegenüber Prozess- und Verständnisorientierung intendieren (vgl. *Baumert et al. 1997; 2000a; 2000b*). Entsprechende Befunde erbrachten auch Untersuchungen des Max-Planck-Instituts für Psychologische Forschung in München (*Weinert & Helmke 1997*). Von diesen Erkenntnissen ausgehend wurde vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) ein Gutachten zur „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ in Auftrag gegeben, das die weitere Unterrichtsentwicklung in Deutschland in den naturwissenschaftlichen Fächern und in Mathematik maßgeblich beeinflusst hat (BLK 1997). Darauf aufbauend wurde ein umfangreiches bundesweites Modellversuchsprogramm zur fachbezogenen Weiterentwicklung von Schulen auf der Unterrichtsebene aufgelegt („SINUS“; vgl. *www.sinus-transfer.de*), welches mittlerweile in der Lehrerfortbildung in fast allen Bundesländern Verbreitung gefunden hat. Hatten bis dahin Programme zur organisatorisch und pädagogisch orientierten Schulentwicklung (maßgeblich Schulprogrammentwicklung) Hochkonjunktur (vgl. Abschnitt 1.2), so erhielten diese durch SINUS ein Korrektiv und ein „Pendant“ auf der Unterrichtsebene.

Im gleichen Zeitraum, in dem die TIMSS-Ergebnisse Verbreitung fanden, sind auch Weiterentwicklungen in der Lehr-Lernforschung zu beobachten (vgl. z. B. *Weinert 1996; Einsiedler 1997; Ditton 2002; Klieme 2002; Doll & Prenzel 2004; Helmke 2005*). Diese sind vor allem auf Aufarbeitungen der Studien des Max-Planck-Instituts für Psychologische Forschung (Classroom Environment Study, LOGIK, SCHOLASTIK), des Max-Planck-Instituts für Bildungsforschung in Berlin (BIJU und TIMSS) sowie angelsächsischer Befunde zurückzuführen, und zwar insbesondere im Hinblick auf eine Abkehr vom lehrerzentrierten, kleinschrittigen und fragengeleiteten Unterricht und eine Hinwendung zu Formen eines problemlösenden, fächerübergreifenden und anwendungsorientierten Lernens (siehe dazu z. B. *Baumert 1998; Weinert 1998*).

Im Zusammenhang mit den TIMSS-Videodaten wurden auch neue Forschungsprojekte initiiert („TIMSS-Video-Repeat“), die sich von den früheren Arbeiten zur Unterrichtsqualität dadurch unterschieden, dass sie nicht nach Merkmalen des „guten Unterrichts“ bzw. des „guten Lehrers“ suchten und weniger die „Oberflächenstruktur“ von Unterrichtsprozessen zum Gegenstand hatten: Stattdessen waren sie in einer konstruktivistischen Perspektive vorrangig auf kognitive Entwicklungen und sinnverstehendes

Lernen ausgerichtet („Tiefenstruktur“) (vgl. dazu z. B. *Klieme* 2003 und *Reusser et al.* 2006).

Für die Bedeutung der Unterrichtsebene sprechen auch Forschungsbefunde, die höhere Zusammenhänge zwischen schulischen Wirkungen und schulischen Bedingungen dann feststellen, wenn die Bedingungsvariablen direkt die Lehr- und Lernprozesse betreffen:

„Sowohl die Forschung zu schulischer Qualität bzw. Effektivität als auch Bemühungen um eine Qualitätskontrolle und -sicherung im Schulwesen müssen sich, wenn sie erfolgreich sein sollen, primär auf die Ebene des Unterrichts beziehen – organisatorisch betrachtet also in der Regel auf die Ebene der einzelnen Schulklasse (vgl. *Ditton & Kreckler* 1995; *Ditton* 1997). Diese Sicht erfordert eine Abkehr von pauschalen Klassifikationen guter Schulen sowie veränderte Forschungs- und Evaluationsdesigns – ohne dass damit nun in einer überzogenen Kehrtwende gleich das Kind mit dem Bade ausgeschüttet werden soll. Guter Unterricht braucht eine unterstützende Umgebung, d. h. fördernde Bedingungen auf der Ebene der Schule sowie darüber hinaus im administrativen und sozial-regionalen Kontext. Dies verweist zugleich darauf, daß Erhebungen zu Faktoren der Effektivität und gleichfalls die Entwicklung von Evaluationssystemen unter Bezug auf Unterrichtssituationen erfolgen sollten“ (*Ditton* 2000a, 73).

### **3.2 Die systematische Erfassung von Lernleistungen**

Im Zuge der ausgerufenen empirischen Wende in der Bildungspolitik und Bildungsplanung begannen einige Bundesländer, standardisierte Lernstandserhebungen einzuführen. Eine Vorreiterrolle übernahm dabei Hamburg, das bereits 1995 eine wissenschaftliche Untersuchung zu Aspekten der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an allen Hamburger Schulen hinsichtlich der Kenntnisse, Fertigkeiten und Fähigkeiten in den Bereichen Sprachverständnis, Leseverständnis, Rechtschreibwissen, Informationsentnahme aus Karten/Tabellen/Diagrammen und Mathematik in Auftrag gegeben hatte (vgl. *Lehmann et al.* 1996). Andere Bundesländer folgten, allerdings nur im Bereich des Fachs Mathematik, nämlich Brandenburg mit „QuaSUm“ (vgl. *Lehmann et al.* 2000; *Peek* 2004) und Rheinland-Pfalz mit „MARKUS“ (*Helmke & Jäger* 2002). Schon von Anfang an waren diese Vorhaben von Kritik begleitet:

So wurde moniert, dass in den Lernstandserhebungen unter Schulqualität in erster Linie gute fachliche Lernleistungen – insbesondere in den traditionellen Hauptfächern – verstanden würden. Eine gute Schule sei nach diesem Verständnis eine Schule mit guten Testleistungen vor allem in den Fremdsprachen, in Mathematik und Deutsch. Mit einer solchen Orientierung würden die Erträge von Schule auf ein enges Spektrum fachlicher Lernleistungen der Schülerinnen und Schüler reduziert. Es wurde nicht nur ein erweitertes Verständnis fachlichen Lernens vermisst, sondern das Fehlen des gesamten Bereichs der überfachlichen Kompetenzen (teilweise als „Schlüsselqualifikationen“ bezeichnet) bemängelt. Durch diese Einengung werde die Vielfalt schulischer Ziele ignoriert. Schule sei demgegenüber gefordert, viele Aspekte des Schülerverhaltens und der Schülerhaltungen zu beeinflussen. Darüber hinaus wurde kritisiert, dass eine Prozess-

orientierung bei der Frage nach der Schulqualität so gut wie gar nicht berücksichtigt worden sei; die Bedingungen, unter denen Schüler und Schülerinnen lernen, gerieten so nicht in den Blick.

Hieraus wurde gefolgert, dass Ansätze einer ausschließlichen und restriktiven „Output“-Steuerung zu kurz griffen. Sie seien in einigen Kreisen zwar sehr populär, weil sie vor allem vermeintlich klare Messwerte lieferten; sie seien jedoch in pädagogischer Hinsicht nicht tragfähig. So werde auch in Expertenkreisen problematisiert, dass vor allem externe Tests, die zur Kontrolle schulischer Arbeit eingesetzt würden, innovationshemmend wirkten. Wenn der Unterricht sich primär nach ihnen ausrichte, könnten sich neue, weiterführende Unterrichtskonzepte kaum entfalten (*Black & Atkin 1996*). Auf einem internationalen Symposium kamen die versammelten Experten verschiedener Länder zu folgender Bewertung (vgl. *Black & Atkin 1996*):

Erstens müssten Leistungsbeurteilungen so angelegt sein, dass sie nicht nur verschiedene Kompetenzen der Schülerinnen und Schüler erfassen, sondern auch differente Lernkulturen und Lernstile berücksichtigen. Sie sollten zweitens kohärent in dem Sinne sein, dass formative, diagnostische Verfahren mit summativen Verfahren kompatibel sind. Dieses Problem sei bislang (bei den diskutierten Modellen) noch nicht zufriedenstellend gelöst worden. Als dritter Punkt wurde die Notwendigkeit intensiver Lehrerfortbildung betont, ohne die eine erfolgreiche Handhabung dieser neuartigen Verfahren der Leistungsbeurteilung selbst dann nicht gelingen könne, wenn Lehrerinnen und Lehrern wie in Frankreich ein elaboriertes Testinstrumentarium zur Verfügung stehe. Schließlich dürfe nicht übersehen werden: Auch die besten Verfahren der Leistungsbeurteilung könnten von sich aus keine Verbesserung von Lernresultaten bewirken. Sie könnten lediglich auf Probleme aufmerksam machen und Anstöße zu deren Bewältigung geben.

Lernstandserhebungen, so wurde argumentiert, seien somit nur dann von Nutzen, wenn mit ihnen entsprechende Maßnahmen zur Sicherung und Verbesserung von Schule und Unterricht einhergingen. Qualitätsevaluation und Qualitätsverbesserung seien eng miteinander zu verschränken; die beste Qualitätssicherung liege in der Qualitätsentwicklung. Nicht die ständige Prüfung des Schulsystems, sondern die stetigen Verbesserungsbemühungen hätten im Mittelpunkt der Bestrebungen zu stehen.

Dementsprechend sollte sich auch der Auftrag der Unterstützungssysteme zur Beratung und Hilfe einerseits sowie zur Sicherung der Standards und der Gleichwertigkeit andererseits auf ein Konzept konzentrieren, das folgenden Gesichtspunkten Rechnung trage (*Arbeitsstelle 1998, 23*):

- „Es muss Lehrerinnen und Lehrern Aufschluss geben über *nachhaltige* Ergebnisse ihres Unterrichts.
- Es muss Lehrerinnen und Lehrer anregen, über Bedingungen und Wege gelingenden Unterrichts (gemeinsam) nachzudenken und konkrete Entwicklungsschritte für die eigene Schule zu beschreiben und zu erproben (Qualitätsevaluation als Teil von Qualitätsentwicklung).

- Es muss es den Schulen ermöglichen, ihre Leistungen durch Vergleich mit anderen Schulen zu verorten und dies bezogen auf Schulumfeld und Leistungsvoraussetzungen angemessen zu interpretieren.
- Es muss Schulaufsicht und die übrigen Unterstützungssysteme instand setzen, einen fundierten – nicht bloß auf punktuellen Erfahrungen und Einschätzungen beruhenden – Einblick in die Situation der Schulen der Region zu gewinnen, um gemeinsam mit den anderen Unterstützungssystemen ein Konzept der Unterstützung von Schulen zu planen.
- Es muss unter ‚Normalbedingungen‘ praktikabel und in seinen Zielsetzungen und Verfahren nachvollziehbar sein – bei übermäßig komplizierten Verfahrensweisen wird der Wunsch nach einfachen und klaren Lösungen (zentrale Abschlussprüfungen, Leistungstests in wenigen Fächern mit Ranking) kaum abzuwenden sein.“

Dabei sei an eine interne Evaluation durch innerschulisch entwickelte Vergleichsarbeiten genauso zu denken wie an extern entwickelte Testaufgaben für verschiedene Fächer und Jahrgangsstufen. In diesen Überlegungen sind neue Formen der externen Evaluation angedacht, wobei prinzipiell zwischen zwei Grundmustern unterschieden werden kann:

Das eine Muster geht primär von verfügbaren Schuldaten aus, die im Rahmen der schuleigenen Entwicklungsarbeit gewonnen werden (Schulprogramme, schulinterne Curriculumentwicklung, interne Evaluation mit Selbst- und Fremdbeteiligung sowie weitere Dokumente der Schulen). Dieses Muster nimmt Bezug auf Ergebnisse aus schuleigenen Analyse-, Reflexions- und Maßnahmenprozessen. Durch eine solche Bezugnahme wäre eine Entwicklung „von unten“ unter Berücksichtigung der Problemstellungen und des Handlungsbedarfs der einzelnen Schule sichergestellt. Unterstützung und Hilfe könnte darauf aufbauen und zielsicher zum Einsatz kommen. Schulaufsicht müsste dafür Sorge tragen, dass Schulen den vorgegebenen Auftrag (ausgedrückt z. B. im Hessischen Schulgesetz in §§ 2 und 3, in der Stundentafel, den Lehrplänen und in weiteren Regularien) angemessen erfüllen. Dabei hat sie vor allem die Aufgabe einer Sicherung der Standards und der Gleichwertigkeit (nicht Gleichartigkeit) zu erfüllen.

Dieses Vorgehen könnte jedoch mit einem prinzipiellen Problem behaftet sein, das gerade im Zuge der öffentlichen Diskussion um die TIMSS-Ergebnisse eine größere Beachtung fand: So kann davon ausgegangen werden, dass die von den Schulen zur Verfügung gestellten Daten zwar die spezifischen Interessen und Anliegen einer Schule zum Ausdruck bringen, aber nicht unbedingt geeignet sind, über die Vergleichbarkeit ihrer Leistungen Auskunft zu geben. Da die schulischen Daten quasi ‚selbstreferentiell‘ erzeugt sind, können sie insbesondere unter Konkurrenz- und Legitimationsgesichtspunkten ‚geschönt‘ sein (Belege dafür liefern Erfahrungen mit den Schulreformbemühungen in den 1970er und 1980er Jahren).

