

## **Anliegen und Selbstverständnis der Fachtagungen zur "empiriegestützten Schulentwicklung (EMSE)"**

Arbeitspapier für die 3. EMSE-Tagung in Esslingen am 21.-22. Dezember 2005

Als vor über einem Jahr die erste EMSE-Tagung konzipiert wurde, war zwar intendiert, aber nicht abzusehen, dass ihr weitere Tagungen folgen werden. Die nunmehr dritte Tagung in Esslingen und die vorliegenden Zusagen für eine vierte und eine fünfte Tagung sind Anhaltspunkte dafür, dass es einen Bedarf für Zusammenkünfte dieser Art gibt und dass sich EMSE zu einem kontinuierlichen Arbeitszusammenhang entwickeln könnte.

Vor diesem Hintergrund und zur Information für EMSE-"Seiteneinsteiger" möchten wir – die Initiatoren – nochmals unser Anliegen skizzieren, das wir mit den EMSE-Tagungen verbinden. Dabei beziehen wir uns auf unser Einladungsschreiben zur ersten Tagung und auf die Einleitung zur Soester Tagung von Ulrich Steffens.

### **1 Anliegen der EMSE-Tagungen**

Im Zuge der "empirischen Wende" in der Bildungspolitik und Bildungsplanung hat sich mittlerweile ein breites Spektrum an empirisch orientierten Vorhaben zur Qualitätssicherung und Qualitätsentwicklung in den einzelnen Bundesländern entfaltet bzw. verstetigt:

- Fachbezogene Schulleistungstest: standardisierte Lernstandserhebungen unterrichtsfachlicher Leistungen in verschiedenen Jahrgangsstufen;
- überfachliche Schulleistungstests – zum Beispiel zum sozialen und kooperativen lernen, zum Problemlösen oder zur Selbstständigkeitsförderung;
- Sekundäranalysen zu internationalen und nationalen Datensätzen – zum Beispiel PISA 2000-Landesdatensätze; Verknüpfung von Ergebnissen aus internationalen, nationalen Vergleichsuntersuchungen, Systemmonitoring und Schulevaluation;
- fokussierte Evaluationen – zum Beispiel Analyse von Schwachstellen des Schulsystems; Auswertungen von Abschlussprüfungen oder schulinternen Vergleichsarbeiten; Studien zur Schulzeitverkürzung oder zur Diagnosekompetenz und zum förderorientierten Verhalten von Lehrpersonen; Arbeiten zur Lehrer- oder Schülerbeurteilung sowie Arbeitsplatzuntersuchungen;
- empirische Analysen von Bildungs- und Schulstatistiken ("Schulberichterstattung");
- Studien zur Schul- und Unterrichtsqualität – zum Beispiel ergänzende empirische Studien zur Klärung von Wirkungszusammenhängen zwischen Lehr- und Lernbedingungen sowie erzieherischen und fachlichen Leistungen ("Konfigurationen" von Schul- und Unterrichtsqualität); Forschungsvorhaben zur Ermittlung von Bewertungsmaßstäben der Schul- und Unterrichtsqualität ("Qualitätsindikatoren"); Identifikation relevanter Kontextbedingungen für schulische Lehr- und Lernprozesse ("Belastungsindikatoren");
- wissenschaftliche Begleitungen von Schulentwicklungsmaßnahmen – zum Beispiel externe Evaluation von Programmen und Wirksamkeitsanalysen;

- sonstige Vorhaben zur empiriegestützten Schul- und Unterrichtsentwicklung – zum Beispiel Transfer- und Folgenforschung; Studien zu Erfolg versprechenden Strategien der Ergebnismeldung an die Schulen; Studien zur Schulevaluation (u.a. Gelingenbedingungen schulinterner Evaluation, Analysen zur Schulinspektion bzw. externen Evaluation; Verzahnung interner mit externer Evaluation); bildungsökonomische Recherchen (Kosten-Nutzen-Analysen).

