

**Thorsten Bohl (2009): Fazit: Unter welchen Bedingungen ist Lernen aus Evaluationsergebnissen möglich?**

Langversion des Fazits aus:

**Thorsten Bohl/Hanna Kiper (Hrsg.): Lernen aus Evaluationsergebnissen – Verbesserungen planen und implementieren. Bad Heilbrunn: Klinkhardt**

URL:  
[http://www.erziehungswissenschaft.uni-tuebingen.de/Abteilungen/Schulpaedagogik/Personal/Bohl\\_Dateien/index.html](http://www.erziehungswissenschaft.uni-tuebingen.de/Abteilungen/Schulpaedagogik/Personal/Bohl_Dateien/index.html)



## **Gliederung der Langversion**

Lernen aus Evaluationsergebnissen: zwischen Einstellungen und konkret veränderten Unterrichtsarrangements	2
Wie müssen Rückmeldeformate aus Evaluationsergebnissen aussehen, damit Lehrkräfte mit ihnen nachhaltig arbeiten?	4
Lehrerinnen und Lehrer und Kollegien innerhalb der Organisation Schule: Innovationspotentiale im Spiegel einer Schulbiographie	5
Unterricht konzeptionell verändern: Innovationspotentiale im Spiegel individueller Möglichkeiten	9
Aufgabentypen zwischen vergleichender Bildungsforschung und individuell-didaktischem Handeln	10
Lernen aus Evaluationsergebnissen: Günstige Bedingungen	14
Zusammenfassung und Schlussbemerkung	20
Literatur	24



*Thorsten Bohl*

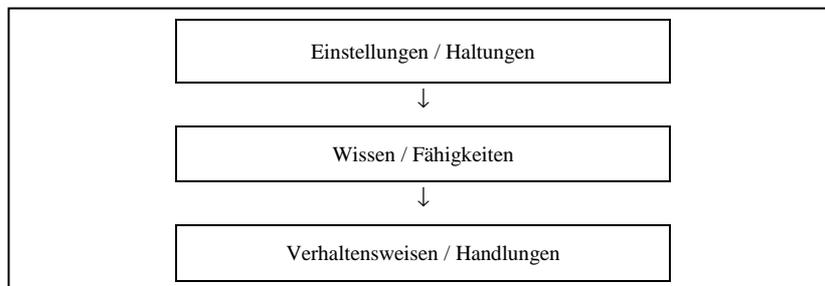
## **Fazit: Unter welchen Bedingungen ist Lernen aus Evaluationsergebnissen möglich?**

In diesem letzten Kapitel werden die bisherigen Beiträge zusammengeführt. Einzelne bisher thematisierte Themenbereiche werden aufgegriffen und vertieft. Ziel dieses Beitrages ist es, die Bedingungen herauszuarbeiten, die günstig erscheinen, um die Lücke zwischen vorliegenden Evaluationsergebnissen und der Implementierung von Innovationen zu schließen.

Das Thema ‚Lernen aus Evaluationsergebnissen‘ erfolgt in diesem Band aus handlungstheoretischer Perspektive (Beiträge von Hacker und Wahl), in national und international vergleichender Perspektive (Beiträge von Hovestadt und Lind), aus organisationspsychologischer und sozialpsychologischer Perspektive (Beitrag von Rosenstiel und Steins), aus der Perspektive der Rezeptionsforschung (Beiträge von Kuper/Schneewind; Neubrand; Kohler; Maier), aus der Perspektive der Schulleitung und Schulaufsicht (Beiträge von Märkl; Ballasch; Bosen), aus professionstheoretischer Perspektive (Beiträge von Bauer; Miller; Landert; Kieschke), aus der Perspektive des Qualitätsmanagements (Beiträge von Böttcher), mit Blick auf das Qualitätsmanagement eines Bundeslandes (Beitrag von Kiper), sowie aus der Perspektive einer Versuchsschule (Beitrag von Hollenbach). Aus der Vielzahl an Überlegungen, Befunden und Diskussionsbeiträgen greife ich im Folgenden zentrale Aspekte heraus. Als „zentral“ erachte ich ein Thema dann, wenn es in der bisherigen Diskussion vernachlässigt wurde oder als besonders relevant für erfolgreiches Lernen aus Evaluationsergebnissen erscheint.

## Lernen aus Evaluationsergebnissen: zwischen Einstellungen und konkret veränderten Unterrichtsarrangements

Mit Blick auf einzelne Lehrerinnen und Lehrer werden in den Beiträgen verschiedene Ebenen angesprochen (Abb. 1), es geht um Einstellungen (gegenüber Evaluationen), um Konzepte (zum Umgang mit Evaluationsergebnissen, z. B. Rückmeldeformate) und um konkrete Handlungsmöglichkeiten (im alltäglichen unterrichten und schulischen Geschehen).



**Abb. 1:** Ebenen des Umgangs mit Evaluationsergebnissen aus individueller und sozialpsychologischer Perspektive

Die drei Ebenen sind hilfreich, weil sie verdeutlichen, an welche Ebene sich verschiedene Maßnahmen richten bzw. auf welcher Ebene Entwicklungsbedarf besteht. Charakteristika und Veränderungsmöglichkeiten der drei Ebenen sind unterschiedlich.

*Einstellungen* sind relativ stabile und erfahrungsbasierte psychische Reaktions-tendenzen auf Situationen (vgl. Stahlberg/Frey 1992, S. 144ff), im Gegensatz zu Meinungen handelt es sich bei Einstellungen um positive oder negative Bewertungen. Sie sind wesentlich für die Erklärung und Vorhersage von Verhaltensweisen. Einstellungen von Lehrerinnen und Lehrern gegenüber Evaluationen sind geprägt von den Erfahrungen, die sie damit hatten. Da Evaluationen vielerorts in den vergangenen Jahren erst eingeführt wurden, ist es möglich, dass hier nur wenige Erfahrungen vorliegen, sondern vielmehr auf andere strukturähnliche Erfahrungswerte ausgewichen wird (z. B. auf die Einführung von Bildungsplänen oder anderen Innovationsmaßnahmen) oder, in Bedenken vor Freiheitseinschränkungen oder Kontrollverlust (vgl. Steins