An dieser Stelle kommt das zweite Grundmuster von Qualitätssicherung zum Tragen, welches das genannte Problem durch die Vorgabe externer Leistungsstandards aufzufangen versucht und dabei maßgeblich auf die Erfassung von Schülerleistungen setzt (vgl. dazu *Lehmann 1997; Klieme et al 2000; Baumert 2001; Weinert 2001,*

*Pekrun 2002; Terhart 2002; Markstahler 2003*). Für einen solchen Weg hatte sich im Anschluss an TIMSS die KMK entschieden und die Beteiligung an weiteren internationalen Vergleichsuntersuchungen (zunächst PISA, dann PIRLS und später auch DESI) sowie an nationalen Vergleichen zwischen den Bundesländern beschlossen. Einzelne Bundesländer, allen voran Hamburg, hatten zwischenzeitlich ebenfalls Leistungstests oder Vergleichsarbeiten in verschiedenen Fächern eingeführt, um dadurch Aufschluss über die Leistungsfähigkeit ihres Schulsystems zu gewinnen. Vor dem Hintergrund dieser Entwicklung wurde 1985 auf Initiative von Rainer Peek und Ulrich Steffens ein bundesweites Netzwerk, „EMSE“, eingerichtet, das zum Themenzusammenhang der empirisch orientierten Schulentwicklung die Fachleute in den Ministerien bzw. Landesinstituten aller Bundesländer zwei Mal im Jahr zu Tagungen versammelt. Der Schwerpunkt des Erfahrungsaustauschs lag bisher im Bereich standardisierter Lernstandserhebungen, die mittlerweile in fast allen Bundesländern in jährlicher Folge durchgeführt werden. Das EMSE-Netzwerk hat bislang fünf Tagungen ausgerichtet und ein „Positionspapier“ zu Lernstandserhebungen verabschiedet (vgl. *Peek et al. 2006*).

Während das Problem des erstgenannten Evaluationsmusters in der mangelnden Vergleichbarkeit liegen könnte, werden die Probleme des zweiten Evaluationsmusters in der mangelnden Ergiebigkeit hinsichtlich der schulpraktischen Konsequenzen vermutet. Die größte Herausforderung für eine Qualitätsentwicklung und Qualitätssicherung von Schule und Unterricht, so der Tenor der Schulentwicklungsdiskussion im deutschsprachigen Raum, liegt aufgrund der genannten Problemstellungen in der Frage nach erfolversprechenden und pädagogisch sinnvollen externen Evaluationsverfahren sowie nach einem geeigneten Zusammenspiel der unterschiedlichen Vorgehensweisen.

#### **4 Die Wiederentdeckung der Systemebene**

Im Zuge der „empirischen Wende“ in der Bildungspolitik und Bildungsplanung hat sich in den einzelnen Bundesländern ein Spektrum an empirisch orientierten Vorhaben zur Qualitätssicherung und Qualitätsentwicklung entfaltet bzw. verstetigt. Diesen Vorhaben liegt das Anliegen zugrunde, neben „Systemwissen“ im Interesse einer fundierten, rationalen und wirksamen Bildungsplanung Handlungswissen für die Schul- und Unterrichtsgestaltung unter Berücksichtigung empirischer Verfahren zu gewinnen.

Im Mittelpunkt der aktuellen Aktivitäten stehen Verfahren der Standardsetzung, die sich bundesweit in der Formulierung von Bildungsstandards (vgl. *Klieme et al. 2003*) und in den einzelnen Bundesländern in der Formulierung von Kompetenzerwartungen oder verbindlichen Anforderungen in (Kern-)Lehrplänen für die Fächer zeigen. Daran gekoppelt spielen in der Diskussion Verfahren der Standardüberprüfung eine große Rolle. In diesem Zusammenhang hat die Kultusministerkonferenz das „Institut zur Qualitätsentwicklung im Bildungswesen (IQB)“ eingerichtet (unter Leitung von Olaf Köller), dessen maßgebliche Aufgaben die Weiterentwicklung, Operationalisierung, Normierung und Überprüfung der Bildungsstandards sind. Darüber hinaus haben sich bundesweit und in den einzelnen Bundesländern Lernstandserhebungen (Vergleichsarbeiten bzw. Orientierungsarbeiten) etabliert, welche die methodischen Prinzipien

(standardisierte Testverfahren) aus Schulleistungsstudien bzw. „large scale assessments“ wie PISA anwenden. Im Unterschied zu deren primärer Zielsetzung des Systemmonitorings zielen sie aber vornehmlich auf Qualitätsentwicklung „vor Ort“.

Vergegenwärtigt man sich die grundsätzlichen Vorbehalte, die in Pädagogenkreisen immer wieder gegenüber empirischen Verfahren geäußert wurden (z. B. „pädagogisches Handeln lässt sich nicht messen“), so überrascht die zunehmende Akzeptanz von Vergleichsuntersuchungen und standardisierten Lernstandserhebungen. Soweit Bedenken bestehen, sind sie weniger prinzipieller Natur. Sie beziehen sich eher auf die mit solchen Erhebungen verbundenen Belastungen für das Lehrpersonal, auf die unzweckmäßige Verwendung der Ergebnisse in Form öffentlicher Rankings und vor allem auf ein Ungleichgewicht zwischen „Messen und Entwickeln“, das deshalb besonders auffällt, weil in einigen Bundesländern gleichzeitig Lehrerfortbildungs- und Schulberatungssysteme zurückgebaut wurden. Zudem ist zu beobachten, dass Bildungsplanung, Bildungsverwaltung und Unterstützungssysteme nur begrenzte Kapazitäten für die Verarbeitung der Forschungsergebnisse besitzen (vgl. *Tillmann* 2001, 11).

Ein solcher Einstellungswandel in Schulwissenschaft und Schulpraxis ist ohne die internationalen und nationalen Vergleichsuntersuchungen kaum vorstellbar: Vor allem die Schulvergleichsstudien TIMSS und PISA haben dem deutschen Bildungssystem einen Spiegel vorgehalten. Durch den von ihnen erbrachten unbestreitbaren empirischen Nachweis von Schwachstellen im Bildungssystem ist – erstmals seit den Bildungsreformen der 1960er Jahre – Bildung wieder zu einem zentralen gesellschaftspolitischen Thema in Deutschland geworden. Infolge der kumulativen Wirkung von TIMSS und PISA scheint sich nunmehr eine bildungspolitische und -planerische Wende abzuzeichnen. Die Befunde dieser und weiterer Studien verweisen auf gravierende strukturelle Probleme des deutschen Bildungswesens. PISA hat also nicht nur primäre Effekte erzielt, indem die Studie Prozess- und Ertragsindikatoren zum Zwecke der Verbesserung von Schulsystemen zur Verfügung gestellt hat (*Deutsches PISA Konsortium* 2001), sondern hat darüber hinaus sekundäre Effekte von großer Tragweite produziert. Diese sind vor allem in einem Bewusstseinswandel hinsichtlich der Notwendigkeit grundlegender struktureller Reformen, und zwar insbesondere in der Hinwendung zu einer evaluations- und wirkungsorientierten sowie an Standards orientierten Steuerung unseres Schulwesens zu sehen. Der Wandel vollzieht sich dabei in allen bildungspolitischen Lagern, gesellschaftlichen Bezugsgruppen sowie auf allen Handlungsebenen des Schulsystems.

Im Kern betrifft dieser Wandel die Unterschiedlichkeit einzelner Schulen, auch und gerade innerhalb ein und derselben Schulform – also die Güte von Schulgestalten – und die unterschiedliche Realisierung schulaufsichtlicher Vorgaben. Die so genannte Input-Steuerung war anerkanntermaßen an ihre Leistungsgrenze gestoßen: Schulen realisierten die rechtlichen Vorgaben auf jeweils eigene Weise, ohne dass die Prozess-Steuerung in Form schulaufsichtlicher Kontrolle eine Chance sah, zu intervenieren. Die Einsicht in die begrenzte Steuerbarkeit schulischer Gestaltungsprozesse durch makro-organisatorische Vorstrukturierungen (z. B. in Form von Lehrplänen und Ausführungsvorschriften für die Arbeits- und Lernorganisation) einerseits und die Wahr-

nehmung der geringen Wirksamkeit schulaufsichtlichen Vollzugs- und Kontrollhandelns andererseits führten – gelinde gesagt – zu dem Verdacht (im Sinne einer generalisierenden Schlussfolgerung), dass klassische Steuerungsinstrumente zumindest in der derzeitigen Ausprägung und Qualität nicht mehr hinreichend greifen. Im Grunde befand sich damit die Schulsystemsteuerung mitten in einer Legitimationskrise.

Die Objektivierung der schulisch-unterrichtlichen Arbeit war unausweichlich: „Im neuen Paradigma wird nicht einfach der Zuteilung solcher ‚Inputs‘ vertraut. Der Erfolg des Systems Schule wird an seinen Resultaten, vor allem den Lernergebnissen der Schüler, gemessen“ (Messner 2004). Die bereits in den 1990er Jahren propagierte Hinwendung zur Outputsteuerung wurde nunmehr – vor dem Hintergrund der ernüchternden Ergebnisse aus den Vergleichsuntersuchungen und der Sinnhaftigkeit externer Evaluation – vollzogen. Dabei ist eine ökonomische Ausrichtung der neueren Reformvorstellungen unverkennbar (vgl. Böttcher 2002).

Auf der Suche nach Alternativen setzte daraufhin ein Öffnungsprozess ein, der bisher vernachlässigten Referenzsystemen und Erfahrungsbereichen zu einer größeren Beachtung verhalf. Dadurch gewannen die empirischen Wissenschaften einerseits und die internationalen Entwicklungen andererseits an Bedeutung für bildungspolitisches und bildungsplanerisches Handeln. Anders als noch wenige Jahre zuvor fanden nun auch Ergebnisse aus internationalen Vergleichsuntersuchungen Beachtung (für einen Forschungsüberblick vgl. Scheerens et al. 2003; Scheerens & Bosker 2005). Das wiederholt festgestellte mittelmäßige Abschneiden deutscher Schülerinnen und Schüler bei internationalen Vergleichen fiel somit erstmals und nachhaltig auf einen fruchtbaren Nährboden. Hinzu kam übrigens, dass die Befunde aus der TIMS-Studie auch öffentlichkeitswirksam in gelungener Weise präsentiert und publiziert wurden.

Im Zuge dieser Entwicklung gerieten auch die Steuerungssysteme anderer Staaten in den Blick (vgl. Döbert et al. 2003). Dies waren zum einen alternative Instrumente der Produktsteuerung durch Leistungstests, für die es in Deutschland bislang keine Tradition gab: So setzen Schulsysteme anderer Staaten z. T. schon seit vielen Jahren standardisierte Tests ein, um Wissen über die Leistungsfähigkeit des Schulwesens insgesamt oder einzelner Schulen zu erhalten („Outputsteuerung“). Zum anderen ging es um die Frage nach einer wirksameren Steuerung schulischer Prozesse („Prozesssteuerung“) und ihrer systematischen Verbindung mit Output- und Inputvariablen durch Schulinspektion als eine Form externer Evaluation (vgl. dazu Posch & Altrichter 1997; Posch 1999; Ackeren 2003 und Kotthoff 2003).

„Der Blick über die nationalen Grenzen zeigt, dass in vielen Schulsystemen Europas die Einzelschulen in den letzten Jahren immer stärker dazu angehalten werden, Zeugnis über ihre Leistungen abzulegen. Zentral organisierte Leistungsmessungen und Evaluationsprogramme spielen dabei eine wichtige Rolle. Mit ihrer Hilfe werden Auskünfte über die Qualität der Einzelschule eingeholt und öffentlich oder/und vertraulich verschiedenen Zielgruppen zur Verfügung gestellt“ (Ackeren 2003, 12). „Der Prozess der ‚Autonomisierung‘ der Schule, so lehrt ein Blick auf die Schulsysteme anderer Staaten, scheint zumindest mittelfristig die Entwicklung von evaluativen Gegengewichten zu erzwingen, um Entwicklungen kontrollierbar und steuerbar zu halten. Insofern erscheint

die schulische Evaluation als der ‚Preis‘ für eine erweiterte Autonomie der Schule“ (Kotthoff 2003, 12).

Die aktuellen Entwicklungen in den Bundesländern lassen unschwer erkennen, dass viele Länder in der Zwischenzeit auf eine „Doppelstrategie“ der Qualitätsentwicklung und Qualitätssicherung setzen: Schulinterne Maßnahmen in Form von Schulprogrammen und schulinterner Evaluation werden durch schulexterne Maßnahmen in Form von standardisierten Lernstandserhebungen, zentralen Abschlussprüfungen und Schulinspektionen bzw. Qualitätsanalysen flankiert. Damit wird zugleich das Beziehungsgefüge zwischen dezentraler Steuerung (Schulebene) und zentraler Steuerung (Systemebene) neu justiert. Bei dieser Neubestimmung erhält die Systemebene wieder einen konstitutiven Stellenwert im Geflecht der Handlungsebenen.