Alle diese und weitere – hier nicht aufgeführten – Vorhaben (z. B. zentrale Abschlussprüfungen, schulinterne Vergleichsarbeiten / Parallelarbeiten) sind von dem Anliegen getragen, 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 und in den einzelnen Bundesländern in der Formulierung von Kompetenzerwartungen oder verbindlichen Anforderungen in (Kern-) Lehrplänen für die Fächer zeigen. Unmittelbar daran gekoppelt spielen in der Diskussion Verfahren der Standardüberprüfung eine große Rolle. In diesem Zusammenhang haben sich bundesweit und in den einzelnen Bundesländern Lernstandserhebungen / Vergleichsarbeiten bzw. Orientierungsarbeiten etabliert, die die methodischen Prinzipien aus Schulleistungsstudien bzw. large scale assessments sensu PISA anwenden (standardisierte Testverfahren), gegenüber deren primärer Zielsetzung des Systemmonitorings aber vornehmlich auf Qualitätsentwicklung "vor Ort" zielen.

Wir möchten die EMSE-Tagungen zunächst auf dieses Feld schulischer Qualitätssicherung und Qualitätsentwicklung fokussieren. Dabei soll es darum gehen,

- eine Diskussion über dieses relativ neue Instrument der Qualitätssicherung zu führen,
- einen Einblick in laufende und geplante Projekte der Standardsetzung und Standardüberprüfung in den Bundesländern zu bekommen und
- erste Erfahrungen mit dem Instrumentarium auszutauschen.

Als Arbeitsmodus schlagen wir einen fachlichen Dialog auf der so genannten "operativen Ebene" vor (Workshopcharakter). Uns geht es dabei nicht um eine Erörterung bildungspolitischer Positionen und deren Legitimationsanstrengungen; vielmehr wollen wir – in einem offenen, um die Sache ringenden Austausch – unterschiedlicher Herangehensweisen kennen lernen, die Implikationen verschiedener methodischer Verfahren eruieren und den Handlungsbedarf an Erfolg versprechenden Ansätzen und Konzepten für die Schul- und Unterrichtsentwicklung ermitteln.

Die Konzeption der Tagungen sieht drei Themenbereiche vor, die sich nach den Erfahrungen der ersten beiden Tagungen bewährt haben:

- (1) Ein Schwerpunktthema des gastgebenden Landes,
- (2) ein "Brennpunktthema", das den EMSE-Teilnehmern von besonderem Interesse ist, sowie
- (3) Länderberichte und Erfahrungsaustausch.

Im Folgenden einen Auszug aus der Einleitung von *Ulrich Steffens* zur ersten EMSE-Tagung in Soest. Der Text soll den Handlungshintergrund und den Handlungsbedarf von EMSE verdeutlichen helfen.

## 2 Die Zeit der empirischen Wende in der Bildungspolitik und Bildungsplanung

Die Konstanzer Beschlüsse der KMK markieren den Ausgangspunkt einer Wende in der Bildungspolitik und Bildungsplanung. So wurde nicht nur festgelegt, dass sich Deutschland an regelmäßigen internationalen Untersuchungen beteiligen und ihnen einen Stellenwert für bildungspolitisches Handeln einräumen wird – namentlich ging es dabei um PISA –, sondern darüber hinaus auch verabredet, unter den Bundesländern Vergleichsuntersuchungen durchzuführen. Dieser Beschluss war insofern kühn, weil mit erheblichen Akzeptanzproblemen innerhalb der eigenen Bildungsverwaltung, in breiten Schichten der Lehrerschaft, aber auch unter Lehrerfortbildnern, Schulentwicklern und in der Erziehungswissenschaft zu rechnen war. Nach BAUMERT (2001, S.24) lassen sich die Vorbehalte gegenüber standardisierten Leistungsmessungen „im Wesentlichen auf zwei Basiseinwände zurückführen: Der erste Einwand behauptet einen Widerspruch zwischen Ganzheitlichkeit von Bildungsprozessen und den eingeschränkten Fragestellungen von Evaluationsmaßnahmen, insbesondere wenn sie mittels standardisierter Leistungserhebungen erfolgten. Der zweite Einwand besagt, dass standardisierte Untersuchungen oftmals den ihnen zu Grunde liegenden Bildungsbegriff oder das voraussetzende Fachverständnis nicht explizieren und Bildungsqualität letztlich mit dem in eins setzen, was ein Test erfasse.“