S. 185ff.), Widerstand in unterschiedlicher Ausprägung erfolgt. Letzteres ist wahrscheinlich, weil dem Thema Evaluation insbesondere bei wenig ausgeprägter Kenntnis (z. B. der Ziele, der Möglichkeiten, der Chancen) eine Vermessungs- und Kontrollstruktur inne liegt (z. B. aufgrund der Begriffe, der notwendigen Engführung aufgrund von Operationalisierungsprozessen, der Bewertung über bestimmte positiv bzw. negativ formulierte Skalierungen). Insofern ist nicht verwunderlich, wenn auf der Ebene der Einstellungen das Thema Evaluation eher skeptisch betrachtet wird. Liegen bereits positive oder negative Erfahrungen mit Evaluationen und Evaluationsergebnissen vor, dann ist zu vermuten, dass diese die Einstellungen stark prägen. Dies fügt sich in das Bild ein, das Bauer (S. 219ff.) auslegt, wenn Lehrkräfte, die gezielt und damit bewusst Innovationen einführen, auch offener mit der Innovation ‚Evaluation‘ umgehen. Die Arbeit an Einstellungen ist gegenüber der Entwicklung von Konzepten oder der Anleitung für Handlungsweisen schwieriger. Das erfolgsversprechendere Ermöglichen von positiven Erfahrungen und positiven Zugängen zu Evaluationen ist aufwändig. Überreden, verbales Überzeugen, Darbietung von Vorteilen etc. kann wertvolle Impulse liefern, reicht jedoch nicht aus. Diese skeptische Betrachtung richtet sich insbesondere an diejenige Seite des „Selbst“ (vgl. Bauer S. 4), die sich nicht an berufstypischen Werten wie die optimale Entwicklung von Lernenden orientiert oder orientierten kann. Wenn Evaluation als selbstverständliche Standardkompetenz und als selbstverständlicher Teil eines professionellen Selbst angesehen wird, ist ein produktiver Umgang höchst wahrscheinlich.

Auf der nächsten Ebene von (hier zusammenfassend beschriebenen) *Wissen und Fähigkeiten* ist zu klären, inwiefern Lehrkräfte, Teams (z. B. Fachkonferenzen) oder Einzelschulen in der Lage sind, eine fachlich angemessene Konzeption zu entwickeln oder mit einer vorliegenden Konzeption (z. B. Rückmeldeformat) fachlich angemessen zu arbeiten. Hier fließen Wissen über Evaluationen (z. B. Phasen der Evaluation, Eigenschaften von Testaufgaben, statistischen Prozeduren) und Fähigkeiten (z. B. Verbinden didaktischer Fähigkeiten mit einem Wissen über Aufgabenformate von Vergleichsarbeiten) zusammen. Diese Ebene verhindert, dass einzelne und isolierte Maßnahmen aus Evaluationen abgeleitet werden, sondern systematische Klärungen und Strategien entwickelt werden.

Die dritte Ebene der *Verhaltensweisen und Handlungen* sichert, dass Konsequenzen aus Evaluationsergebnissen bei Schülerinnen und Schülern in Form konkret veränderter und weiterentwickelter Lernarrangements ankommen. Dies erfolgt etwa durch veränderte Aufgabenformate im Mathematikunter-

richt, durch konkrete kumulativ aufgebaute Angebote zur Entwicklung selbständigen Lernens.

Optimalerweise sind Übergang und Passung zwischen den drei Ebenen geschmeidig und stimmig. Der Umgang mit Evaluationsergebnissen ist dann geprägt von einer konstruktiven und proaktiven Einstellung („Evaluationen und Evaluationsergebnisse bringen mich in meiner Kompetenzentwicklung weiter“), einer konzeptionellen Klärung beim Umgang mit Evaluationsergebnissen (z. B.: schulische Konzeption für das Erarbeiten von Konsequenzen aus Vergleichsarbeiten im Fach Mathematik) und entsprechenden Verhaltensweisen und Handlungen (z. B. konkrete Änderungen bei der Unterrichtsplanung und -gestaltung im Fach Mathematik, Klasse 7, Thema Bruchrechnen). Bei der Analyse der Maßnahmen, die in verschiedenen Bundesländern ergriffen werden, um erfolgreiche Evaluationen zu ermöglichen, entsteht der Eindruck, dass die Wissens Ebene besonders ausführlich bedient wird. Sie ist naturgemäß leicht zugänglich. Dies wird über die Hinweise, Verfahren, Instrumente (z. B. Bögen zur Selbstevaluation) deutlich, die an den Landesinstituten für Schulentwicklung in den Bundesländern erarbeitet und verfügbar gemacht werden. Diese Ebene ist systematisch zentralisiert bedienbar, während die Ebene der Handlungen deutlich individualisiert und alltagsnah zu gestalten ist und die Ebene der Einstellungen sich einem unmittelbaren technizistischem Zugriff entzieht.

### **Wie müssen Rückmeldeformate aus Evaluationsergebnissen aussehen, damit Lehrkräfte mit ihnen nachhaltig arbeiten?**

Im Rahmen nationaler und internationaler Schulleistungsvergleichsstudien, die innerhalb von Deutschland durchgeführt wurden oder an denen Deutschland beteiligt war, wurde bisher fast durchweg Daten an die beteiligten Schulen zurückgemeldet, allerdings in anfangs wenig ausdifferenziertem Format. Auf der Grundlage vorliegender Studien (vgl. Beitrag von Kuper/Schneewind, S. 113ff.) können Ergebnisrückmeldungen wie folgt formatiert werden. Die Ergebnisse

- sollten zeitnah zur Datenerhebung an die einzelnen beteiligten Lehrkräfte rückgemeldet werden,
- sie sollten leicht erhaltbar sein,
- sie sollten technisch einfach zu handhaben sein,
- sie sollten verständlich sein und keine statistischen Grundkenntnisse erfordern,

- die Möglichkeiten und Grenzen der Evaluation deutlich machen,
- graphisch attraktiv und übersichtlich sein,
- nicht zu viele Informationen enthalten,
- Ansprechpartner für weitere Fragen enthalten,
- in einem unterstützenden Duktus verfasst sein,
- die verwendeten Begriffe sollten für Lehrkräfte verständlich sein.

Diese Hinweise sind meines Erachtens noch vollständig unter ‚handwerklichen‘ Verbesserungen von Evaluationsverfahren zu sehen. Komplizierter wird der nächste Schritt: Wie können optimale Rückmeldeformate in veränderte Unterrichtsarrangements überführt werden?