Zwischen der von *Heinrich Roth* ausgerufenen empirischen Wende in der Pädagogik, der von Staatsrat *Hermann Lange* proklamierten empirischen Wende in der Bildungspolitik und Bildungsplanung und der allem Anschein nach nunmehr tatsächlich vollzogenen Wende liegen mehrere Jahrzehnte. In dieser Zwischenzeit ist vieles geschehen, was als Nährboden bzw. Voraussetzung für die aktuellen Entwicklungen in der Bildungspolitik und Bildungsplanung anzusehen ist.

Obwohl die empirische Bildungsforschung bereits vor über drei Jahrzehnten Einlass in bildungsplanerische Handlungszusammenhänge erhalten hatte, blieben empirische Befunde als Planungsparameter für bildungspolitische und bildungsplanerische Entscheidungen zunächst weitgehend randständig. So hatten bereits Ende der 1970er Jahre große empirische Schulforschungsprojekte relevante Erkenntnisse für die Bildungsplanung und Schulgestaltung bereitgestellt – etwa im Rahmen des Sonderforschungsbereichs Bildungsforschung der Deutschen Forschungsgemeinschaft (DFG) an der Universität Konstanz oder des Dortmunder Instituts für Schulentwicklungsforschung, aber auch der Max-Planck-Institute für Bildungsforschung in Berlin und für Psychologische Forschung in München. Rückblickend betrachtet lässt sich feststellen, dass diese Forschungen zwar zunächst keine unmittelbare handlungspraktische Bedeutung für die Bildungsverwaltung erlangten, wohl aber einen gewissen „Depot-Effekt“ auslösten, der vermutlich erst seit Mitte der 1990er Jahre Wirkung zeitigt. Mittlerweile ist festzustellen: „Die Evaluierbarkeit des Schulwesens auf allen Ebenen wird kaum mehr grundsätzlich in Frage gestellt“ (*Leschinsky & Cortina* 2003, 46).

Die TIMSS-Befunde führten – abgesehen von der skizzierten Unterrichtsentwicklung (vgl. Abschnitt 3.1) – zum Einstieg in eine systematische Erfassung von Lernleistungen in Deutschland (vgl. Abschnitt 3.2). In konsequenter Ergänzung dazu stellt PISA einen weiteren Meilenstein der deutschen Bildungsplanung dar: *Pekrun* (2002, 112) charakterisiert PISA als ein outputorientiertes deskriptives Systemmonitoring zu bestimmten Aspekten des deutschen Schulsystems. Insofern übte PISA eine Katalysatorfunktion bei der Einführung eines systematischen Konzepts von Qualitätsüberprüfung aus. Das Beobachten und Vermessen des Schulsystems entwickelte sich seit PISA zu einer wesentlichen neuen Aufgabe der Bildungsadministration. Der Bildungsforschung kommt dabei die Aufgabe zu, „die Bildungswirklichkeit an ihren gesellschaftlich definierten Ansprüchen zu messen“ sowie „Bildungsziele mit der Bildungswirklichkeit zu konfrontieren, und zudem Hinweise (zu) liefern, wie man die Kon-

gruenz von Anspruch und Praxis erhöhen kann“ (*Leschinsky & Cortina* 2003, 28, 46). Eine maßgebliche Rolle spielt dabei die 2003 begonnene nationale Bildungsberichterstattung (vgl. *Avenarius et al.* 2003; *Konsortium Bildungsberichterstattung* 2006 sowie den Beitrag von *Döbert & Avenarius* in diesem Band). Die „wichtigste Grundlage für Leistungstransparenz im System“ (*Baumert et al.* 2003, 147), nämlich ihre systematische Beobachtung, scheint damit weitgehend etabliert zu sein.

Wiederum waren es wegweisende Beschlüsse der KMK, die zu diesen Entwicklungen geführt haben, nämlich die „Eisenacher Beschlüsse“ von 2002 zur Einführung einer deutschen Bildungsberichterstattung und zur Entwicklung nationaler Bildungsstandards. Die Bildungsstandards stellen das Input-Pendant zur regelmäßigen Outputmessung durch Leistungsfeststellungen und Schulberichterstattung dar. Die Schulqualitätsdiskussion hat damit (wieder) die Systemebene erreicht.

## 5 Bilanz und Perspektiven

### 5.1 Zur Programmatik der Schulqualitätsdiskussion

Vor dem Hintergrund der Schulqualitätsdiskussion in Deutschland ist zu fragen, was sie kennzeichnet, was ihr gemeinsamer Nenner ist und wie sich diese Diskussion in den verschiedenen Stadien entwickelt hat. Meines Erachtens lassen sich in der Schulqualitätsdiskussion unschwer zwei zentrale Verwendungszusammenhänge ausmachen: Von Qualität ist die Rede,

- 1) wenn unterschiedliche Potenziale, Konfigurationen und Realisierungen von Schulgüte thematisiert werden und
- 2) wenn Definitionen über Aufgaben und Leistungen des Schulsystems auf den unterschiedlichen Handlungsebenen zur Debatte stehen.

In einer **ersten Phase** der Diskussion interessierten vor allem Fragen, wie Schulen mit den Rahmenbedingungen der Systemebene (und des lokalen Umfeldes) umgehen, wie Schulen ihre innere Organisation handhaben und als Rahmenbedingung für unterrichtliche Prozesse gestalten. Allgemein gesprochen ging es um eine Erfassung unterschiedlicher Realisierungen makro-organisatorischer Vorstrukturierungen auf der Handlungsebene der einzelnen Schule. Vor diesem Hintergrund sollten Charakterisierungsmerkmale und Gütekriterien zur Bestimmung von Schul- und Unterrichtsqualität sowie Gelingensbedingungen für die Schulentwicklung identifiziert werden.

In einer **zweiten Phase** der Diskussion – ausgelöst durch die Ergebnisse der TIMS-Studie – stand die Frage der Leistungsfähigkeit des Schulsystems auf der Unterrichtsebene im Mittelpunkt, und zwar in einer doppelten Perspektive: Einerseits ging es um eine Verbesserung der Unterrichtsqualität (vorrangig in Mathematik und den naturwissenschaftlichen Fächern) und andererseits um eine Verbesserung der empirischen Basis für die Erfassung von Schülerleistungen (internationale und nationale Vergleichsuntersuchungen bzw. Lernstandserhebungen).

Im Zuge der PISA-Studie und anderer Vergleichsuntersuchungen kann von einer **dritten Phase** der Qualitätsdebatte gesprochen werden, in der nun Standards schuli-

scher Wirkungen und Prozesse im Zentrum der Betrachtungen stehen. Hierbei geht es um die Definition und Erfassung von Fachleistungen in den Kernfächern („Bildungsstandards“) einerseits und von Organisations-, Lehr- und Lernprozessen („Schulinspektion“) andererseits, um Auskunft über die Güte des Schulwesens – insbesondere auf der Systemebene – zu erhalten („Systemmonitoring“).

Gemeinsame Kennzeichen für alle Phasen der Schulqualitätsdiskussion sind zwei Wesensmerkmale, die eine normative Komponente implizieren:

- Die Erfassung, Beschreibung und Bestimmung von Schulqualität ist empirisch ausgerichtet. Frühere Bemühungen – vor allem in den 1960er Jahren und teilweise bis weit in die 1980er Jahre hinein – waren demgegenüber nahezu ausschließlich normativ ausgerichtet.
- Die Schulqualitätsdiskussion folgt einem praktischen Erkenntnisinteresse. Es geht – ganz allgemein formuliert – um eine Bestimmung dessen, was Schule leisten und wie sie arbeiten soll.

Nimmt man die Diskussionen des „Arbeitskreises Qualität von Schule“ zum Maßstab, dann ist unverkennbar, dass der Begriff Schulqualität in den Kontext von Schulentwicklung einzuordnen ist (vgl. *Steffens & Bargel* 1993, 8ff.). Seine Besonderheit liegt jedoch in seiner spezifischen Ausrichtung. In einer sozial-ökologischen Perspektive wird nach den Handlungskontexten auf den verschiedenen Ebenen des Schulsystems und nach den Möglichkeiten ihrer Realisierung gefragt: „Was ist eine gute Schule und wie ist sie zu verwirklichen?“ (*Steffens & Bargel* 1993, 9). Mit dieser Fragerichtung ist ein integrativer Ansatz verknüpft, der erstens verschiedene Forschungstraditionen und -methoden, zweitens verschiedene Entwicklungsansätze und -konzepte sowie drittens verschiedene normative Vorstellungen und Diskussionsstränge zu verbinden versucht, um die Gestaltungsmöglichkeiten und Gelingensbedingungen für Schule und Unterricht zu identifizieren und Orientierung für die Schulentwicklung zu geben.

Dabei schien eine Begriffsklärung bislang von nachgeordneter Bedeutung zu sein: In der Schulforschung und Schulentwicklung gibt es wohl kaum ein anderes Thema, über das so viel geschrieben und gesprochen wurde wie über die Schulqualität – ohne dass eine klare Verständigung darüber erzielt wurde, was darunter zu verstehen sei. Beispielsweise kann man sich folgenden Sichtweisen nicht verschließen:

- „Aus schultheoretischer Sicht ist festzuhalten, dass mit dem Begriff wertende Positionen zusammenhängen“; „mit dem Begriff Qualität wird eine große Spannweite unterschiedlicher Intentionen zusammengefasst“ (*Kleinschmidt* 1999, 100 bzw. 102).
- „Recht pragmatisch ist die japanische Definition: Qualität ist alles, was man verbessern kann“ (*Posch* 1999b, 199).
- „Schulqualität ist zu entdecken, nicht zu erfinden“ (*Szaday* 1999, 231).
- „(...) die Qualitätsdiskussion (ist) tatsächlich nichts anderes als die Diskussion über Bildung und Erziehung“ (*Liket* 1993, 83).
- „Qualität ist das Leitziel einer pädagogischen Schulentwicklung“ (*OECD* 1991).

*Fend* (1999, 138) erläutert den Begriff wie folgt: „Der Begriff der Qualität führt in den pädagogischen Diskurs lediglich eine formale Differenz ein. Er besagt vorläufig nicht mehr, als die Bevorzugung von etwas Höherwertigem gegenüber etwas Wenigerwertigem. Pädagogik kann generell als Strategie der Kultivierung einer Differenz zwischen Natur und Kultur, zwischen Geschehendem und Gestaltetem verstanden werden. Die Differenz, die durch ‘Qualität’ induziert wird, ist in Bezug auf den pädagogischen Diskurs noch höchst formal. Es ist deshalb unumgänglich, zu einer inhaltlichen Formulierung dieser Differenz zu kommen.“

Nach *Heid* (2000, 41) ist Qualität „keine beobachtbare Eigenschaft oder Bezeichnung eines Objekts, sondern das Resultat einer Bewertung der Beschaffenheit eines Objekts“. *Posch & Altrichter* (1997, 28) schlagen im Anschluss an *Burrows et al.* (1992, 5f.) eine „pragmatische Definition von Qualität“ vor:

„Es wäre vergebliche Mühe, *ein einheitliches* Qualitätskonzept zu definieren, weil Qualität ein *relativer* Begriff ist und nur im Hinblick auf die Werte der verschiedenen Interessengruppen (stakeholders) näher bestimmt werden kann. Nach *Burrows* muss Qualität im Sinne von Qualitäten definiert werden, wobei eine Institution *im Hinblick auf einen Faktor hohe Qualität aufweisen kann, während sie in Bezug auf einen anderen Faktor von niedriger Qualität sein kann. Man kann nicht mehr erreichen, als jene Kriterien, die jeder stakeholder bei seinen Qualitätseinschätzungen benutzt, so klar wie möglich zu definieren und diese - zueinander im Wettbewerb stehenden - Sichtweisen zu berücksichtigen, wenn Qualitätsbeurteilungen vorgenommen werden*“ (kursiv im Original).

*Böttcher* (2002, 92ff.) erläutert den Qualitätsbegriff wie folgt: „Qualität ‚an sich‘ gibt es nicht. Auch nicht mit Blick auf die Schule. Ohne Ziel- bzw. Aufgabenbestimmung hängt die Schulqualität ‚in der Luft‘ (vgl. auch *Oelkers* 1997) (...) Schule ist eine gesellschaftliche Einrichtung und in Demokratien können (oder sollten) die Ziele der Schule nicht einfach ‚von oben‘ gesetzt, sondern sie müssen in öffentlichen Diskursen ausgehandelt werden.“

Man muss den vielen kritischen Kommentaren und Analysen zur deutschen Schulqualitätsdiskussion und zum Begriff Schulqualität daher zustimmen, wenn sie feststellen, dass man von einer eigenständigen Forschungsdisziplin Schulqualität nicht sprechen kann, dass es enorme Theoriedefizite gibt und dass der Begriff Schulqualität vage ausfällt, für unterschiedliche Zwecke benutzt wird und bis heute nicht klar definiert ist (vgl. dazu u. a. *Büeler* 1996; *Harvey & Green* 2000; *Heid* 2000; *Helmke et al.* 2000; *Böttcher* 2002; *Helmke* 2005; *Galiläer* 2005; *Steffens & Boenicke* 2000). So geht auch die OECD bei der Definition von Schulqualität pragmatisch vor: „Qualität wird in einem doppelten normativen Sinne erfahren: einerseits kennzeichnet der Begriff einen (zeitweisen) Zustand im Hinblick auf die Vorstellung von ‘Perfektion’. Zweitens erlaubt er eine ordinale Attribuierung (besser, sehr gut, von mittlerer Qualität...). Es wird auch noch erwähnt, dass Qualität nicht den Antagonisten zu Quantität bildet. Erstens können rein quantitative Merkmale als Kriterien von Qualität fungieren. Zweitens können hinter qualitativen Wertungen durchaus quantitative und/oder mit quantitativen Verfahren erhobene Daten stecken. Die OECD verzichtet auf eine inten-

sivere Begriffsdiskussion, sondern betrachtet stattdessen aktuelle Ansätze, ihre Dimensionen und Perspektiven (*OECD 1991, S. 39*)“ (*Böttcher 2002, 87*).