In diesem Zusammenhang muss auch die Initiative „Entwickeln statt Vermessen“ gesehen werden, die maßgeblich von Lehrerfortbildnern, Schulberatern und diesen Handlungsfeldern nahe stehenden Erziehungswissenschaftlern getragen wurde. Mit dem Slogan dieser Initiative verband sich zudem die Absicht, auf die Gefahren aufmerksam zu machen, dass durch eine Konzentration auf das Messen die notwendigen Investitionen in Schulentwicklung und Lehrerfortbildung zu kurz kommen könnten. So Unrecht hatten diese Initiatoren nicht, wenn man sich vor Augen führt, dass die Lehrerfortbildungs- und Schulberatungssysteme in einigen Bundesländern merklich zurückgebaut wurden.

Ich möchte mit diesen Hinweisen allerdings keine Debatte über das Für und Wider und die Alternativen von Leistungsmessungen und Vergleichsuntersuchungen auslösen. Zumal aus meiner Sicht sich nicht die Alternative Entwickeln *oder* Vermessen stellt; vielmehr muss die Formel lauten: Entwickeln *und* Vermessen – was im Übrigen auch das Begriffspaar Qualitätsentwicklung und Qualitätssicherung zum Ausdruck bringt.

Der Ausgangspunkt für diese eingeflochtenen Hinweise war das erwartete Akzeptanzproblem, den der KMK-Beschluss auslösen würde. Was bewog die KMK, trotzdem so zu entscheiden? Einer der Kultusminister, der maßgeblich am Zustandekommen dieses Beschlusses beteiligt war, gab mir in einer kleinen Gesprächsrunde einmal zu verstehen, dass man sich „mit dem Rücken an der Wand“ gefühlt hatte.

Um diese Äußerung besser verstehen zu können, bedarf es meines Erachtens eines Rückblicks auf die vorausgegangenen Entwicklungen. So liegen zwischen der von Staatsrat Lange proklamierten „empirischen Wende in der Bildungspolitik“ und jener von Heinrich Roth ausgerufenen „empirischen Wende in der Pädagogik“ mehrere Jahrzehnte. In dieser Zwischenzeit ist Vieles geschehen, was als Nährboden – also die Voraussetzungen – für die aktuellen Entwicklungen in der Bildungspolitik und Bildungsplanung anzusehen sind.

Obwohl die empirische Bildungsforschung bereits vor über drei Jahrzehnten Einlass in bildungsplanerische Handlungszusammenhänge erhalten hatte,

- beispielsweise durch die Gutachten und Studien des Deutschen Bildungsrates,
- beispielsweise im Rahmen der Strukturreformen in den 1970er Jahren (maßgeblich die empirische Erforschung des Gesamtschulsystems),
- beispielsweise in den verschiedenen Programmen der Bund-Länder-Kommission (BLK) mit seinen Modellversuchen,

blieben trotz dieser Handlungszusammenhänge empirische Befunde als Planungsparameter für bildungspolitische und bildungsplanerische Entscheidungen weitgehend randständig. So scheiterte beispielsweise Ingenkamps Initiative für bundesweite Leistungsvergleiche in den 1980er Jahren am Widerstand der Kultusministerien.

Warum aber haben empirische Befunde erst jetzt einen Stellenwert für die Bildungsverwaltung erlangt? So hatten ja bereits Ende der 1970er Jahre große empirische Schulforschungsprojekte relevante Erkenntnisse für die Bildungsplanung und Schulgestaltung bereit gestellt – 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 große handlungspraktische Bedeutung für die Bildungsverwaltung erfahren hatten, wohl aber einen gewissen „Depot-Effekt“ ausgelöst hatten, der vermutlich erst seit Mitte der 1990er Jahre Wirkung zeitigt. Von entscheidender Bedeutung erwies sich dabei der Befund über die Unterschiedlichkeit einzelner Schulen trotz einheitlicher makro-organisatorischer Vorstrukturierungen durch die Bildungsverwaltung bzw. Schulaufsicht. Auf diesen Sachverhalt hatte erstmals in Deutschland die Konstanzer Schulforschung unter Leitung von Helmut Fend hingewiesen. Dieser Sachverhalt war auch dann der Ausgangspunkt für eine Schulqualitätsbewegung, wie sie der „Arbeitskreis Qualität von Schule“ ausgelöst hatte (vgl. STEFFENS & BARGEL 1987). „Schrittmacher“ dieser Entwicklung in Deutschland – so Jürgen Baumert und Eckhard Klieme – waren der bereits erwähnte Arbeitskreis Qualität von Schule und das Institut für Schulentwicklungsforschung in Dortmund (BAUMERT 2001; KLIEME u.a. 2000).