### **Lehrerinnen und Lehrer und Kollegien innerhalb der Organisation Schule: Innovationspotentiale im Spiegel einer Schulbiographie**

Im Beitrag von Rosenstiel (vgl. S. 43ff.) wird das Verhältnis von Subjekt und Organisation angesprochen. Dies zeigt sich etwa in typisch personalpsychologischen Themen wie Eignungsdiagnostik, Personalauswahl oder Personalentwicklung, die an Schulen seit einigen Jahren eine zunehmende Rolle spielen. Ein anderer Themenbereich verweist jedoch auf ein Forschungsdefizit in der Schulforschung und ein Desiderat der Personalentwicklung: Im Kapitel ‚Sozialisation in und durch Organisationen‘ verdeutlicht von Rosenstiel, wie das Subjekt von der Organisation im Laufe der Jahre beeinflusst wird. Die Arbeit in Organisationen bedingen psychosomatische Erkrankungen, wirken sich auf Werte aus oder beeinflussen die Soziabilität (Rosenstiel S. 8 xxx). Insofern lässt sich die gesundheitsförderliche Gestaltung der Organisation als Personalentwicklung verstehen. Es ist daher nicht verwunderlich, wenn eine Organisation als soziales System, mit all ihren vielfältigen Beziehungsstrukturen und Individuen, die konsequentes Change Management betreibt, gleichzeitig auch eine psychologische Intervention betreibt (Rosenstiel S. 43ff.).

Über diese allgemeine Anbindung der Organisationspsychologie an die Schule und die Arbeit von Lehrkräften hinaus, zeitigt das Verhältnis von Innovation, Subjekt und Organisation für den Umgang mit Evaluationsergebnissen konkrete Folgen. Wenn Schulen Leitbilder, Programme oder Curricula entwickeln, spiegelt sich darin der jeweilige Stand der Konsensbildung in einem Kollegium wieder. Entwürfe zu einem Schulprogramm werden innerhalb von Kollegien zunächst zumeist unterschiedlich gesehen: der einen Gruppierung

ist der Entwurf zu radikal, einer anderen zu rückständig, der dritten genau passend. Im Prozess der Konsensbildung erfolgt eine Annäherung bis – im optimalen Falle – alle das Programm unterstützen. Insofern sind veröffentlichte Dokumente (z. B. Schulprogramm) Ausdruck des derzeitigen Innovationspotentials eines Kollegiums. Die Frage, auf welchem „Innovationsniveau“ ein Konsens möglich und wie verbindlich eine Innovation für das gesamte Kollegium ist, hängt auch vom der Biographie der einzelnen Lehrkräfte innerhalb ihrer Schule ab. Dies sei an einem Beispiel verdeutlicht.

Betrachten wir eine Realschule. Sie hat ca. 60 Lehrkräfte und ca. 1000 Schülerinnen und Schüler. Sie liegt in einer kleineren Stadt und ist insofern eine relativ normale Realschule. An dieser Realschule entwickelte bereits in den 70er Jahre eine engagierte Gruppe von Lehrkräften eine Konzeption für offenen Unterricht. Diese Konzeption wird im Kern, wenn auch immer wieder weiterentwickelt, relativ unverändert bis heute realisiert. An mehreren Stunden in der Woche arbeiten die Klassen dieser ca. 10 Lehrkräfte, die sich immer in Teams organisieren, in einer bestimmten Wochenplan- und Freiarbeitskonstellation. Obschon die Gruppe sehr offen ist, sind im Laufe der Zeit nur wenige weitere Lehrkräfte dazu gekommen, die in ähnlich konsequenter Weise offenen Unterricht praktizieren. Dies ist insofern erwähnenswert, weil die Gruppe selbst bundesweit immens viele Fortbildungen für andere Lehrkräfte zum offenen Unterricht durchgeführt hat. Aber offensichtlich ist es nur in sehr begrenztem Maße gelungen, die Innovation ‚offener Unterricht‘ im eigenen Kollegium zu verbreiten. Weshalb ist dies Situation erwähnenswert? Die Situation steht exemplarisch für ein schulinternes Innovationspotential, dessen weitere schulinterne Verbreitung über Jahrzehnte hinweg nicht gelungen ist.

Nun führt diese Schule mit den Instrumenten des baden-württembergischen Landesinstituts für Schulentwicklung eine Selbstevaluation durch. U. a. füllen die Schülerinnen und Schüler einen Fragebogen aus, auf dem sie eintragen (Viererskalierung) wie viele ihrer Lehrkräfte diese Unterrichtsformen einsetzen. Tab. 1

Bitte kreuze in jeder Zeile das Kästchen an, das am meisten zutrifft.		(fast) alle Lehrer/innen	mehr als die Hälfte der Lehrer/innen	weniger als die Hälfte der Lehrer/innen	(fast) keine Lehrerin/ kein Lehrer	weiß nicht
		4	3	2	1	
9 a	Die Lehrerin/der Lehrer spricht lange vor der Klasse und wir hören nur zu.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g	Wir bearbeiten in einer Gruppe eine Aufgabe (Gruppenarbeit).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h	Jeder von uns kann an einer selbst gewählten Aufgabe arbeiten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j	Wir arbeiten nach einem Wochenplan.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Tab. 1:** Basisinstrument Selbstevaluation / Fragebogen für Schülerinnen und Schüler (Auszug)

Die Ergebnisse zeichnen folgendes Bild: Ca. 20% der Schülerinnen und Schüler kreuzen „mehr als die Hälfte der Lehrerinnen und Lehrer“ bei den Items g, h und j an. Ca. 70% der Schülerinnen und Schüler kreuzen hier an „fast kein Lehrer“. Die Ergebnisse spiegeln das Bild einer deutlichen methodisch-didaktischen Spaltung des Kollegiums – zumindest mit Blick auf offenen Unterricht und Differenzierung. Die Daten liegen nun vor. Die seit vielen Jahren stabile Situation kommt über die Selbstevaluation an die Oberfläche und auf den Tisch des Kollegiums. Die Ergebnisse sind relativ eindeutig, dem Kollegium ist der Hintergrund klar. Nun stellt sich eine Frage, die das Kollegiums seit den 70er Jahren begleitet: Wie kann diese methodisch-didaktische ‚Vielfalt‘ angeglichen werden? Soll sie angeglichen werden? Wenn ja – in welche Richtung und in welcher Weise? Insofern sind die Evaluationsergebnisse dem Kollegium bekannt. Zwischen der Gruppe ‚offener Unterricht‘ und dem restlichen Kollegium besteht in methodisch-didaktischer Hinsicht eine deutliche Distanz, und zwar seit vielen Jahren. Abb. 2 verdeutlicht die Situation dieser Schule.