Vor dem Hintergrund der deutschen Schulqualitätsdiskussion entsteht der Eindruck, dass die Bezeichnung Schulqualität für eine Bewegung steht. Schulqualität ist ein Programm in praktischer Absicht, das Forschungsbefunde im Hinblick auf ihren möglichen Nutzen für die Verbesserung des Schulwesens auf den unterschiedlichen Handlungsebenen (Makro-, Meso- und Mikroebene) und in seinen verschiedenen Handlungsstadien (Input, Prozess, Output/Outcome) befragt und nach bestmöglichen Antworten sucht – auf der Basis geeigneter Forschungsdesigns und dementsprechend praktizierter Schul(entwicklungs)forschung (vgl. dazu auch *Fend 1998; Holtappels 2005*).

Die normative Komponente ist aus dieser Bewegung bzw. diesem Programm nicht wegzudenken, genauso wenig wie die Notwendigkeit einer Verständigung über ihre normativen Implikationen. Letztendlich geht es um eine „praktische Philosophie“ der Schulentwicklung. Auf die gleiche Weise, wie die Medizinwissenschaft eine vorherrschende Lehrmeinung kundtut und praktiziert, benötigt auch die Erziehungswissenschaft eine praktische Erziehungslehre im Sinne „bewährter Hypothesen“. Bildungsplanung und Schulentwicklung sind auf die bestmögliche Expertise angewiesen, wenn es um Standards für die Lehrerbildung, für erzieherische Wirkungen und erzieherische Prozesse geht.

Verschiedene Schulrechtsexperten haben in Fachkreisen wiederholt darauf hingewiesen, dass es Aufgabe vor allem der Erziehungswissenschaft sei, Kriterien für Schulqualität bereitzustellen, die als Orientierungsmaßstab für die fachaufsichtliche Überprüfung heranzuziehen seien. In diesem Zusammenhang wurde wiederholt bemängelt, dass die Schulforschung einerseits zwar Ergebnisse zu Tage fördere, sich aber andererseits davor hüte, diese Ergebnisse als vorherrschende Lehrmeinung auszuweisen. Offenbar, so könnte zugespitzt gefolgert werden, ist es ein auffallendes Kennzeichen der deutschen Erziehungswissenschaft, mit praktischen Folgerungen aus vorherrschenden Lehrmeinungen sehr zurückhaltend zu sein. Die angewandte Schulforschung (vor allem an den Landesinstituten) ist jedoch aufgrund der Handlungserfordernisse genötigt, Stellung zu beziehen. Sie muss Positionierungen dazu vornehmen, welches die zentralen Erkenntnisse in praktischer Absicht sind.

Es ist unbestritten, dass „man von empirischen Studien (...) keine definitiven Antworten über ‚Qualität von Schule‘ erwarten (kann). Empirie ist ihrem Wesen nach deskriptiv, die Frage nach der ‚Qualität von Schule‘ normativ“ (*Rolff 1991*). In praktischer Perspektive bleibt somit nur die Wahl, die normativen Entscheidungen so rational und transparent wie möglich zu gestalten und diesen Prozess als Entscheidungsakt – nach dem Evaluationsverständnis von *Stuffelbeam (1972)* – zu kennzeichnen. Für ein demokratisch gestaltetes Gemeinwesen ist dabei die Konsensfähigkeit entscheidend.

Meines Erachtens liegt ein entscheidendes Verdienst des Schulqualitätsansatzes darin, dass er durch seine empirische Vorgehensweise wesentlich dazu beigetragen hat, die jahrelang geführten ideologischen Grabenkämpfe zwischen den verschiedenen pädagogischen Lagern als häufig „jenseits scheinbarer Alternativen“ liegend zu erken-

nen (*Bohnsack* 1991), und dass er zugleich die hohe Übereinstimmung in wesentlichen Fragen von Schulqualität zu Tage gefördert hat (vgl. *Boenicke & Steffens* 1999). Dies gilt selbst für die stärker normativ orientierte Schulpädagogik (vgl. z. B. *Meyer* 2004).

Im Rückblick auf die Schulqualitätsdiskussion von drei Dekaden wird offensichtlich, dass sich hinter den unterschiedlichen thematischen Phänomenen – neben oder unbeschadet einer allgemeinen Schulkritik – eine leitende Tiefenstruktur verbirgt, die von zwei Faktoren ‚gespeist‘ wird: Dies sind die Steuerbarkeit und die Wirksamkeit des Schulsystems, die sich in praktischer Absicht zu einem gemeinsamen Handlungszusammenhang verbinden lassen. Wir kommen damit auf den Ausgangspunkt der deutschen Schulqualitätsdiskussion zurück, der heute noch so aktuell wie damals ist: Die Frage nach den Wirkungen von Schule, ihren Steuerungsmöglichkeiten und Gelingensbedingungen (*Melzer* 1997, *Fend* 2000).

Aufgrund ihrer Genese und Rezeption in der Bildungsplanung und Schulentwicklung ist die ‚Schulqualitätsbewegung‘ untrennbar mit der Einzelschule als pädagogischer Handlungsebene verbunden. Diese Bewegung wäre jedoch grundlegend missverstanden, wollte man daraus einen Primat der Einzelschule über die anderen Handlungsebenen ableiten. Das Schulsystem stellt ein komplexes System dar, „in dem sowohl Elemente auf verschiedenen Ebenen der Gestaltung (Makro-, Meso- Mikroebene) als auch Relationen zwischen verschiedenen Elementen ausgezeichnet werden können, ja schließlich die Stimmigkeit der ‚Gesamtgestalt‘ eines Bildungssystems zur Debatte steht“ (*Fend* 1998, 16). Die Qualitätsfrage stellt sich damit auf allen Handlungsebenen. Jede Ebene hat eigene Aufgaben, die jeweils ihren Beitrag im Gesamtsystem leisten; gleichzeitig sind die einzelnen Ebenen wechselseitig miteinander verbunden (vgl. dazu bereits *Barr & Dreeben* 1983). „Heute hat sich die Vorstellung durchgesetzt, daß die Besonderheiten der jeweiligen Ebene sichtbar gemacht werden müssen, die nicht in anderen aufgehen können, und daß die Formen des Zusammenwirkens verschiedener Ebenen selber von großer Bedeutung sind“ (*Fend* 1998, 358). Entsprechende Analysen von *Ditton* (1997) und *Fend* (1998 und 2006) belegen dies eindrücklich.

## 5.2 Erträge der Schulqualitätsdiskussion: Erfolge und Ernüchterungen

In einer vorläufigen Bilanz der Schulqualitätsdiskussion in Deutschland ist eine ambivalente Forschungs- und Entwicklungssituation zu konstatieren, die durch Erfolge einerseits und Ernüchterungen andererseits geprägt zu sein scheint:

Zunächst einmal ist der **Erkenntnisgewinn** offensichtlich, der durch die empirische Ausrichtung der Fragestellung erzielt wurde. Dies gilt vor allem für die erste Phase der Qualitätsdiskussion, in der maßgebliche Befunde verschiedener Forschungsrichtungen daraufhin befragt wurden, was sie zu der zentralen Frage nach den relevanten Variablen von Schulqualität beitragen können.

Eine damit verbundene erste Folgefrage – „does school matter?“ – schien in der Auseinandersetzung mit den Forschungsbefunden vor allem von *Coleman et al.* (1966) und *Jencks* (1973) einerseits sowie *Rutter et al.* (1979), *Brookover et al.* (1979) und *Fend* (1980) andererseits zunächst recht rasch beantwortet. Diese Frage war zugleich

der Ausgangspunkt für die deutsche Qualitätsbewegung. Auch bei der zweiten Folgefrage – „what works?“ – kam man relativ rasch voran. In zahlreichen Publikationen, maßgeblich in Form synoptischer Darstellungen von Forschungsergebnissen, aber auch in zusätzlich in Auftrag gegebenen spezifischen Analysen und Fallstudien wurde eine breite Palette an Befunden über gestaltungsrelevante Merkmale zusammengestellt. In der Bündelung solcher Forschungsbefunde ist der besondere Ertrag des „Arbeitskreises Qualität von Schule“ zu sehen. Zur dritten Folgefrage, nämlich der Frage nach den Kennzeichen erfolgreicher Schulen, trug insbesondere die angelsächsisch geprägte „School Effectiveness“-Forschung maßgebliche Erkenntnisse bei, die auch die deutsche Diskussion nachhaltig beeinflusst haben.

Erst bei der sich daran anschließenden vierten Folgefrage, welche Variablen aus der Vielfalt der Forschungsbefunde und Qualitätskataloge im Einzelnen und konkret für die Gestaltung und Verbesserung von Schule und Unterricht zentral und unabdingbar seien, zeigte sich die defizitäre Situation, dass trotz der umfangreichen Erhebungen und zahlreichen Studien weiterhin ein erheblicher Forschungsbedarf besteht. Dies wird insbesondere daran erkennbar, dass die als relevant ausgewiesenen Bedingungsvariablen in keinem systematischen Zusammenhang mit Wirkungsvariablen stehen, dass sich keine eindeutigen Wenn-dann-Beziehungen zeigen und sogar widersprüchliche Befunde auftreten können:

„Die Kenntnis von Fakten allein führt (...) nicht notwendigerweise auch zu einem gesicherten Wissen darüber, was denn verändert werden müsste, und sie gibt noch keinen Anhaltspunkt dafür, wie Veränderungen bewirkt werden können (Gray et al. 1996)“ (Ditton 2000b, 20).

Dieses Erkenntnisdefizit erschwert auch die Beantwortung der bildungsökonomischen Frage, wie Ressourcen und Wirkungen im Hinblick auf Schulqualität zusammenspielen. Das Missverhältnis zwischen öffentlichem Aufwand und Ertrag sowie die geringe Wirksamkeit investierter Mittel – in Deutschland erstmals im Zusammenhang mit den TIMSS-Ergebnissen von einer breiten Öffentlichkeit wahrgenommen – führten zu dem „globalen Phänomen“ einer „massiven ökonomischen Kritik an Schulen“ und zur Kritik am „Mangel an Wissen über Funktionsweise und Wirkungen der Schule“ (Böttcher 2002, 27 bzw. 24).

Lässt man einmal die drei Dekaden Revue passieren, so ist enttäuschend, wie wenig an neuen Erkenntnissen im Verlauf der Jahre hinzugewonnen wurde. Auch die Schulforschung in Deutschland konnte keine erheblichen Fortschritte bei der Identifikation gestaltungsrelevanter Variablen verzeichnen (vgl. Ditton 2000a; Galiläer 2005). Selbst für die auf die Wirksamkeit von Schule ausgerichtete „School Effectiveness“-Forschung wird eine ernüchternde Bilanz gezogen (Mortimor 1997; Scheerens & Bosker 1997).

„Ob der Komplexität des Gegenstandes geschuldet oder aufgrund der notwendigen normativen Standards einer Bewertung, offenbar wurde nur wenig Wissen über Schule und ihre Wirkungen generiert. So stellen Block und Klemm in ihrer oben zitierten Studie fest: ‚Bei der Durchsicht der deutschen Forschungsliteratur hat uns überrascht, wie spärlich die Befunde zu den Effekten institutionalisierter Bildung ausfallen‘ (Block &

*Klemm* 1997, 11). Die deutsche erziehungswissenschaftliche Forschungstradition ist eher wenig darauf orientiert, in der Schule erworbenes Wissen, Einstellungen oder Fähigkeiten systematisch zu untersuchen“ (*Böttcher* 2002, 24).

Bei der Bewertung der vorliegenden Forschungsbefunde muss allerdings unterschieden werden zwischen einer „Kartografierung“ relevanter Gestaltungsvariablen im Kontext von Schulqualität einerseits (im Zusammenhang mit den internationalen Vergleichsuntersuchungen sind hier im Übrigen weitere Erkenntnisse hinzugewonnen worden) und den Möglichkeiten der Vorhersage von Ergebnissen bei systematischer Variation der Gestaltungsparameter andererseits. Intentionen letztgenannter Art werden beim derzeitigen Erkenntnisstand zwangsläufig zu Enttäuschungen führen.

Diese **Ernüchterungen** haben vor allem mit der relativ geringen technologischen Verwertbarkeit der Forschungsbefunde zu tun, die nur selten klare ‚Produktionszusammenhänge‘ zwischen Input und Output zeigten. Wer die Erwartung hegt, er könne einzelne Lehr- und Lernbedingungen hinreichend präzise aus dem „Gesamtkörper“ von Schulgestalten herauspräparieren, um sie in erwünschte Ziel-Mittel-Zusammenhänge einzupassen, verkennt die Komplexität schulischer Organisationen (vgl. jüngst *Merkens* 2006). Diese Komplexität lässt sich zum Beispiel an ihrer „loosely coupled system“-Struktur (*Weick* 1976) und an ihrem komplexen gesellschaftlichen Auftrag (vgl. *Steffens & Bargel* 1992) ablesen.