Meine These ist, dass die Unterschiedlichkeit der einzelnen Schulen, auch und gerade zwischen ein und derselben Schulform, also die unterschiedliche Güte von Schulgestalten, den Auslöser für ein Umdenken und damit für einen Bewusstseinswandel innerhalb der Führungsebenen der Bildungsverwaltung darstellt. Die unterschiedliche Güte, also die unterschiedliche Realisierung von schulaufsichtlichen Vorgaben – neudeutsch: „Input-Steuerung“ – war nicht mehr zu übersehen und konnte ganz einfach nicht weiterhin ignoriert werden. Die so genannte Input-Steuerung war anerkanntermaßen an ihre Leistungsgrenze gestoßen: Schulen realisierten die rechtlichen Vorgaben auf jeweils ihre eigene Weise, ohne dass die Prozess-Steuerung in Form schulaufsichtlicher Kontrolle eine Chance sah, zu intervenieren. Allenfalls bei extremen Verstößen dienst- oder rechtsaufsichtlicher Belange – so scheint es – konnte Schulaufsicht noch kontrollierend eingreifen. Im übrigen wollten viele Aufsichtspersonen diese Funktion auch nicht ausüben, sahen sie sich doch in ihrem Selbstverständnis zunehmend in der Rolle des beratenden und unterstützenden Schulaufsichtsbeamten, wie die entsprechenden Positionspapiere und empirischen Untersuchungen belegen (vgl. STEFFENS & BARGEL 1998). Empirische Studien zeigten aber gleichzeitig, dass trotz dieses Selbstbildes von Schulaufsicht das Fremdbild damit nicht übereinstimmte. Schulaufsicht wurde nämlich – sowohl von Schulleitungspersonen als auch von der Lehrerschaft – weiterhin primär als Kontrollinstanz identifiziert, obwohl die Schulaufsicht faktisch diese unterstellte Kontrollfunktion nur noch in einschränkender Weise ausgeübt hatte.

Diese hier nur angedeuteten Argumente mögen vielleicht ausreichen, um deutlich zu machen, dass es maßgeblich zwei „Schubkräfte“ in unserem Schulsystem sind, die im Wesentlichen ein Umdenken in der Systemsteuerung begünstigt haben bzw. begünstigen:

- erstens der Sachverhalt über die Unterschiedlichkeit von Schulen, die die eingeschränkten Steuerungsmöglichkeiten im so genannten Input-Bereich erfahrbar werden ließen;
- zweitens die begrenzten Möglichkeiten der Prozess-Steuerung aufgrund der praktizierten Formen und Strukturbedingungen der Schulaufsicht.

Die Einsicht in die begrenzte Steuerbarkeit schulischer Gestaltungsprozesse durch makroorganisatorische Vorstrukturierungen (beispielsweise in Form von Lehrplänen, von Ausführungsvorschriften für die Arbeits- und Lernorganisation) einerseits und die Wahrnehmung der geringen Wirksamkeit schulaufsichtlichen Vollzugs- und Kontrollhandelns andererseits führten – gelinde gesagt – zu dem Verdacht (sozusagen als generalisierende Schlussfolgerung), dass die klassischen Steuerungsinstrumente zumindest in ihrer derzeitigen Ausprägung und Qualität nicht mehr hinreichend greifen. Im Grunde befand sich damit die Schulsystemsteuerung mitten in einer Legitimationskrise.