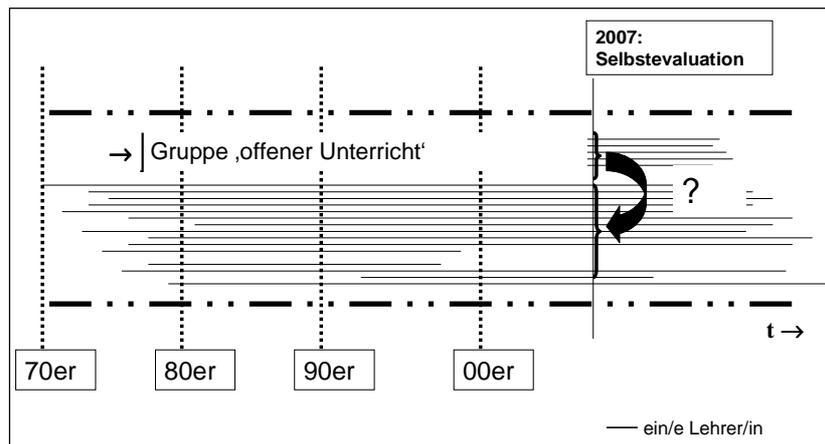


Abb. 2: Berufsbiographien einzelner Lehrkräfte innerhalb einer Einzelschule

Nun könnte man argumentieren, es gehöre zum professionellen Selbstverständnis von Lehrerinnen und Lehrern, ihr Methodenrepertoire anzupassen und kontinuierlich zu verändern. Forschungsbefunde zeigen jedoch, dass sich das Methodenrepertoire von Lehrkräften, insbesondere an Sekundarstufen, nur sehr langsam verändert (Bohl 2000; Helmke/Jäger 2002). Der Blick in die Biographie dieser Einzelschule offenbart schwierige Entwicklungen: Das Kollegium wird gemeinsam älter, bei der gegenwärtigen Altersstruktur von Kollegien ist das besonders wirksam. Man kennt sich. Jeder weiß, wen er ansprechen könnte, wenn es um Innovationen geht. Es ist bekannt, wer für was zu haben ist. Alle Lehrkräfte, die sich nicht der Gruppe offener Unterricht angeschlossen haben, hatten vielfach Gelegenheit dazu. Sie haben es jedoch nicht getan. Die Gründe hierfür sind vielfältig, sie liegen im mikropolitischen Bereich (Altrichter/Posch 1996), sind mit zeitlicher Belastung oder psychischer Belastung (vgl. Beitrag von Kieschke) zu erklären, mit Unsicherheit, mit schlichter Ablehnung.

Lernen aus Evaluationsergebnissen findet insofern immer als Verbindung zwischen Subjekt und Organisation statt. Diese Verbindung ist geprägt von der Erfahrung, die einzelne Lehrkräfte im Umgang mit Innovationen in der

Schule bereits hatten. Die Implementierung von Innovationen sollte daher sensibel gegenüber diesen bisherigen Erfahrungen sein.

### **Unterricht konzeptionell verändern: Innovationspotentiale im Spiegel individueller Möglichkeiten**

Wie im Beitrag von Märkl (S. 283ff.) zur Schulinspektion in Niedersachsen deutlich wird, taucht mit einer systematischen Evaluation ein bekanntes Problem der Unterrichts- und Schulentwicklung auf: Im Bereich des Lehrerhandelns im Unterricht ist einerseits der Entwicklungsbedarf besonders deutlich, andererseits ist die Veränderung des Lehrerhandelns im alltäglichen Unterricht, und nicht in abgrenzbaren Phasen wie Projektwochen, die größte Herausforderung. Die Befunde aus Niedersachsen verdeutlichen positive Einschätzungen für Zielorientierung/Strukturierung und Pädagogisches Klima, weniger positiv sind die Einschätzungen jedoch hinsichtlich Differenzierung und Unterstützung bei aktiven Lernprozessen. Dies verweist auf Defizite bei der Förderung von Selbständigkeit und Teamfähigkeit (vgl. Beitrag von Märkl S. 283ff.).

Insofern trifft das Thema dieses Bandes, Lernen aus Evaluationsergebnissen' auf ein grundlegendes Thema der Unterrichtsentwicklung: Unterrichtsentwicklung ist gut möglich, wenn...

- nur einzelne Bausteine in den eigenen Unterricht zu integrieren sind (z. B. Bausteine zum Methodentraining), oder
- relativ geschmeidig an bereits vorhandenen Erfahrungen angesetzt werden kann (z. B. kleinere Referate zu Buchvorstellungen werden mittel- und langfristig in Präsentationen und diese etwa in ein Konzept ‚Lernen durch Lehren‘ überführt), oder
- bereits bekannte Konzeptionen lediglich variiert oder in einigen Aspekten verändert werden (z. B. Gruppenarbeit wird bereits regelmäßig durchgeführt und jetzt ‚optimiert‘, z. B. mit Blick auf Auswertungsphasen).

Schwieriger wird die Unterrichtsentwicklung

- wenn neue Unterrichtskonzepte (z. B. selbstorganisiertes Lernen, Varianten offenen Unterrichts, differenzierte Lernumgebung) eingeführt werden sollen, die in ihrem Kern bisher nicht realisiert wurden, oder
- grundlegende Aspekte des Classroom-Managements entwickelt werden müssen (z. B. Allgegenwärtigkeit).

Insofern wird deutlich, dass innerhalb eines Lehrerkollegiums differenziert werden muss (vorausgesetzt die Bereitschaft zur Unterrichtsentwicklung ist vorhanden): Wenn eine Lehrerin bisher nicht in der Lage war, ihr Classroom-Management zu verbessern, wird sie dazu auch aufgrund von Evaluationsergebnissen nicht in der Lage sein. Wenn ein Lehrer sich bisher nicht dazu in der Lage sah, selbstorganisiertes und selbstständiges Lernen konzeptionell zu entwickeln, wird er es vermutlich auch nicht tun, wenn seine Schülerinnen und Schüler im Rahmen einer Evaluationsrückmeldung hier negativ abgeschnitten haben. Die Frage der Veränderungsbereitschaft und -fähigkeit auf individueller Ebene zeigt sich ebenso auf der Ebene der Einzelschulen. Um Unterricht systematisch zu verändern und mit den Maßnahmen nicht die bereits offenen und engagierten Lehrkräfte zu erreichen, sind vermutlich alle Register einer begleitenden Unterstützung, Beratung, Kooperation und Fortbildung zu ziehen. Erste Schritte in diese Richtung werden im Beitrag von Ballasch (S. 299 ff.) beschrieben: Gemeinsame Vorbereitung von Unterrichtsprojekten, gemeinsames Erstellen von didaktische Materialien, gemeinsam gezielt Fortbildungen organisieren, Schülergruppen oder einzelne Schüler gezielt fördern u. a.