Obwohl wir bereits seit einiger Zeit über eine Reihe von Ergebnissen bezüglich gelingender Reformprozesse verfügen (vgl. z. B. *Holtappels* 1995b; *Skibba & Steffens* 1995; *Haenisch* 1996; *Rolff* 1999a und 1999b), ist nicht zu bestreiten, dass die Erforschung schulischer Entwicklungsprozesse noch im Anfangsstadium begriffen ist. Der Vorwurf von *Huber* (1999), dass sich die Schulforschung zu wenig mit der Frage befasse, wie Schulen tatsächlich lernen, ist insofern berechtigt. Offenbar sind die Forschungsmethoden und -designs nach wie vor nicht elaboriert genug und fehlt es an den erforderlichen Forschungsmitteln. Das gilt auch für den „School Effectiveness“-Ansatz, dessen Befunde bezüglich wirksamer Schulen nicht immer konsistent und reliabel ausfallen (vgl. *Scheerens & Bosker* 2005).

Welche Bedeutung in strategischer und erkenntnisleitender Hinsicht dem Ausmaß an Forschung beizumessen ist, wird daraus ersichtlich, dass in allen Phasen der Schulqualitätsentwicklung in Deutschland vor allem solchen Forschungsprojekten eine große Bedeutung zukam, die an internationalen und nationalen Großuntersuchungen beteiligt waren bzw. Systemvergleiche im staatlichen Auftrag durchführten. Das betrifft in den 1970er Jahren in besonderer Weise die Konstanzer Schulforschung (*Helmut Fend* und Mitarbeiter/innen), aber auch Arbeiten des Instituts für Schulentwicklungsforschung (IfS) in Dortmund. In den 1980er und 1990er Jahren gilt das vor allem für die Max-Planck-Institute für Bildungsforschung in Berlin und für Psychologische Forschung in München, in den 1990er Jahren insbesondere für die Projekte BIJU und TIMSS und in den letzten Jahren vornehmlich für PISA. Die großen Vergleichsstudien sind insofern als Meilensteine anzusehen, mit denen jeweils entscheidende Phasen der Schulqualitätsdiskussion in Deutschland verbunden sind.

Meines Erachtens ist es dringend geboten, mit ähnlichem Aufwand, wie derzeit Schulsystemmonitoring in seinen verschiedenen Formen betrieben wird (Vergleichsuntersuchungen, Lernstandserhebungen, Schulberichterstattung), auch die Schulentwicklungsforschung zu forcieren (Messner 2003). Experten stimmen darin überein, dass die praktische Umsetzung von PISA ihrerseits ein Forschungs- und Entwicklungsproblem darstellt, das dem Aufwand der Schulvergleichsstudien nicht nachsteht. „Die eigentliche Arbeit“, so Baumert (2001, 33), „beginnt in der Schule, den sie unterstützenden Einrichtungen und der Politik erst nach der Untersuchung“. Zudem ist zu konstatieren, dass es in Deutschland bislang weder eine Wirksamkeitsforschung noch eine ausgeprägte Evaluationsforschung gibt – und erst recht keine Schulqualitätsforschung.

## 6 Literatur

- Achs, Oskar; Gruber, Karl Heinz; Kral, Paul & Tesar, Eva (Hrsg.) (1995): Schulqualität – Facetten und Felder einer Entwicklung. Vorträge und Diskussionen anlässlich des 1. europäischen Bildungsgespräches ‚94. Wien: ÖBV Pädagogischer Verlag.
- Ackeren, Isabell van (2003): Evaluation, Rückmeldung und Schulentwicklung. Erfahrungen mit zentralen Tests, Prüfungen und Inspektionen in England, Frankreich und den Niederlanden (Studien zur International und Interkulturell Vergleichenden Erziehungswissenschaft, Band 2). Münster u. a.: Waxmann.
- Allmann, V. (1994): Die Rolle der Schulaufsicht im Entwicklungsprozeß von Schule und bei der Umsetzung von Bildungsplänen. Ein neues Leitbild von Schulaufsicht durch Dialog mit den Schulen. In: Buchen, H./Horster, L./Rolf, H.-G. (Hrsg.): Schulleitung und Schulentwicklung. Erfahrungen, Konzepte, Strategien. Berlin u.a.: Raabe, H. 2.1.
- Altrichter, Herbert (1998): What makes a good school tick? Einige Überlegungen und Diskussionsanregungen aus dem OECD/CERI-Projekt „Teacher Quality“. In: Stefens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 123-135.
- Arbeitsstelle Bildungskonzeptionen, Qualitätssicherung und Qualitätsentwicklung (1998): Konzeptionelle Überlegungen zum „Regionalen Aktionsprogramm Schulentwicklung (RASCH)“. Wiesbaden: Hessisches Landesinstitut für Pädagogik (HeLP), Arbeitsstelle Bildungskonzeptionen.
- Aurin, Kurt (Hrsg.) (1990): Gute Schulen – Worauf beruht ihre Wirksamkeit? Bad Heilbrunn: Klinkhardt.
- Aurin, K.; Stark, G. & Stobberg, E. (1977): Beratung im Schulbereich. Aufgabenfelder, Strukturprobleme, Entwicklungstendenzen und Empfehlungen. Weinheim & Basel: Beltz.
- Avenarius, Hermann; Baumert, Jürgen; Döbert, Hans & Füssel, Hans-Peter (Hrsg.) (1998): Schule in erweiterter Verantwortung. Positionsbestimmungen aus erziehungswissenschaftlicher, bildungspolitischer und verfassungsrechtlicher Sicht. Neuwied & Kriftel/Ts.: Luchterhand.
- Avenarius, Hermann; Ditton, Hartmut; Döbert, Hans; Klemm, Klaus; Klieme, Eckhard; Rürup, Matthias; Tenorth, Heinz-Elmar; Weishaupt, Horst & Weiß, Manfred (2003): Bildungsbericht für Deutschland. Erste Befunde. Opladen: Leske+Budrich.
- Averch, H. A. et al. (1974): How Effective Is Schooling? A Critical Review of Research. Englewood Cliffs, N. J.: Rand Corporation.
- Barr, Rebecca & Dreeben, Robert (1983): How Schools Work. Chlicago, Il.:University of Chicago Press.
- Baumert, Jürgen (1998): Fachbezogenes – fachübergreifendes Lernen. Erweiterte Lern- und Denkstrategien. Einführung in die Thematik. In: Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (Hrsg.): Wissen und Werte für die Welt von morgen. München und Donauwörth: Auer, S. 213-225.
- Baumert, Jürgen (2001): Vergleichende Leistungsmessung im Bildungsbereich. In: Zeitschrift für Pädagogik, 43. Beiheft, 13-36.
- Baumert, Jürgen; Lehmann, Rainer; Lehrke, Manfred; Schmitz, Bernd; Clausen, Marten; Hosenfeld, Ingmar; Köller, Olaf & Neubrand, Johanna (1997): TIMSS – Mathematisch naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske+Budrich.
- Baumert, Jürgen; Bos, Wilfried; Watermann, Rainer (1998): TIMSS/III. Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich. Zusammenfassung deskriptiver Ergebnisse (Studien und Berichte, Band 64). Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, Jürgen; Bos, Wilfried & Lehmann, Rainer (Hrsg.) (2000 a): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Bildung am Ende der Pflichtschulzeit. Opladen: Leske+Budrich.

- Baumert, Jürgen; Bos, Wilfried & Lehmann, Rainer (Hrsg.) (2000 b):
- TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie am Ende der Schullaufbahn. Band 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen: Leske+Budrich.
- Baumert, Jürgen; Cortina, Kai S. & Leschinsky, Achim (2003):
- Grundlegende Entwicklungen und Strukturprobleme im allgemein bildenden Schulwesen. In: Cortina, Kai S. et al., a.a.O., 52-147.
- Becker, H. (1956): Kulturpolitik und Schule. Probleme der verwalteten Schule. Stuttgart.
- Becker, H. & Hager, F. (1992): Aufklärung als Beruf. Gespräche über Bildung und Politik. München: Piper.
- Berg, Hans Christoph & Steffens, Ulrich (Hrsg.) (1991): Schulgüte und Schulvielfalt. Das Saarbrücker Schulgütesymposium '88 (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 5). Wiesbaden. Hessisches Institut für Bildungsplanung und Schulentwicklung (HIBS).
- Bessoth, R. (1974): Schulverwaltungsreform. Entwürfe zur Neugliederung der Schulaufsicht in einzelnen Bundesländern und ihre Kritik. Weinheim/Basel: Beltz.
- Bessoth, R. (1977): Managementdefizite der Schulverwaltung. In: Bessoth, R./Braune, G.: Schule und Management. Braunschweig: Westermann, 92-130.
- Bessoth, R. (1978): Positive Schulverwaltungsreform. Leistungsfähige Alternativen zur fragmentierten Schulverwaltung in der Bundesrepublik. Frankfurt/M: Max-Traeger-Stiftung.
- Bessoth, R. (1995): Zehn Thesen zur Bildungsverwaltung. In: Deutsche Gesellschaft für Bildungsverwaltung (DGBV): Lernen in einer offenen Gesellschaft. Dokumentation der 15. DGBV-Jahrestagung v. 15.-17.09.1994 in Bayreuth (DGBV-Jahrestagungen, Band 15). Frankfurt/M., 143-165.
- Biewer, W. (1993): Steuerung und Kontrolle öffentlicher Schulen. Neuwied: Luchterhand.
- Black, Paul & Atkin, J. Myron (Hrsg.) (1996): Changing the Subject. Innovations in Science, Mathematics and Technology Education. London, New York: Routledge.
- Block, Rainer & Klemm, Klaus (1997): Lohnt sich Schule? Reinbek: Rowohlt.
- Böhrs, Gerburg; Daschner, Peter; Meyer-Dohm, Peter; Rosenbusch, Heinz & Strittmatter, Anton: Externe Abschlussequivalenten der Fortbildungsmaßnahme QUESS. In: Ministerium für Schule und Weiterbildung & Landesinstitut für Schule und Weiterbildung: Schulentwicklung und Schulaufsicht – Qualitätsentwicklung und Qualitätssicherung von Schule. Ergebnisse und Materialien aus der Fortbildungsmaßnahme. Soest: Verlag für Schule und Weiterbildung, 49-94.
- Boenicke, Rosemarie & Steffens, Ulrich (1999): Qualitätskriterien und die Arbeit am Schulprogramm. In: Schulverwaltung spezial, Sonderausgabe Nr. 1/99, 18-22.
- Böttcher, Wolfgang (2002): Kann eine ökonomische Schule auch eine pädagogische sein? Schulentwicklung zwischen Neuer Steuerung, Organisation, Leistungsevaluation und Bildung. Weinheim und München: Juventa.
- Bohnsack, Fritz (1987): Pädagogische Strukturen einer „guten“ Schule heute. In: Steffens, Ulrich & Bargel, Tino (Hrsg.) (1987a), a.a.O., 105-118.
- Bohnsack, Fritz (1991): Strukturen einer guten Schule heute – Versuch einer normativen Begründung. In: Berg, Hans Christoph & Steffens, Ulrich (Hrsg.), a.a.O., 23-30.
- Brookover, W. B. & Lezotte, L. W. (1977): Changes in School Characteristics Coincident with Changes in Student Achievement. East Lansing Michigan State University, College of Urban Development.
- Brookover, W. B. et al. (1976): Elementary School Climate and School Achievement. Michigan State University, College of Urban Development.
- Brookover, W. B.; Beady, C.; Flood, P.; Schweitzer, J. & Wisenbaker, J. (1977): Schools can make a difference. East Lansing: Michigan State University, College of Urban Development.
- Buchen, H.; Burkard, C. & Eikenbusch, G. (1995): Fortbildung und Unterstützung für eine erweiterte Selbständigkeit von Schule. Beispiele aus der nordrhein-westfälischen Lehrerfortbildung. Soest: Landesinstitut für Schule und Weiterbildung.
- Buer, Jürgen van; Lehmann, Rainer H. & Venter, György (1997): Erweiterte Autonomie für die Schule – Einführung in das Thema, die Ziele und die Struktur der internationalen Sommerakademie. In: Venter, György; Buer, Jürgen van & Lehmann, Rainer H. (Hrsg.): Erweiterte Autonomie für Schule. Grundlagen und nationale Sichtweisen (Studien zur Wirtschafts- und Erwachsenenpädagogik aus der Humboldt-Universität zu Berlin, Band 13.1). Berlin: Humboldt-Universität, 17-26.
- Büeler, Xaver (1995): Die Verwirklichung guter Schulen: Trendbericht zur Schulqualitäts- und Schulentwicklungsforschung im deutschsprachigen Raum. Zürich: Universität Zürich, Pädagogisches Institut.
- Büeler, Xaver (1996): Prozeß oder Produkt? Grundprobleme bei der Erfassung und Beschreibung von Schulqualität. In: Bildung und Erziehung, 49, Heft 2, 135-154.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung – BLK (Hrsg.) (1982): Modellversuche mit Gesamtschulen. Auswertungsbericht der Projektgruppe Gesamtschule. Buhl/Baden: Konkordia Verlag.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung – BLK (Hrsg.) (1997): Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ (BLK-Materialien zur Bildungsplanung und Forschungsförderung, Heft 60). Berlin: BLK.
- Burkard, Christoph & Kohlhoff, Walter (1998): Schulentwicklung durch Schulaufsicht? Qualitätsentwicklung und -sicherung im Rahmen der Maßnahme „Schulentwicklung und Schulaufsicht“ des Landesinstituts für Schule