Auf der Suche nach Alternativen setzte darauf hin ein Öffnungsprozess ein, der bisher vernachlässigte Referenzsysteme und Erfahrungsbereich zu einer größeren Beachtung verhalf. Dadurch fanden die empirischen Wissenschaften einerseits und die internationalen Entwicklungen andererseits ihre Relevanz für bildungspolitisches und bildungsplanerisches Handeln. Nun erhielten auch Ergebnisse aus internationalen Vergleichsuntersuchungen, anders als noch wenige Jahre zuvor, Beachtung. 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 internationalen Perspektive gerieten auch die Steuerungssysteme anderer Staaten in den Blick. Dies betraf alternative Instrumente der Produktsteuerung durch Leistungstests, für die es ja bislang in Deutschland keine Tradition gab. So setzen Schulsysteme anderer Staaten teilweise bereits seit vielen Jahren standardisierte Tests ein, um Wissen über die Leistungsfähigkeit des Schulwesens insgesamt oder einzelner Schulen zu erhalten. Die Leistungstests stießen dabei deshalb auf eine besondere Resonanz, weil sie ein Instrumente aus einem für Deutschland neuen Steuerungsbereich, nämlich dem der Produkt- oder Output-Steuerung, stammen und möglicherweise – so die Hoffnung –, das Steuerungsproblem im Input- und Prozess-Bereich kompensieren könnten.

Spätestens für dieses Stadium der einsetzenden Entwicklung könnte die „Wiederentdeckung der Systemebene“ als Handlungsregulativ konstatiert werden. Ich spiele mit der Bezeichnung „Wiederentdeckung“ auf einen Topos der Schulqualitätsforschung an, nämlich auf die von Helmut Fend stammende Formel der „Wiederentdeckung der einzelnen Schule als pädagogischer Handlungseinheit“ (FEND 1987). Mit dieser Wiederentdeckung der Systemebene ist meines Erachtens eng die Einsicht in die Steuerungsnotwendigkeit in allen drei Bereichen (Input-, Prozess- und Produktbereichen) verbunden, wobei es entscheidend darauf ankommt, dass diese Trias in einem koordinierten, konsistenten und kontinuierlichen Gesamtzusammenhang verwoben wird. Gleichzeitig bedarf es dabei einer Neukonzipierung der Systemsteuerung, die ein ausbalanciertes Verhältnis von Qualitätssicherung und Qualitätsentwicklung anzustreben hat. Denn dieses Verhältnis ist nach meinen Beobachtungen in eine Schiefelage zu Gunsten der Kontrollstrategien und zu Lasten der Entwicklungsstrategien im

Sinne von Schulentwicklung und Schulberatung geraten. Schulleistungstests stoßen in der Lehrerschaft nur dann auf Akzeptanz, wenn sie konstitutiv mit Praxisunterstützungen für die Schul- und Unterrichtsentwicklung verkoppelt werden.

Die Zeit von 1997 bis 2004 war das Thema meines zweiten Teils. In diesen sieben Jahren ist die empirische Wende vollzogen worden. Wir müssen aufpassen, dass sie uns nicht vergessen lässt, dass die Sau vom Wiegen allein nicht fetter wird. Damit ist bereits das Thema für meinen dritten Teil aufgerufen, in dem ich – aus Zeitgründen – nur noch kurz einige zentrale Herausforderungen aus meiner Sicht skizzieren möchte.

### **3 Die proklamierte empirische Wende in der Bildungspolitik und Bildungsplanung ist eingetreten – was nun?**

Norbert Maritzen hatte in 1998 für die Amtschefkommission Qualitätssicherung eine Bestandsaufnahme zu Verfahren der Qualitätsevaluation im Schulwesen durchgeführt und ausgewertet (MARITZEN 1998). Vergleicht man die Ergebnisse von damals mit der aktuellen Situation in den Bundesländern, so belegen die zwischenzeitlichen Entwicklungen meine These, dass die empirische Wende eingetreten ist: Bundesländer-Vergleiche werden mit PISA 2003-E zum zweiten Mal durchgeführt, und viele Bundesländer haben zwischenzeitlich standardisierte Lernstandserhebungen eingeführt, andere Länder stehen kurz davor.