### **Aufgabentypen zwischen vergleichender Bildungsforschung und individuell-didaktischem Handeln**

Aufgaben und die Orientierung an Aufgaben haben in den vergangenen Jahren die Methodik der Bildungsforschung erheblich beeinflusst. Über das erfolgreiche oder erfolglose bewältigen von Testaufgaben werden Schülerinnen und Schülern bestimmte Kompetenzen bzw. Kompetenzstufen zugeschrieben. (Test-)Aufgaben werden jeweils bestimmte Eigenschaften zugeschrieben (z. B. fachliches Anspruchsniveau, Reflexion, Transfer, Lebensnähe).

	<b>primäre Funktion</b>	<b>Beispiel</b>	<b>vorrangige Protagonisten der Aufgabenentwicklung</b>	<b>Kriterien der Aufgabenauswahl</b>
<b>Testaufgaben</b>	Steuerungswissen generieren	Aufgabe ‚Graffiti‘ (Lesekompetenz PISA 2000)	Wissenschaftler (Textexpertise und Fachexpertise)	testtheoretisch begründet, z. B.: Zugehörigkeit zu einem Testmodell, Itemanalyse

Fazit: Unter welchen Bedingungen ... 11

<b>Prüfungsaufgaben</b>	Selektionsentscheidungen legitimieren	Abituraufgabe Fach Deutsch	erfahrende Praktiker	curriculare Validität, fachliche und bildungstheoretische Zielsetzung des Unterrichts
<b>Förderdiagnostische Aufgaben</b>	Informationen über den individuellen Stand erhalten, aus denen Maßnahmen für die weitere Förderung und Qualifizierung eines Lernenden abgeleitet werden können	vom Lehrer entwickelte oder von Praktikern/Experten bereitgestellte i.d.R. fachspezifische Aufgaben	Lehrer/Praktiker	
<b>Förderdiagnostisch eingesetzte Tests</b>	Testbasierte Informationen über den individuellen Stand erhalten, aus denen Maßnahmen für die weitere Förderung eines Lernenden abgeleitet werden können	wissenschaftliche Tests, TEDDY-PC (Test zur Diagnose von Dyskalkulie)	Wissenschaftler (Textexpertise und Fachexpertise)	testtheoretisch begründet, z. B.: Standardisierung, Validität, Objektivität
<b>Intentionale Lernaufgaben</b>	Lernen, bewusst unterstützt von Planung und Rahmenbedingungen, i.d.R. alltäglicher Unterricht zur Generierung von neuen Erkenntnissen von Schülern/Klassen.	Lehrer entwickelt eine komplexe fachspezifische Aufgabe (Problemaufgabe oder Übungsaufgabe); Lehrer stellt anregende Lernumgebung zur Verfügung (z. B. elaboriertes Montessori-Material)	Lehrer	curriculare Validität, Mittleres Anspruchsniveau,
<b>Inzidente Lernaufgaben</b>	Funktion im Rahmen der Entwicklung, vorrangig	eine Jugendliche beobachtet, wie ein Nachbar einen	„das Leben“	keine

## 12 Thorsten Bohl

	außerschulisch	Ölwechsel durchführt, hilft ihm dabei und macht es schließlich beim Auto der Eltern alleine		
--	----------------	---	--	--

**Tab. 2:** Aufgabentypen und ihre Merkmale

Die Übersicht über verschiedene Aufgabentypen (Tab. 2) hat vorrangig analytischen Charakter. Je stärker die Mikroebene von Planungs- und Durchführungsprozessen des Unterrichts in den Blick geraten, desto deutlicher werden die Verbindungen und Überschneidungen zwischen den verschiedenen Aufgabentypen:

- Die Aufgaben selbst müssen sich nicht zwangsläufig unterscheiden, sie erhalten ihre Funktion über die Zielsetzung und den Kontext. So kann dieselbe (zumindest strukturgleiche) Mathematikaufgabe als Lernaufgabe fungieren (z. B. bei der Vorbereitung auf eine Abschlussprüfung) und in diesem Kontext keinerlei Selektionsfunktion innehaben, dieselbe (bzw. strukturgleiche) Aufgabe kann dann innerhalb der Abschlussprüfung eine Selektionsfunktion erfüllen.
- Dieselbe Aufgabe kann als Lernaufgabe, als Prüfungsaufgabe oder als diagnostisch eingesetzte Aufgabe fungieren. Die Änderung der Funktion hängt damit zusammen, welche Priorität die Lehrkraft hier legt. Die Aufgabenbearbeitung des Schülers Peter kann von der Lehrkraft kontrolliert und bewertet werden – nicht mehr (Prüfungsaufgabe) – oder: die Lehrkraft kann anschließend aufgrund der gelösten und nichtgelösten Aufgaben differenziert herausarbeiten, welche weiteren Aufgaben dem Schüler Peter angeboten werden sollen, damit er zielgerichtet seine Defizite bearbeiten kann und daraufhin einen individuellen Förderplan entwickeln. Die Aufgabe erhält dann einen förderdiagnostischen Charakter.
- Nach Veröffentlichung der TIMSS-Studie 1997 setzte in der Mathematikdidaktik<sup>1</sup> eine Diskussion um eine veränderte Aufgabenkultur ein, bis hin zum Projekt SINUS. Aufgaben, die im Rahmen der Bildungsforschung als

---

<sup>1</sup> z. B.: Zs. Mathematik Lehren. Heft 90/1998. Themenheft: TIMSS – Anstöße für den Mathematikunterricht; Zs. Mathematik Lehren. Heft 100/ 2000. Themenheft: Offene Aufgaben

Testaufgaben eingesetzt wurden, wurden über Multiplikatorenstrukturen in den Unterrichtsalltag ‚transferiert‘, also zu Lernaufgaben umgewandelt. Grob betrachtet lassen sich *Testaufgaben* mit wissenschaftlichem Anspruch und *didaktische Aufgaben*, die vom Lehrer alltagsnah entwickelt werden, unterscheiden. Der Übergang von den großen nationalen und internationalen Studien zum didaktischen Handeln von Lehrkräften lässt sich zunächst in einer zu fordernden komplementären Perspektive beschreiben: „Auf die ‚Leistungen im System‘ zu blicken, erfordert nämlich auch, Konzepte und Instrumente als die Basis der wissenschaftlichen Beobachtung so zu entwickeln, dass die unter anderer Perspektive erzeugten didaktischen Kategorien erhalten bleiben.“ (Neubrand, S. 97ff.). Im Idealfall ergänzen sich also beide Perspektiven. Dass zwischen beiden Perspektiven bzw. Aufgabentypen vermittelt werden kann, erscheint sehr wahrscheinlich. Dies wird besser gelingen, wenn die Vermittlung zwischen beiden Perspektiven sorgfältig konzeptionell ausgestaltet und beratend begleitet wird (vgl. Abb. 2, Schritt 4): Welche Begriffe können zwischen beiden Perspektiven vermitteln? Welche Aufgaben bzw. Aufgabenteile wurden bereits in ähnlicher Weise im Unterricht realisiert? Was kann leicht übernommen werden, wo werden Probleme gesehen? Welcher Art sind die Probleme (Einstellungen? Konzeption? Handlungsebene? Zeit? Curriculum?)? Bei intensiver Vermittlung wird sich die Schnittmenge erhöhen. Allerdings wird es, trotz Übereinstimmungen, auch klare Abgrenzungen geben.