- und Weiterbildung. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 93-105.
- Burkard, Christoph & Rolf, Hans-Günter (1994): Steuerleute auf neuem Kurs? Funktionen und Perspektiven der Schulaufsicht für die Schulentwicklung. In: Jahrbuch der Schulentwicklung, Band 8. Weinheim/München: Juventa.
- Coleman, J.S.; Campbell, E.Q.; Hobson, C.J.; McPartland, J.; Mood, A.M.; Weinfield, F.D. & York, R.L. (1966): Equality of Educational Opportunity. Washington D.C.: U.S. Government Printing Office.
- Cortina, Kai S.; Baumert, Jürgen; Leschinsky, Achim & Mayer, Karl Ulrich (Hrsg.) (2003): Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick. Reinbek b. Hamburg: Rowohlt.
- Creemers, Bert; Peters, Ton & Reynolds, Dave (ed.) (1989): School Effectiveness and School Improvement. Proceedings of the Second International Congress Rotterdam 1989. Amsterdam u. a.: Swets & Zeitlinger.
- Dalin, P; Rolf, H.-G. & Buchen, H. (1990): Institutionelles Schulentwicklungsprogramm. Eine neue Perspektive für Schulleiter, Kollegium und Schulaufsicht. Soest: Soester Verlagskontor.
- Daschner, Peter; Rolf, Hans-Günter & Stryck, Tom (Hrsg.) (1995): Schulautonomie – Chancen und Grenzen. Impulse für die Schulentwicklung. Weinheim & München: Juventa.
- Deutscher Bildungsrat (1973): Zur Reform von Organisation und Verwaltung im Bildungswesen. Teil I: Verstärkte Selbständigkeit der Schule und Partizipation der Lehrer, Schüler und Eltern (Empfehlungen der Bildungskommission). Stuttgart: Klett.
- Deutsches PISA-Konsortium: Baumert, Jürgen; Klieme, Eckhard; Neubrand, Michael; Prenzel, Manfred; Schiefele, Ulrich; Schneider, Wolfgang; Stanat, Petra; Tillmann, Klaus-Jürgen & Weiß, Manfred (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske+Budrich.
- DGBV (Deutsche Gesellschaft für Bildungsverwaltung (Hrsg.) (1998): Qualität im Bildungssystem. Dokumentation der 18. DGBV-Jahrestagung vom 06. bis 08. November 1997 in Weinheim). Wetzlar: DGBV.
- Ditton, Hartmut (1997): Wirkung und Wirksamkeit der Einzelschule. Analysen zur Bedeutsamkeit der verschiedenen Ebenen des Schulsystems. In: Lehmann, Rainer H. u.a. (Hrsg.): Erweiterte Autonomie für Schule – Bildungscontrolling und Evaluation. Berlin-Nyiregyhaza, S. 91-116.
- Ditton, Hartmut (2000 a): Qualitätskontrolle und -sicherung in Schule und Unterricht. Ein Überblick über den Stand der empirischen Forschung. In: Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Hrsg.), a.a.O., 73-92.
- Ditton, Hartmut (2000 b): Elemente eines Systems der Qualitätssicherung im schulischen Bereich. In: Weisshaupt, Horst (Hrsg.), a.a.O., 13-35.
- Ditton, Hartmut (2002): Unterrichtsqualität. Konzeptionen, methodische Überlegungen und Perspektiven. In: Unterrichtswissenschaft, 30, Heft 3, 197-212.
- Döbert, Hans; Klieme, Eckhard; Kopp, Botho von; Weiß, Manfred et al. – Arbeitsgruppe „Internationale Vergleichsstudie“ (2003): Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten. Berlin: Deutsches Institut für Internationale Pädagogische Forschung (DIPF).
- Dörger, Ursula (1992): Projekt Lehrkooperation. Eine pädagogische Konzeption zur Weiterentwicklung von Gesamtschulen. Weinheim & München: Juventa.
- Dörger, Ursula & Steffens, Ulrich (1986): Die eigene Gesamtschule pädagogisch weiterentwickeln. In: Westermanns Pädagogische Beiträge, 38, Heft 7-8, 24-29.
- Doll, Jörg & Prenzel, Manfred (Hrsg.) (2004): Bildungsqualität von Schule: Lehrerprofessionalität, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung. Münster: Waxmann.
- Dubs, Rolf (1994a): Marktwirtschaft im Bildungswesen – eine Alternative? In: Schweizer Schule, Heft 6, 11-17.
- Dubs, Rolf (1994b): Die Führung einer Schule. Leadership und Management. Zürich: SKV.
- Dubs, Rolf (1996): Schule, Schulentwicklung und New Public Management. St. Gallen: Universität St. Gallen, Institut für Wirtschaftspädagogik.
- Edelhoff, Christoph (1988): Lehrerfortbildung im Schnittfeld der Interessen und Bedürfnisse. Instrumente zur Veränderung der Schule oder Serviceeinrichtung für Lehrer? In: Pädagogik 40, Heft 6, 8-12.
- Edmonds, R. R. (1979): Effective schools for the urban poor. In: Educational Leadership 37, 15-27.
- Edmonds, R. R. (1981): Making Public Schools Effective. In: Social Policy 12, 56-60.
- EFQM – The European Foundation for Quality Management (1995): The European Quality Award. Brüssel: EFQM.
- Einsiedler, Wolfgang (1997): Unterrichtsqualität und Leistungsentwicklung. Ein Literaturüberblick. In: Weisert, Franz E. & Helmke, Andreas (Hrsg.), a.a.O., 225-240.
- Ellwein, T. (1964): Die verwaltete Schule. In: Das Argument 6, Heft 31.
- Ermert, Karl (Hrsg.) (1986): „Gute Schule“ – Was ist das? Aufgaben und Möglichkeiten der Lehrerfortbildung. 6. überregionale Fachtagung der Lehrerfortbildner. Rehrburg-Loccum: Evangelische Akademie Loccum.
- Fend, H. (1977): Schulklima. Soziale Einflußprozesse in der Schule. Weinheim, Basel: Beltz.
- Fend, H. (1980): Theorie der Schule. München u.a.: Urban & Schwarzenberg.
- Fend, H. (1982): Gesamtschule im Vergleich. Bilanz der Ergebnisse des Gesamtschulversuchs. Weinheim/Basel: Beltz.
- Fend, H. (1987): "Gute Schulen - schlechte Schulen" - Die einzelne Schule als pädagogische Handlungseinheit. In:

- Steffens, U./Bargel, T. (Hrsg.): Erkundungen zur Wirksamkeit und Qualität von Schule (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 1). Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung, 55-80.
- Fend, H. (1990): Bilanz der empirischen Bildungsforschung. In: Zeitschrift für Pädagogik 36, 687-709.
- Fend, Helmut (1998): Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistungen. Weinheim & München: Juventa.
- Fend, Helmut (1999): Thesen zum Workshop. In: Grogger, Günther & Specht, Werner (Hrsg.), a.a.O., 136-138.
- Fend, Helmut (2000): Qualität und Qualitätssicherung im Bildungswesen. Wohlfahrtsstaatliche Modelle und Marktmodelle. In: Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Hrsg.).
- Fend, Helmut (2006): Neue Theorien der Schule. Einführung in das Verstehen von Bildungssystemen. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fend, H./Dreher, E./Haenisch, H. (1980): Auswirkungen des Schulsystems auf Schulleistungen und soziales Lernen - Ein Vergleich zwischen Gesamtschule und dreigliedrigem Schulsystem. In: Zeitschrift für Pädagogik 26, 673-698.
- Fleischer-Bickmann, Wolff (1998): Glanz der Profile – Elend der Programme? Kontexte der Theorie und Praxis von Schulprogrammen. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 83-92.
- Friedeburg, L. von (1989): Bildungsreform in Deutschland. Geschichte und gesellschaftlicher Widerspruch. Frankfurt/M.: Suhrkamp.
- Frommelt, Bernd & Steffens, Ulrich (1998): Schulautonomie – Auf dem Weg zu einem neuen Verständnis von Schulgestaltung. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 25-25.
- Fürstenau, P. (1969): Zur Psychoanalyse der Schule als Institution. In: Fürstenau, P. u.a. (Hrsg.): Zur Theorie der Schule. Weinheim/Basel: Beltz, 9-25.
- Fullan, Michael (1982): The Meaning of Educational Change. New York & London: Teachers College, Columbia University.
- Galiläer, Lutz (2005). Pädagogische Qualität. Perspektiven der Qualitätsdiskurse über Schule, Soziale Arbeit und Erwachsenenbildung. Weinheim und München: Juventa.
- Gampe, H. (1994): Kooperation zwischen Schulaufsicht und Schule. Untersuchungen zur pädagogischen und rechtlichen Schulratsfunktion. Neuwied: Luchterhand.
- Greber, U.; Maybaum, J.; Priebe, B. & Wenzel, H. (Hrsg.) (1991): Auf dem Weg zur „Guten Schule“. Schulinterne Lehrerfortbildung. Bestandsaufnahme, Konzepte, Perspektiven. Weinheim & Basel: Beltz.
- Grogger, Günther & Specht, Werner (Hrsg.) (1999): Evaluation und Qualität im Bildungswesen. Problem- analyse und Lösungsansätze am Schnittpunkt von Wissenschaft und Bildungspolitik. Dokumentation eines internationalen Workshops (Blumau, Steiermark, 18.-21. Feber 1999). Graz: Zentrum für Schulentwicklung.
- Haenisch, Hans (1996a): Qualitätssichernde Maßnahmen in der Schule. Ein Trendbericht auf der Grundlage der Ergebnisse einer internationalen Tagung zur Schulqualität. In: Bildung und Erziehung, 49, Heft 2, S. 183-196.
- Haenisch, Hans (1996b): Bausteine für die Entwicklung und die Profilbildung von Schulen. In: Landesinstitut für Schule und Weiterbildung (Hrsg.): Jahrbuch des Landesinstituts für Schule und Weiterbildung: Schule auf neuen Wegen. Anstöße, Konzepte, Beispiele. Bön- nen/Westf.: Verlag für Schule und Weiterbildung, 21-52.
- Hage, Klaus; Bischoff, Heinz; Dichanz, Horst; Eubel, Klaus-D.; Oehlschläger, Heinz-Jörg & Schwittmann, Dieter (1985): Das Methoden-Repertoire von Lehrern. Eine Untersuchung zum Unterrichtsalltag in der Sekun- darstufe I. Opladen.
- Haller, Ingrid & Wolf, Hartmut (1973): Neuorganisation der Lehrerfortbildung als Teilstück schulnaher Curriculum- entwicklung. In: curriculum konkret, 1, Heft 1.
- Harvey, L. & Green, D. (2000): Qualität definieren. Fünf verschiedene Ansätze. In: Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Hrsg.), a.a.O., 17-39.
- Heid, H. (2000): Qualität. Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie. In: Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Hrsg.), a.a.O., 41-51.
- Heipcke, Klaus (1977): Kommunikation über Unterricht – Gefährdungen pädagogischen Alltagshandelns (Beiheft 7 zu curriculum konkret). Frankfurt/M.: Regionale Leh- rerfortbildung.
- Hentig, H. von (1990): Bilanz der Bildungsreform in der Bundesrepublik Deutschland. In: Neue Sammlung 30, 366-384.
- Helmke, Andreas (2005): Unterrichtsqualität – erfassen, bewerten, verbessern. Seelze: Kallmeyer.
- Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (2000): Qualität und Qualitätssicherung im Bildungsbe- reich. Zur Einleitung in das Beiheft. In: Helmke, Andre- as; Hornstein, Walter & Terhart, Ewald (Hrsg.), a.a.O.
- Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Hrsg.) (2000): Qualität und Qualitätssicherung im Bil- dungsbereich: Schule, Sozialpolitik, Hochschule. 41. Beiheft der Zeitschrift für Pädagogik. Weinheim und Basel: Beltz.
- Helmke, Andreas & Jäger, Reinhold S. (Hrsg.) (2002): Die Studie MARKUS – Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext. Grundlagen und Perspektiven. Landau: Verlag Empirische Pädagogik.
- Hentig, Hartmut von (1987): „Humanisierung“ – eine verschämte Rückkehr zur Pädagogik? Andere Wege zur Veränderung der Schule. Stuttgart: Klett-Cotta.
- Hentig, Hartmut von (1990): Bilanz der Bildungsreform in der Bundesrepublik Deutschland. In: Neue Sammlung, 30, Heft 3, 366-384.
- Hoffmann, E. (1988): „Die Zusammenarbeit müßte besser werden, aber ...“ Veränderungen von Schulen durch