Trotz der empirischen Wende sind wir aber nach wie vor empirisch nicht so weit, um erklären zu können, warum beispielsweise Baden-Württemberg bei PISA besser als Bremen abgeschnitten hatte und warum Hessen auch in PISA-Hinsicht in der Mitte Deutschlands liegt. So wissen wir nicht, warum der Landkreis Fulda immer besonders gut bei den hessischen Lernstandserhebungen in Mathematik abschneidet. Es gibt sicherlich einige plausible Hypothesen. Eine davon betrifft den göttlichen Segen durch den dortigen Erzbischof. Diese Vermutung ist vielleicht deshalb gar nicht so falsch, weil sie auf regionsspezifische Sozialisationsmuster verweist.

Mit diesen Hinweisen soll deutlich werden, dass wir noch nicht über einen theoretischen Referenzrahmen oder ein Bezugssystem zur Erklärung von Schulleistungsunterschieden verfügen. Erklärungen dieser Art sind deshalb von praktischer Bedeutung, weil sie uns Hinweise auf die Güte von Schulsystemen oder von einzelnen Schulen geben können. Selbst eine Schule mit guten Testleistungen ihrer Schüler muss deshalb noch nicht eine gute Schule sein. So könnte sie, gemessen an den Lernvoraussetzungen ihrer Schüler, unterhalb der Leistungserwartungswerte liegen. Trotz vergleichsweise guter Ergebnisse bliebe eine solche Schule unterhalb ihrer Leistungspotenziale. Umgekehrt könnte eine Schule mit niedrigen Testwerten in Anbetracht ihrer hoch problematischen Schülerzusammensetzung deutlich über den Leistungserwartungswerten liegen. Welche Schule ist nun besser? In einem Ranking-Verfahren, das die Kontextbedingungen der Lehr- und Lernprozesse nicht berücksichtigt, hat die Schule mit den besseren Leistungsdaten in der Öffentlichkeit ein höheres Ansehen – meines Erachtens zu Unrecht, weil die Indikatoren für Schulleistungen nicht in Relation zu den Leistungspotenzialen einer Schule gesehen werden. Hier besteht eine große Herausforderung für zukünftige Lernstandserhebungen, nämlich die Kontextbedingungen schulischer Lehr- und Lernprozesse als mit in den Blick zu nehmen.

Eine weitere Herausforderung sehe ich darin, neben fachlichen Testleistungen auch Schülerleistungen im so genannten „überfachlichen Bereich“ zu erfassen. Bei PISA werden Leistungen dieser Art als „cross curricular competencies“ bezeichnet und betreffen soziale Kompeten-

zen und Problemlösen. Diese Lernleistungen müssten in zukünftigen Erhebungen ebenfalls einen Stellenwert erhalten.

Ferner ist es meines Erachtens für Schulevaluationen unabdingbar, auch die Lehr- und Lernbedingungen in Unterricht und Schule in den Blick zu nehmen und die Qualitätsfrage auch auf die schulischen Gestaltungsprozesse zu beziehen.

Insgesamt betrachtet, benötigen wir für Schulevaluationen einen Kriterienrahmen und Indikatoren für Schul- und Unterrichtsqualität.

Ein eigenes Kapitel der erforderlichen Entwicklungen müsste sich auf die Ausschärfung von Funktionen und Aufgabenrollen bei der Qualitätssicherung und Qualitätsentwicklung beziehen. Etwa Abgrenzungen zwischen wissenschaftlichen Untersuchungen, System- und Schul-Monitoring, Schulevaluationen unterschiedlichster Machart sowie Programm- und Wirkungsstudien. Auch was die Schulevaluation anbelangt, bedarf es einer weiteren Klärung ihrer Funktionen. Denn von der jeweiligen Funktion ist auch der Bearbeitungsmodus abhängig. Geht es um Legitimation und Rechtfertigung, so sind andere Vorgehensweisen angeraten, als wenn es um die Erzeugung von Steuerungswissen geht – sei es mit dem Interesse der Kontrolle oder dem der Entwicklung. Auch die Frage der beauftragenden Instanz bzw. die Frage der Datenhoheit ist für den zu wählenden Evaluationsmodus von erheblicher Bedeutung.