In den Testaufgaben großer Studien liegt zweifellos ein großes Anregungspotential (Neubrand, S. 97ff.). Inwiefern eine enge Verbindung bzw. große Schnittmenge gelingt, liegt nicht nur an den Aufgaben selbst, sondern vielmehr in deren Einbettung

- in eine didaktische und fachdidaktische Konzeption des Unterrichts (z. B.: bildungstheoretisches Verständnis, Leistungsverständnis, Kompetenzverständnis, zentrale Begriffe),
- in externe und interne Unterstützungsmaßnahmen: didaktische Aufbereitung der Ergebnisse, Unterstützung bei der ‚Übersetzung‘ von Testergebnissen, Fortbildungen, Fachkonferenzen, schulinterne Klärungen.

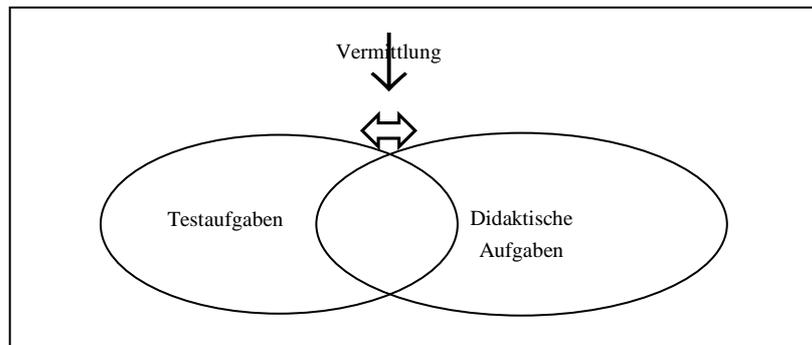


Abb. 3: Schnittmenge zwischen Testaufgaben und didaktischen Aufgaben

Wie verwirrend die verschiedenen Aufgabentypen und ihre Funktionen für Lehrkräfte sowie Eltern sind, wird in der Studie von Maier (vgl. Beitrag Maier S. 131ff.) deutlich. Insofern sind nicht nur Vermittlungsmöglichkeiten zwischen den Aufgaben, sondern zunächst ihre unterschiedlichen Funktionen zu klären.

### **Lernen aus Evaluationsergebnissen: Günstige Bedingungen**

Auf der Grundlage theoriebezogener und empirischer Forschung können günstige Bedingungen in zweierlei Hinsicht ermittelt werden. Erstens in allgemeiner, schulunspezifischer Betrachtung zum Themenbereich ‚Vom Wissen zum Handeln‘, zweitens mit spezifischem Blick auf Handeln im Anschluss an Evaluationen.

- (1) Um zu erklären und zu verstehen, wie Lehrkräfte mit Evaluationsergebnissen umgehen, können unterschiedliche Theorien aus benachbarten Disziplinen herangezogen werden.
  - Wenn Individuen sich in ihrer Freiheit und ihren Verhaltensalternativen eingeschränkt sehen, entsteht das Bedürfnis, diese Freiheitsgrade zu erhalten oder wieder herzustellen. Es kommt zu Widerstand, der in unterschiedlicher Ausprägung erfolgen kann (Reaktanztheorie, vgl. Beitrag von Steins S. 185ff.).
  - Individuen und Personengruppen (Schulaufsicht, Lehrkräfte, Eltern) neigen dazu, unerfreuliche Resultate ursächlich den anderen Personen-

gruppen zuzuweisen. Wenig erfreuliche Ergebnisse werden ursächlich am wenigsten bei der eigenen Gruppe lokalisiert (Attributionstheorie, vgl. Beitrag Kohler S. 81ff.).

- Kuper (2006, S. 7) beobachten gesamtgesellschaftlich eine Tendenz, dass unbegrenzt verfügbares wissenschaftliches Wissen auf begrenzte Verarbeitungskapazitäten stößt und dadurch die Hoffnung erschöpft würde, über Wissen Handlungsunsicherheiten reduzieren zu können.
- Die Kluft zwischen Wissen und Handeln als generelles Problem von Schule und Hochschule zeigt sich besonders bei Wissen, das nicht mit der eigenen Handlungslogik verbunden werden kann (Gruber u. a. 2000).
- Handlungstheoretisch gesehen ist die die Umsetzung von Wissen in Handeln unwahrscheinlich (vgl. Beitrag von Wahl S. 155ff.). Es müssen beispielsweise zunächst die handlungssteuernden inneren Bilder modifiziert werden.
- In systemtheoretischer Betrachtung ist bekannt „dass zwar von außen interveniert werden kann, aber das System selbst ‚entscheidet‘, wie es diese Interventionen verarbeitet. Das gilt auch für das System Einzelschule.“ (Rolf 2001, S. 338)

Der Umgang mit Evaluationsergebnissen stellt ein prototypisches Beispiel für die Problematik des Umgangs mit Wissen dar. Von den Ergebnissen, die in der Regel schriftlich vorliegen, bis zu alltagsresistenten, darauf beruhenden innovativen Handlungen liegt ein weiter Weg. Exemplarisch seien an dieser Stelle Hinweise aufgegriffen, die Steins in ihrem Beitrag aus der Reaktanztheorie formuliert (vgl. Beitrag von Steins, S. 185ff.):

- Wenn Freiheitseinengungen notwendig sind, sollten sie gut begründete werden, damit sie eher akzeptiert werden.
- Die Verantwortlichen und Akteure der Evaluation müssen glaubwürdig sein.
- Gerade unsichere Personen beharren besonders auf Ihre Freiheit, sie akzeptieren Einschränkungen eher, wenn sie diese als fair und unterstützend wahrnehmen.
- Erzwungene Teilnahme und schnelle Umsetzung erhöhen die Wahrscheinlichkeit der Entstehung einer psychologischen Reaktanz.