- Kollegiumsberatung? In: Neue Sammlung 28, Heft 3, 387-403.
- Holtappels, Heinz Günter (Hrsg.) (1995a): Entwicklung von Schulkultur. Ansätze und Wege schulischer Erneuerung. Neuwied & Kriftel/Ts.: Luchterhand.
- Holtappels, Heinz Günter (1995b): Innere Schulentwicklung: Innovationsprozesse und Organisationsentwicklung. In: Rolff, Hans-Günter (Hrsg.), a.a.O., 327-354.
- Holtappels, Heinz Günter (1999): Schulentwicklungsforschung auf neuen Wegen. In: Rösner, Ernst (Hrsg.), a.a.O., 175-218.
- Holtappels, Heinz Günter (2005): Bildungsqualität und Schulentwicklung. In: Holtappels, Heinz Günter & Höhmann, Katrin (Hrsg.): Schulentwicklung und Schulwirksamkeit. Systemsteuerung, Bildungschancen und Entwicklung der Schule. Weinheim & München: Juventa, 27-47.
- Hopes, Clive (1998): Transnationale Kooperation auf dem Gebiet der Beurteilung, Evaluation und Sicherung der Qualität von Schulen. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 107-121.
- Hopf, Ch./Nevermann, K./Richter, I. (1980): Schulaufsicht und Schule. Eine empirische Analyse der administrativen Bedingungen schulischer Erziehung (Veröffentlichungen des Max-Planck-Instituts für Bildungsforschung). Stuttgart: Klett-Cotta.
- Huber, Stephan G. (1999): School Effectiveness: Was macht Schulen wirksam? In: Schulmanagement, 30, Heft 2, S. 10-17.
- Jencks, C. (1973): Chancengleichheit. Reinbek: Rowohlt.
- Klafki, Wolfgang (1982): Zur Pädagogischen Bilanz der Bildungsreform. In: Die Deutsche Schule, 74, Heft 5, 339-352.
- Klafki, Wolfgang (1991): Perspektiven einer humanen und demokratischen Schule. In: Berg, Hans Christoph & Steffens, Ulrich (Hrsg.) (1991), a.a.O., 31-41.
- Kleinschmidt, Gottfried (1999): Bericht aus dem Diskussionsforum „Ringelgrün“ des internationalen Workshops „Evaluation und Qualität im Bildungswesen“. In: Grogger, Günther & Specht, Werner (Hrsg.), a.a.O., 97-102.
- Klemm, Klaus (1998): Steuerung der Schulentwicklung durch zentrale Leistungskontrollen? In: Rolff, Hans-Günter; Bauer, Karl-Oswald; Klemm, Klaus & Pfeiffer, Hermann (Hrsg.): Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven, Band 10. Weinheim & München: Juventa, 271-294.
- Klemm, K./Rolff, H.-G./Tillmann, K.-J. (1986): Bildung für das Jahr 2000. Bilanz der Reform, Zukunft der Schule. Reinbek bei Hamburg: Rowohlt.
- Klieme, Eckhard (2002): Was ist guter Unterricht? In: Bergsdorf, W. et al. (Hrsg.): Herausforderungen der Bildungsgesellschaft. 15 Vorlesungen. Weimar: Rhino Verlag, 89-113.
- Klieme, Eckhard (2003): Unterrichtsqualität und mathematisches Verständnis in verschiedenen Unterrichtskulturen. Frankfurt/M.: Deutsches Institut für Internationale Pädagogische Forschung (DIPF).
- Klieme, Eckhard; Baumert, Jürgen & Schwippert, Knut (2000): Schulbezogene Evaluation und Schulleistungsvergleiche. Eine Studie im Anschluss an TIMSS. In: Rolff, Hans-Günter; Bos, Wilfried; Klemm, Klaus; Pfeiffer, Hermann & Schulz-Zander, Renate (Hrsg.). Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven, Band 11, 387-419.
- Klieme, Eckhard; Avenarius, Hermann; Blum, Werner; Döbrich, Peter; Gruber, Hans; Prenzel, Manfred; Reiss, Kristina; Riquarts, Kurt; Rost, Jürgen; Tenorth, Heinz-Elmar & Vollmer, Helmut (2003): Expertise zur Entwicklung nationaler Bildungsstandards. Berlin: Deutsches Institut für Internationale Pädagogische Forschung (DIPF).
- Klitgaard, R. E. & Hall, G. R. (1973): A statistical search for unusually effective schools. Santa Monica, CA
- Köller, Franz (1991): Gute Schulen in der Sicht hessischer Schulverwaltung. Welchen Gestaltungsspielraum brauchen gute Schulen? In: Berg, Hans Christoph & Steffens, Ulrich (Hrsg.), a.a.O., S. 81-84.
- Köller, Franz (1992): Pädagogische und politische Überlegungen zu einem Hessischen Schulgesetz. Grundlegende Ansätze und wichtigste Akzente. In: Schindehütte, M. (Hrsg.): Schule in Hessen. Eigenverantwortung und Selbstverwaltung. Gestaltungsperspektiven für die kommenden Jahre (Hofgeismarer Protokolle, Heft 295). Hofgeismar: Evangelische Akademie, S. 6-19.
- Koetz, A. G. et al. (1994): Reorganisation der Staatlichen Schulaufsicht des Landes Nordrhein-Westfalen (Schlußgutachten). Düsseldorf: Kienbaum Unternehmensberatung.
- Kohlhoff, Walter (1995): Qualitätssicherung von Schule. Möglichkeiten einer externen Evaluation durch die Schulaufsicht. In: Buchen, H.; Horster, H. & Rolff, H.-G. (Hrsg.): Schulleitung und Schulentwicklung. Erfahrungen – Konzepte – Strategien. E 5.1. Stuttgart.
- Kotthoff, Hans-Gerog (2003): Bessere Schulen durch Evaluation? Internationale Erfahrungen (Studien zur International und Interkulturell Vergleichenden Erziehungswissenschaft, Band 1). Münster u. a.: Waxmann.
- Konsortium Bildungsberichterstattung (2006): Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration (im Auftrag der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland und des Bundesministeriums für Bildung und Forschung). Bielefeld: W. Bertelsmann Verlag.
- Kuhlmann, C. (1972): Lernen und Verwalten. Zusammenhänge von Bildungsreform und Bildungsverwaltung dargestellt am Strukturplan des Deutschen Bildungsrates (Gutachten und Studien der Bildungskommission, Bd. 23). Stuttgart: Klett.
- Lange, Hermann (1999): Qualitätssicherung in Schulen. In: Die Deutsche Schule, 91, Heft 2, 144-159.
- Lehmann, Rainer H. (1997): Überregionale Leistungsmessung und der individuelle Anspruch auf optimale Lernförderung. In: Empirische Pädagogik, 11, S. 85-94.

- Lehmann, Rainer H.; Peek, Rainer & Gänsfuß, Rüdiger (1996): Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lehmann, Rainer H.; Peek, Rainer; Gänsfuß, Rüdiger; Lutkat, Sabine; Mücke, Stephan & Barth, Ingola (2000):
- QuaSUM – Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg (Schulforschung in Brandenburg, Heft 1). Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- Lengen, M. (1989): Schulrat und Innovation. Handlungsspielräume der unteren Schulaufsicht. Aurich: Gesellschaft zur Förderung und Erforschung modellhafter Unterrichtsrealisation.
- Leschinsky, Achim & Cortina Kai S. (2003): Zur sozialen Einbettung bildungspolitischer Trends in der Bundesrepublik. In: Cortina, Kai S. et al. (Hrsg.): a.a.O., 20-51.
- Lightfoot, Sara (1983): The Good High School – Portraits of Character and Culture. New York: Basic Books.
- Liket, Theo M. E. (1993): Freiheit und Verantwortung. Das niederländische Modell des Bildungswesens. Gütersloh: Bertelsmann Stiftung.
- Lind, Georg (1993): Die „Gerechte Gemeinschaft-Schule“ – Erziehung in Demokratie. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 73-95.
- Madden, J. V./Lawson, D. R./Sweet, D. (1976): School Effectiveness Study. State of California.
- Maritzen, Norbert (1998): Schulinspektion: Qualitätssicherung im Regelkreis von Schulprogramm, interner und externer Evaluation. Das Bremer Modell. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 65-81.
- Markstahler, Jürgen (1999): Neue Konzepte der Qualitätssorge als Antwort auf die Krise des öffentlichen Schulwesens. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O. 21-33.
- Markstahler, Jürgen (2003): Vergleichsuntersuchungen und Schulqualität. In: Steffens, Ulrich & Messner, Rudolf (Hrsg.), S. 17-31.
- Martini, Renate (1997): „Schulautonomie“. Auswahlbibliographie 1989-1996 (Materialien zur Bildungsforschung, Band 1). Frankfurt/M.: Gesellschaft zur Förderung Pädagogischer Forschung & Deutsches Institut für Internationale Pädagogische Forschung.
- Mayer, Hilbert (2004): Was ist guter Unterricht? Berlin: Cornelsen.
- Melzer, Wolfgang (1997): Der Beitrag der Schulforschung zur Qualitätssicherung und Entwicklung von Schule. In: Erziehungswissenschaft, 8, Heft 16, 16-23.
- Merkens, Hans (2006): Pädagogische Institutionen. Pädagogisches Handeln im Spannungsfeld von Individualisierung und Organisation. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Messner, Rudolf (1998): Pädagogische Schulentwicklung zwischen neuer Lernkultur und wachsenden Modernisierungsdruck. In: Beiträge zur Lehrerbildung, 16, Heft 1, 86-110.
- Messner, Rudolf (2003): Konsequenzen der PISA-Ergebnisse für die Qualitätsentwicklung an Schulen – Überlegungen aus der Perspektive der Schulentwicklung. In: Steffens, Ulrich & Messner, Rudolf (Hrsg.): Macht PISA Schule? Perspektiven der Schulentwicklung (Folgerungen aus PISA für Schule und Unterricht, Heft 1). Wiesbaden: Hessisches Landesinstitut für Pädagogik (HeLP), 45-57.
- Messner, Rudolf (2004): Was Bildung von Produktion unterscheidet – oder : Die Spannung von Freiheit und Objektivierung und das Projekt der Bildungsstandards. In: Erziehung und Unterricht, 154, Heft 7-8, 693-716.
- Meyer, Hilbert (2004): Was ist guter Unterricht? Berlin: Cornelsen Verlag.
- New York State Department of Education (1974): Performance Review. School Factors Influencing Reading Achievement. A Case Study of Two Inner City Schools. New York: Office of Education.
- OECD (1991): Schulen und Qualität. Ein internationaler OECD-Bericht. Frankfurt/M. u.a.: Lang.
- OECD (1997): Prepared for life? How to measure Cross-Curricular Competencies. Paris.
- Oser, Fritz (1993): Lernen durch die Gestaltung des Schullebens – Der Ansatz der „Gerechten Gemeinschaft“. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 27-37.
- Peek, Rainer (2004): Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QuaSUM) – Klassenbezogene Ergebnisrückmeldungen und ihre Rezeption in Brandenburger Schulen. In: Empirische Pädagogik, 18, Heft 1, 82-114
- Peek, Rainer; Steffens, Ulrich & Köller, Olaf (2006): Positionspapier des Netzwerkes Empiriegestützte Schulentwicklung (EMSE) zu: Zentrale standardisierte Lernstandserhebungen. 5. EMSE-Tagung, Berlin, 8. Dezember 2006.
- Pekrun, Reinhard (2002): Vergleichende Evaluationsstudien zu Schülerleistungen: Konsequenzen für zukünftige Bildungsforschung. In: Zeitschrift für Pädagogik, 48, S. 111-128.
- Philipp, Elmar (1992): Gute Schule verwirklichen. Ein Arbeitsbuch mit Methoden, Übungen und Beispielen der Organisationsentwicklung. Weinheim & Basel: Beltz.
- Picht, Georg (1965): Die deutsche Bildungskatastrophe. München.
- Pieper, A. (1986): Verbesserung der Zusammenarbeit im Lehrerkollegium als Aufgabe einer systembezogenen schulpsychologischen Beratung. Frankfurt/M u. a.: Lang.
- Posch, Peter (1999a): Erfahrungen mit dem Inspektionssystem in England und Schottland. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), 108-121.
- Posch, Peter (1999b): Qualitätsevaluation und Qualitätsentwicklung im Schulwesen. In: Grogger, Günther & Specht, Werner (Hrsg.), a.a.O., 197-210.