Die größte Herausforderung im Zusammenhang mit Schulevaluationen und Lernstandserhebungen sehe ich gegenwärtig in der Ausarbeitung eines klugen Designs, das die unterschiedlichen Interessen und Absichten ausbalanciert. Systemmonitoring ist zwar ein wichtiges Instrument der verantwortlichen Instanz in der Bildungsverwaltung, liefert aber den Schulen in der Regel nur indirekt Orientierungen und Hilfen für die eigene Weiterentwicklung. Wie müssen schulspezifische Ergebnismeldungen angelegt sein, dass sie von den Schulen rezipiert und konstruktiv genutzt werden können? Welche flankierenden Maßnahmen der Unterstützung, der Prozess- und Fachberatung sind dabei erforderlich? Und wie müssen interne und externe Evaluation aufeinander bezogen sein, damit sie synergetisch genutzt werden können?

Ich wünsche mir sehr, dass sich unsere Tagungen auch mit diesen angesprochenen Herausforderungen und Fragen beschäftigen werden.

## Literatur

Baumert, Jürgen (2001):  
Vergleichende Leistungsmessung im Bildungsbereich. 43. Beiheft der Zeitschrift für Pädagogik: Zukunftsfragen der Bildung. Hrsgg. v. J. Oelkers. Weinheim und Basel: Beltz.

Fend, Helmut (1987):  
„Gute Schulen – schlechte Schulen“ – Die einzelne Schule als pädagogische Handlungseinheit. In: Steffens, Ulrich und 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 (HIBS), 1987, S. 55-79.

Klieme, Eckhard, Baumert, Jürgen und Schwippert, Knut (2000):  
Schulbezogene Evaluation und Schulleistungsvergleiche. Eine Studie im Anschluss an TIMSS. Jahrbuch der Schulentwicklung, Bd. 11. München: Juventa.

Maritzen, Norbert (1998):

Bestandsaufnahme zu Verfahren der Qualitätsevaluation im Schulwesen. Berlin: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.

Steffens, Ulrich & Bargel, Tino:

„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 (HIBS), S. 1-20.

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).



ANLAGE:

**Fachtagungen „Empiriegestützte Schulentwicklung“:**

**Fachbezogene Schulleistungstests: standardisierte Lernstandserhebungen unterrichtsfachlicher Leistungen in verschiedenen Jahrgangsstufen –  
Checkliste: Mögliche Gesichtspunkte für Länderberichte**

## **1 Entwicklungsstadium**

In Vorbereitung? In der Entwicklung und Erprobung? Ausgearbeitete Testinstrumente? Regeleinsatz?

In welchem Jahr wurde mit der Entwicklungsarbeit standardisierter Leistungserhebungen begonnen?

## **2 Verwendungszweck und „funktionale Ebene“ der Evaluationsmaßnahmen**

- Verwendungszweck:

Zur Verortung der Leistungsfähigkeit des Schulsystems? Als Steuerungsdaten für bildungsplanerische Entscheidungen? Als Orientierungs-, Normierungs- oder Korrektivfunktion für Schulen (benchmarking)? Als Informationsquelle für fach-, dienst- oder rechtsaufsichtliche Maßnahmen? Zur Bewertung der Schülerleistungen (Klassenarbeiten)?

- Zuständigkeit (Datenhoheit) auf Schulebene (dezentral): Individuelle Bewertung und Zertifizierung von Schülerinnen und Schülern? Schulinterne Evaluation (zur Optimierung von Schulentwicklung)? Externe Evaluation von Programmen?

- Zuständigkeit (Datenhoheit) auf Systemebene (zentral): Individuelle Zertifizierung durch zentrale Tests oder Prüfungen mit Berechtigungsfunktion? Flächendeckende Evaluation von Einzelschulen (bzw. ausgewählten Jahrgängen an Einzelschulen)? Systemmonitoring (auf Stichprobenbasis)?