(2) Der nationale und internationale Forschungsstand zur schulischen Nutzung von Leistungsdaten wurde von verschiedenen Autorinnen und Autoren zusammengetragen (z. B. v. Ackeren 2003; Bonsen/Gathen 2004;

Gathen 2006) und auch im vorliegenden Band thematisiert (Steins S. 185ff. Böttcher S. 207ff.; Maier S. 131ff.; Lind S. 61ff.). Dabei hinkt der nationale Forschungsstand dem internationalen zum Teil erheblich hinterher. Während in manchen Ländern, z. B. USA und England seit vielen Jahren Erfahrungen mit Leistungsmessung und Evaluationen bestehen, ist in Deutschland sowohl die Entwicklung von Rückmeldeformaten als auch die Rezeptionsforschung noch wenig entwickelt. Die bisherige Rezeption von Evaluationsergebnissen in Deutschland zeigt (in diesem Band: Kohler S. 81ff., Maier S. 131ff.), dass Lehrkräfte Daten nur sehr wenig für die Entwicklung ihres Unterrichts nutzen. Im Folgenden sei jedoch nicht dieser bekannte Sachverhalt thematisiert, sondern vielmehr einige Befunde herausgearbeitet, die Hinweise zur Verbesserung der Bedingungen im Umgang mit Evaluationsergebnissen bieten. Die herangezogenen Studien sollen das Spektrum der verfügbaren Befunde verdeutlichen.

In der Untersuchung von Stamm (2003) werden in qualitativen Analysen vier Typen der Evaluationsnutzung gebildet: „Innovation“, „Reaktion“, „Alibi“, „Blockade“. In dieser differenzierten Analyse zeigt sich, dass lediglich der Typ „Innovation“ engagiert Evaluationsergebnisse nutzt. Die anderen drei Typen werden eher als problematisch beschrieben. Mit Blick auf Konsequenzen aus Evaluationsergebnissen macht diese Studie, dank ihrer qualitativen, vertiefenden Analyse deutlich, dass eine deutlichere und differenziertere Adressatenorientierung notwendig ist. Dies vereinfacht die Situation nicht. Aber es wird deutlich, dass der Typ „Innovation“ mit geringeren Anleitungsmaßnahmen auskommt als etwa der Typ „Blockade“ und insofern eine einheitliche Unterstützungskonzeption wenig Ertrag verspricht.

Im Rahmen der PISA 2000 – Studie entwickelte Hessen eine eigene Rückmeldekonzepktion (Scharz/Steffens 2003). Schulen erhielten auf Wunsch externe Unterstützung (Schulpsychologe des eigenen oder eines anderen Schulamtes). Die Einschätzungen zu weiten Teilen dieser Beratung waren positiv bzw. entsprachen den Erwartungen. Allerdings wurde kritisch eingeschätzt, inwiefern aus der Beratung heraus Anregungen für die weitere Arbeit resultierten. Aus dieser Studie heraus wird deutlich, dass Unterstützung bei der Rückmeldung positiv gesehen wird, aber nur ein erster Schritt darstellt und damit noch nicht entsprechende Weiterentwicklungsmaßnahmen einhergehen.

Ditton/Arnoldt (2004) fanden in ihrer Studie zu Schülerbefragungen im Fachunterricht heraus, dass Lehrkräfte, die kooperationsbereit sind, eine

höhere berufliche Zufriedenheit haben und sich häufiger fortbilden, eher von externen Evaluationen erreicht werden. Einzelschulen, an denen allgemeine Schulklima positiv eingeschätzt wurde, setzen sich mit schulbezogenen Daten intensiver auseinander (QUASUM-Studie, vgl. Peek 2003, S. 71). Evaluationsbefragungen im Rahmen der VERA-Studie (vgl. Groß Ophoff u. a. 2007) zeigen, dass externe Unterstützung besonders dann erfolgreich ist, wenn sie kontinuierlich angeboten wird und eng mit den Ergebnissen der Vergleichsarbeiten verbunden wird. Zwar sind durchaus Konsequenzen für den Unterrichtsentwicklungsprozess erkennbar, allerdings werden insbesondere Maßnahmen verstärkt, die bereits vor den Vergleichsarbeiten vorhanden waren, d.h. eine deutliche konzeptionelle Änderung des Unterrichts scheint nicht damit einherzugehen. Alle drei Studien verdeutlichen folgendes. Kontinuierliche Unterstützung scheint hilfreich zu sein, allerdings bleiben zwei Problemfelder: Erstens gelingt es damit noch nicht alle Lehrkräfte bzw. Einzelschulen zu erreichen, sondern lediglich diejenigen, die einem besonders ‚offenen‘ Typ entsprechen (höhere berufliche Zufrieden, Kooperationsbereitschaft, häufigere Fortbildungen; bzw. auf Einzelschulebene: allgemeines positives Schulklima); zweitens scheinen Lehrkräfte, wenig überraschend, stark an ihren bereits stabil vorhandenen Handlungsmustern und Maßnahmen gebunden zu sein. Sollen also möglichst viele Lehrkräfte aus einem Kollegium erreicht, sowie deutlich veränderte, neue Unterrichtskonzeptionen eingeführt werden, zu denen bisher noch keine Erfahrungen gewonnen werden konnten, dann sind differenziertere und massivere Unterstützungsmaßnahmen erforderlich.

Die qualitative Studie von Maier (vgl. S. 131ff.) erweitert die bisher benannten Befunde. Maier zeigt, dass die Hälfte der befragten Lehrkräfte ihren Unterricht nicht verändern werde und dies mit dem guten Abschneiden ihrer Schülerinnen und Schüler begründen. Dies ist insofern fatal, als exakt an dieser Nahtstelle die Unterscheidung zwischen fairen und unfairen Schulleistungsvergleichen greift: Bei unfairen Leistungsvergleichen werden die Ausgangsbedingungen (z. B. Intelligenzniveau, soziale Herkunft) nicht berücksichtigt werden. Gute Leistungen sind bei unfairen Vergleichen mit Blick auf das unterrichtliche Lernangebot völlig aussagemäßig. Unfaire Leistungsvergleiche verschließen damit eher Entwicklungsmöglichkeiten: „Wir hatten ein Gespräch mit dem Schulleiter und haben uns die Aufgaben angekuckt. Und da wir relativ gut abgeschnitten haben, war jetzt nicht so, dass wir was ändern müssten. Wir haben uns das angekuckt und haben gesagt, wir sind auf dem richtigen

Weg und wir arbeiten so weiter.“ (Maier, S. 131ff.) Die ungerechte und mit den Daten nicht zu rechtfertigende Einteilung von Lehrerkollegien und Einzelschulen in diejenige, die über oder unter dem Durchschnitt liegen bzw. gute oder schwache Ergebnisse vorweisen, trägt letztlich zu einer Deprofessionalisierungstendenz bei. Aus der Studie von Maier wird erkennbar, dass unfaire Leistungsvergleiche Ungerechtigkeiten zwischen Lehrkräften oder Einzelschulen verstärken und damit letztlich einer gezielten, adressatenangemessenen Verbesserung der Schülerleistungen entgegenstehen.