- Posch, Peter & Altrichter, Herbert (1997): Möglichkeiten und Grenzen der Qualitätsevaluation und Qualitätsentwicklung im Schulwesen (Bildungsforschung des Bundesministeriums für Unterricht und kulturelle Angelegenheiten, Band 12). Innsbruck & Wien: StudienVerlag.
- Purkey, Stewart C. & Smith, Marshall S. (1983): Effective Schools: A Review. In: *The Elementary School Journal*, Heft 4, S. 427-452.
- Purkey, Stewart C. & Smith, Marshall S. (1990): Wirksame Schulen – Ein Überblick über die Ergebnisse der Schulwirkungsforschung in den Vereinigten Staaten. In: *Aurin* (1990), S.13-45.
- Reusser, Kurt; Pauli, Christine & Waldis, M. (2006): Mathematikunterricht und Mathematiklernen in Schweizer Schulen. Ergebnisse einer nationalen und internationalen Video-Studie. Zürich: Universität Zürich, Pädagogisches Institut.
- Richter, Ingo (1975): *Bildungsverwaltungsforschung*. In: Deutscher Bildungsrat (Hrsg.): *Bildungsforschung. Probleme - Perspektiven - Prioritäten. Teil 1* (Gutachten und Studien der Bildungskommission, Bd. 50). Stuttgart: Klett, 341-402.
- Roeder, Peter Martin (1987): Qualitative Schulentwicklung am Beispiel einer Berliner Gesamtschule: Zielkonflikte und unbeabsichtigte Nebenfolgen. In: Steffens, Ulrich & Bargel, Tino (Hrsg.) (1987c), a.a.O., 87-120.
- Roeder, Peter Martin (1992): Innere und äußere Schulreform. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 95-104.
- Rösner, Ernst (Hrsg.) (1999): *Schulentwicklung und Schulqualität*. Dortmund: Institut für Schulentwicklung, IFS-Verlag.
- Rolff, Hans-Günter (1991): Schulentwicklung als Entwicklung von Einzelschulen? Theorien und Indikatoren von Entwicklungsprozessen. In: *Zeitschrift für Pädagogik*, 37, Heft 6, 865-886.
- Rolff, Hans-Günter (1993): *Wandel durch Selbstorganisation. Theoretische Grundlagen und praktische Hinweise für eine bessere Schule*. Weinheim & München: Juventa.
- Rolff, Hans-Günter (Hrsg.) (1995): *Zukunftsfelder von Schulforschung*. Weinheim: Deutscher Studien Verlag.
- Rolff, Hans-Günter (1999a): Selbstorganisation, Gesamtsteuerung und Koppelungsprobleme. In: Risse, Erika & Schmidt, Hans-Joachim (Hrsg.): *Von der Bildungsplanung zur Schulentwicklung*. Rainer Brockmeyer zu ehren. Neuwied, Krieffel: Luchterhand, S. 63-72.
- Rolff, Hans-Günter (1999b): Pädagogisches Qualitätsmanagement: Schulentwicklung und Schulentwicklungsforschung vor neuen Herausforderungen. In: Rösner, Ernst (Hrsg.), a.a.O., 15-34.
- Rosenbusch, Heinz S. (1992): *Lehrer und Schulleiter - Ein strukturell gestörtes Verhältnis. Bericht über eine empirische Untersuchung zur Einschätzung der Schulaufsicht durch Lehrer* (Erziehungswissenschaftliche Forschungsbeiträge der Otto-Friedrich-Universität Bamberg). Bamberg: Otto-Friedrich-Universität Bamberg, Lehrstuhl für Schulpädagogik.
- Rosenbusch, Heinz S. (1994): *Lehrer und Schulleiter. Ein strukturell gestörtes Verhältnis. Berichte und organisationpädagogische Alternativen zur traditionellen Schulaufsicht*. Bad Heilbrunn: Klinkhardt.
- Rosenbusch, Heinz S. & Schlemmer, Elisabeth (1999): Die Rolle der Schulaufsicht bei der pädagogischen Entwicklung von Einzelschulen. In: Steffens, Ulrich & Bargel, Tino (Hrsg.), a.a.O., 52-62.
- Rumpf, H. (1966): *Die Misere der höheren Schule*. Neuwied: Luchterhand.
- Rumpf, H. (1968): Die administrative Verstörung der Schule. In: *Neue Pädagogische Bemühungen*, Band 30.
- Rumpf, H. (1969): Der Unterrichtsbeamte. Die aktuelle Diskrepanz zwischen Lehrerrolle und Lehreraufgabe. In: *Frankfurter Hefte*, Heft 6.
- Rutter, Michael; Maughan, Barbara; Mortimore, Peter; Ouston, Janet & Smith, Alan: *Fifteen Thousand Hours. Secondary schools and their effects on children*. Somerset: Open Books.
- Rutter, Michael; Maughan, Barbara; Mortimore, Peter; Ouston, Janet & Smith, Alan: *Fünfzehntausend Stunden - Schule und ihre Wirkung auf die Kinder*. Weinheim, Basel: Beltz.
- Scheerens, Jaap; Glas, Cees & Thomas, Sally M. (2003): *Educational Evaluation, Assessment, And Monitoring. A Systemic Approach*. Lisse u. a.: Swets & Zeitlinger.
- Scheerens, Jaap & Bosker, Roel J. (2005): *The Foundations of Educational Effectiveness*. Oxford u.a.: Pergamon.
- Scheffer, Walther (1977): *Kooperation von Lehrern und Unterrichtspraxis* (Beiheft 6 zu *curriculum konkret*). Frankfurt/M.: Regionale Lehrerfortbildung.
- Schlömerkemper, Jörg & Winkel, Klaus (1987): *Lernen im Team-Kleingruppen-Modell. Biographische und empirische Untersuchungen zum Sozialen Lernen in der Integrierten Gesamtschule Göttingen-Geismar* (Studien zur Bildungsreform, Band 14). Frankfurt/M. & Bern: Lang.
- Schöning, W. (1988): Schritte zu einer Beratungskultur der Schule durch schulinterne Lehrerfortbildung. In: *Die Deutsche Schule*, 81, Heft 1, 26-37.
- Schatz, Michael (1996): *Die Rolle der Schulaufsicht in der autonomen Schulentwicklung. Eine Untersuchung über Selbstbild, Rollenklärung und Fortbildungsbedarf im Hinblick auf die Autonomisierung des österreichischen Schulwesens*. Innsbruck: Universität Innsbruck, Institut für Erziehungswissenschaften.
- Skibba, Inge & Steffens, Ulrich (1995): *Schule im Spannungsfeld von schulaufsichtlicher Vorstrukturierung und autonomer Weiterentwicklung. Untersuchung im Rahmen des OECD-Projektverbands „The Effectiveness of Schooling and Educational Resource Management“*. Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung (HIBS).
- Specht, Werner & Thonhauser, Josef (Hrsg.) (1996): *Schulqualität. Entwicklungen, Befunde, Perspektiven*. Innsbruck & Wien: Studien Verlag.

- Steffens, Ulrich (1979):* Zur Problematik von Schulreformverläufen (Arbeitsmaterialien und Berichte zur Sekundarstufe I, Bd. 6). Paderborn u.a.: Schöningh.
- Steffens, Ulrich & Bargel, Tino (1987a):* "Qualität von Schule" - Ein neuer Ansatz der Schulentwicklung. In: Steffens, Ulrich & Bargel, Tino (Hrsg.): Erkundungen zur Wirksamkeit und Qualität von Schule (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 1). Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung, 1-20.
- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1987b):* Erkundungen zur Wirksamkeit und Qualität von Schule (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 1). Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.
- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1987 c):* Fallstudien zur Qualität von Schule (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 2). Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.
- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1987 d):* Untersuchungen zur Qualität des Unterrichts (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 3). Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.
- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1988):* Schulleben und Schulorganisation (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 4). Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.
- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1992):* Schulentwicklung im Umbruch. Analysen und Perspektiven für die zukünftige inhaltliche Gestaltung von Schule (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 6). Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung (HIBS).
- Steffens, Ulrich & Bargel, Tino (1993):* Erkundungen zur Qualität von Schule. Neuwied/Kriftel/Berlin: Luchterhand.
- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1998):* Schule zwischen Autonomie und Aufsicht (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 8). Wiesbaden: Hessisches Landesinstitut für Pädagogik (HeLP).
- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1999):* Qualitätsentwicklung und Qualitätssicherung von Schulen. Strategien, Instrumente und Erfahrungen (Beiträge aus dem Arbeitskreis Qualität von Schule, Heft 10). Wiesbaden: Hessisches Landesinstitut für Pädagogik (HeLP).
- Steffens, Ulrich & Boenicke, Rosemarie (2000):* Koordinatensysteme der Schulqualität. In: Haan, Gerhard de; Hamm-Brücher, Hildegard & Reiche, Norbert (Hrsg.): Bildung ohne Systemzwänge. Innovationen und Reformen. Neuwied: Luchterhand, 269-289.
- Steuer, E. (1983):* Organisationsentwicklung für die Schule. Leitbild, Strategie und Verwirklichung schulinterner Organisationsentwicklung. Frankfurt/M. u. a.: Lang.
- Stufflebeam, D. L. (1972):* Evaluation als Entscheidungshilfe. In: Wulf, C. (Hrsg.): Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen. München: Piper, 113-145.
- Szaday, Christopher (1995):* Creating effective schools for all: Trends in school effectiveness and school improvement research. Ebikon: Zentralschweizer Beratungsdienst für Schulfragen.
- Szaday, Christopher (1999):* Fallbeispiel Schulhaus Schachen. In: Grogger, Günther & Specht, Werner (Hrsg.), a.a.O., 231.
- Szaday, C.; Büeler, X. & Favre, B. (1996):* Trendbericht zur internationalen Schulqualitäts- und Schulentwicklungsforschung. Aarau: SKBF.
- Terhart, Ewald (2002):* Wie können Ergebnisse von vergleichenden Leistungsstudien systematisch zur Qualitätsverbesserung in Schulen genutzt werden? In: Zeitschrift für Pädagogik, 48, S. 91-110.
- Tillmann, K.-J. (1987):* Zwischen Euphorie und Stagnation. Erfahrungen mit der Bildungsreform. Hamburg: Bergmann & Helbig.
- Tillmann, Klaus-Jürgen (Hrsg.) (1989):* Was ist eine gute Schule? (PB Buch 13 der Zeitschrift Pädagogische Beiträge). Hamburg: Bergmann+Helbig.
- Tillmann, Klaus-Jürgen (1995):* Autonomie der Schule – Illusion oder realistische Perspektive? In: Tillmann, Klaus-Jürgen: Schulentwicklung und Lehrerarbeit – Nicht auf bessere Zeiten warten. Hamburg: Bergmann + Helbig, 31-45.
- Tillmann, Klaus-Jürgen (2001):* Leistungsvergleichsstudien und Qualitätsentwicklung oder: Auf dem Weg zu holländischen Verhältnissen? In: Journal für Schulentwicklung, 5, Heft 2, S. 9-17.
- Vierlinger, Rupert (Hrsg.) (1989):* Eine gute Schule – was ist das? 10. Europäisches Symposium an der Universität Passau vom 27. Juli bis 4. August 1989. Passau: Universität Passau.
- Weber, G. (1971):* Inner-City Children Can Be Taught To Read. Four Successful Schools. Washington, D. C.: Council for Basic Education.
- Weick, K. E. (1976):* Educational Organizations as Loosely Coupled Systems. In: Administrative Science Quarterly 21, 1-19.
- Weinert, Franz E. (1996):* Psychologie des Lernens und der Instruktion. Göttingen: Hogrefe.
- Weinert, Franz E. (1998):* Neue Unterrichtskonzepte zwischen gesellschaftlichen Notwendigkeiten, pädagogischen Visionen und psychologischen Möglichkeiten. In: Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (Hrsg.): Wissen und Werte für die Welt von morgen. München und Donauwörth: Auer, S. 101-125.
- Weinert, Franz E. (2001):* Perspektiven der Schulleistungsmessung – mehrperspektivisch betrachtet. In: Weinert, Franz E. (Hrsg.), a.a.O., 353-365.
- WEINERT, Franz E. (Hrsg.) (2001):* Leistungsmessungen in Schulen. Weinheim und Basel: Beltz.
- Weinert, Franz E. & Helmke, Andreas (1987):* Schulleistungen – Leistungen der Schule oder der Kinder? In:

- Steffens, Ulrich & Bargel, Tino (Hrsg.) (1987d), a.a.O., 17-31.
- Weinert, F. E. & Helmke, Andreas (Hrsg.) (1997): Entwicklung im Grundschulalter. Weinheim: Psychologie Verlags Union.
- WEISHAUPT, Horst (Hrsg.) (2000): Qualitätssicherung im Bildungswesen. Problemlage und aktuelle Forschungsbefunde. Dokumentation einer Tagung der Kommission Bildungsorganisation, Bildungsplanung, Bildungsrecht der Deutschen Gesellschaft für Erziehungswissenschaft am 10.-11. März 1999 in Erfurt (Erfurter Studien zur Entwicklung des Bildungswesens, Band 13). Erfurt: Pädagogische Hochschule.
- WEISS, Manfred (1985): Schuleffekt-Forschung: Ergebnisse und Kritik empirischer Input-Output-Untersuchungen. In: Twellmann, W. (Hrsg.): Handbuch Schule und Unterricht. Band 7.1. Düsseldorf: Schwann, 1060-1094.
- WELLENDORF, F. (1969): Formen der Kooperation von Lehrern in der Schule. In: Fürstenau, P. u.a. (Hrsg.): Zur Theorie der Schule. Weinheim/Basel: Beltz, 47-66.
- WENZEL, Hartmut; WESEMANN, Matthias & BOHN-SACK, Fritz (Hrsg.) (1990): Schulinterne Lehrerfortbildung. Ihr Beitrag zu schulischer Selbstentwicklung (Studien zur Schulpädagogik und Didaktik, Band 4). Weinheim & Basel: Beltz.
- Wolf, Hartmut & Klüsche, Norbert (1977): Abschlußbericht der Projektleitung für den Auftraggeber (Beiheft 9 zu curriculum konkret). Frankfurt/M.: Regionale Lehrerfortbildung.*
- Zedler, Peter (1990): Bilanz der Bildungsplanung. Veranlassungen, Problemtypen und Überforderungen staatlicher Reformplanung seit 1945 (Erziehungswissenschaftliche Studien und Berichte der Fernuniversität Hagen). Hagen: Fernuniversität, Fachbereich Erziehungswissenschaft.*