## **3 Beteiligte Bezugsgruppen**

- Trägereinrichtung, die für die Lernstandserhebungen zuständig ist (z.B. Ministerium, Landesinstitut, Beauftragung von Wissenschaftlern bzw. Forschungseinrichtungen; Mischformen)

- Einbeziehung von Gutachtern / Experten (z.B. Fachdidaktiker, Testentwickler, erfahrene Lehrpersonen, Lehrplan-Verantwortliche)

- Anhörung von Interessengruppen?

## **4 Konzeptionelle Ausrichtung der Tests**

- Norm- bzw. Lehrplanorientierung?

- Wissensbasierte Tests? Anwendungs-, Problemorientierung?

- Fähigkeits-, kompetenzorientierte Tests (z.B. PISA-Ansatz)?

- Orientierung an Mindeststandards, Regelstandards, KMK-Standards?

- Erfassung des Leistungsspektrums eines Faches: Weit- oder Engführung (z.B. Basiskompetenzen)?

## 5 Fächer, Jahrgangsstufen, Schulformen

mit Bezeichnung des jeweiligen Tests

## 6 Methodischer Anspruch der Testentwicklung

- pragmatisch entwickelte Aufgaben (z.B. von Lehrergruppen)?
- Verankerung der Tests im internationalen, nationalen oder regionalen Vergleich? Übernahme von Testaufgaben aus anderen Erhebungen (z.B. TIMSS, PISA)?
- Skalierung nach Kompetenzstufen, Leistungsdimensionen; testtheoretische Konstrukte?
- Evaluation der Testverfahren? Validitäts- und Reliabilitätsprüfungen?
- Berücksichtigung von Kontextvariablen bzw. Wirkungszusammenhängen: außerschulische Umfeldbedingungen (Elternhaus)? Schulische Rahmenbedingungen? Unterrichtliche Lehr- und Lernbedingungen?

## 7 Erhebungsmodus

- Erhebungsturnus: Jährlich, zweijährlich, ..., sporadisch? Wechselnde Jahrgangsstufen und Fächer?
- Freiwillige oder verpflichtende Teilnahme?
- Repräsentative Landesstichprobe? Landesweit verpflichtend? Fakultatives Angebot für Schulen (z.B. für schulinterne Vergleichsarbeiten)?

## 8 Beobachtungseinheiten bei der Datenerfassung und Aggregationsebenen der Ergebnisdarstellung

Schülerebene („Individual-N“)? Klassen-, Jahrgangsebene? Schulebene, kommunale Ebene, regionale Ebene (Schulaufsichtsbereich)? Landesebene?

## 9 Art der Ergebnisveröffentlichung und ihre Funktionen

- öffentliche Darstellung von klassen- oder schulbezogenen Ergebnissen („Ranking“)? „Best practice“-Veröffentlichungen (nur die Leistungsbesten werden namentlich genannt)? Öffentliche Darstellung von Ergebnissen in aggregierter Form (z.B. Schulformvergleiche, Stadt-Land-Befunde, Landesergebnisse)?
- Ergebnisberichte nur für den dienstinternen Gebrauch? (z.B. analog zu den Ergebnisrückmeldungen an die PISA-Schulen, z.B. Ergebnisrückmeldung nur an den jeweiligen Fachlehrer der betreffenden Klasse, z.B. Ergebnisrückmeldung an die zuständige Schulaufsicht)

## 10 Verwendungszusammenhang der Ergebnisse: Stellenwert für die Schul- und Unterrichtsentwicklung

-Strategien zur Ergebnisdarstellung und -verarbeitung auf Landesebene und Schulumtsebene? (z.B. im Zusammenhang mit Bildungsstandards, Schulberichterstattung, Schulentwicklungsmaßnahmen)

- Strategien zur Ergebnisrückmeldung und -verarbeitung auf Schulebene? (z.B. Verkoppelung mit Schul- und Fachberatung, externe Unterstützung bei der Schulprogrammarbeit und schul-internen Evaluation)
- Wie werden die vorliegenden Untersuchungsergebnisse genutzt? (Als Steuerungsdaten für welche bildungsplanerische Entscheidungen? Als Referenzdaten für welche Form der Qualitätsevaluation auf Schulebene? Als Bezugspunkt/ Anlass für welche fach-, dienst-, rechtsauf-sichtlichen Interventionen?)