Internationale Befunde (vgl. Steins, S. 185ff.; Böttcher S. 207ff.; Lind S. 61ff.) belegen, dass Evaluationen in nicht unerheblichem Maße zu Betrug und zu Belastungen von Lehrkräften und Schülerinnen und Schülern führen können. Auf Kontrolle zielende Evaluationen können damit negative Effekte zeitigen. Insofern ist gerade angesichts der noch geringen Erfahrungen in Deutschland darauf zu achten, dass nicht Strukturen entwickelt werden, die die Evaluationsschraube überdrehen.

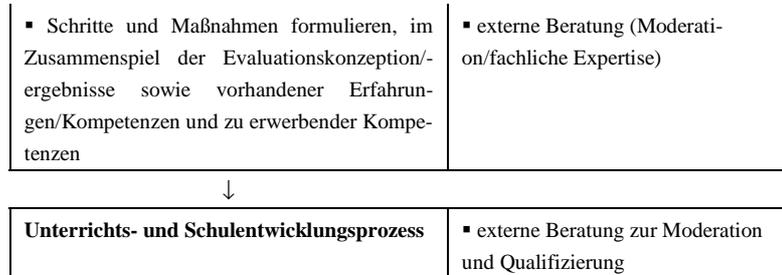
Die skizzierten Befunde verdeutlichen, dass bisher von sehr geringem, einseitigen, skeptischem bis ablehnenden Umgang mit Evaluationsergebnissen ausgegangen werden muss. Gleichzeitig wird deutlich wie geringschätzig bisher der gesamte Bereich der Rezeption von Evaluationsergebnissen sowohl aus wissenschaftlicher als auch aus bildungspolitischer Perspektive behandelt wurde. Insofern besteht ein immenses Entwicklungspotential in schlicht ‚handwerklicher‘ Hinsicht: Evaluationskonzeptionen, Bedingungen und Rückmeldeformate können noch deutlich verbessert werden.

Aus den bisherigen Überlegungen ergeben sich günstige Bedingungen, die die einzelnen Handlungsschritte im Evaluationsverlauf stabilisieren (Abb. 3).

Sämtliche der genannten Schritte bewegen sich (vgl. Abb. 1) auf einer konzeptionellen Ebene, z. T. je nach Grad der Konkretisierung auf einer Handlungsebene. Eine transparente, schlüssige, sensibel dargestellte Bearbeitung der einzelnen Schritte wird zu einer positiven Akzeptanz der Evaluation beitragen und kann insofern möglicherweise zur Veränderung von Einstellungen beitragen.

Fazit: Unter welchen Bedingungen ... 19

	<b>Handlungsschritte</b>	<b>günstige Bedingungen</b>
1	<b>Konzeption einer Evaluationsstudie</b>	<ul style="list-style-type: none"> <li>▪ fairer Vergleich,</li> <li>▪ Transparenz der Möglichkeiten und Grenzen, adressatenfreundliche Konzeption (z. B. zentrale Begriffe, Aufgabentypen, Leistungsverständnis)</li> <li>▪ Klärung des ‚Dateneigentums‘</li> </ul>
	↓	
2	<b>Information bzw. Kenntnis und Entscheidung über Teilnahme an einer Evaluation</b>	<ul style="list-style-type: none"> <li>▪ Freiwilligkeit der Teilnahme bzw. Mitbestimmung bei Konzeption und Verfahren (z. B. Auswahl von Teilbereichen).</li> </ul>
	↓	
3	<b>Ablauf der Evaluation</b> <ul style="list-style-type: none"> <li>▪ Vorbereitung</li> <li>▪ Durchführung</li> <li>▪ Auswertung</li> <li>▪ Beschreiben und Analysieren der Ergebnisse</li> </ul>	<ul style="list-style-type: none"> <li>▪ Nutzen und Verfahren für Adressaten transparent</li> <li>▪ standardisiertes Verfahren</li> <li>▪ standardisiertes Verfahren, externe Korrekturen</li> <li>▪ adressatenfreundliches Rückmeldeformate (Sprache, didaktische Hinweise, zeitnah, anschaulich), externe Bearbeitung</li> </ul>
	↓	
4	<b>Umgang mit Ergebnissen und Konsequenzen aus Ergebnissen (Rückmeldeformate)</b> <ul style="list-style-type: none"> <li>▪ Interpretation und Diskussion der Ergebnisse</li> <li>▪ Gewichtung herstellen: Welche Befunde sind bedeutsamer?</li> <li>▪ Welche Ergebnisse dienen als Grundlage weiterer Entwicklungsprozesse?</li> </ul>	<ul style="list-style-type: none"> <li>▪ externe Beratung (Evaluatoren)</li> <li>▪ externe Beratung (Moderation)</li> <li>▪ externe Beratung (Moderation)</li> </ul>



**Abb. 4:** Handlungsschritte, Akteure und günstige Bedingungen beim Umgang mit Evaluationsergebnissen

## Zusammenfassung und Schlussbemerkung

Folgende Bedingungen erscheinen aus den Beiträgen dieses Band heraus günstig für einen erfolgreichen Umgang mit Evaluationsergebnissen:

- (1) Bildungspolitische Ebene: Das Beispiel Amerika warnt vor einem Überdrehen der Evaluationsschraube (vgl. Beitrag von Lind). Daher sind die Funktion von Evaluationsverfahren i.w.S. (z. B. Vergleichsarbeiten, Diagnosearbeiten, Beteiligung an Schulleistungsvergleichsstudien) und die daraus resultierenden Konsequenzen sensibel und im Spannungsfeld von Kontrolle/Vergleich und Unterstützung deutlich in Richtung Unterstützung zu gewichten. In einigen Bundesländern wird Wert darauf gelegt, dass die ermittelten Daten Eigentum der Schule bleiben – eine unabdingbare Voraussetzung für vertrauensbasierte Evaluationsprozesse (vgl. Beitrag von Ballasch). Inwiefern bzw. auf welcher Ebene Rankings (Länder, Regionen, Einzelschule, Klassen, Lehrkräfte, Fächer...) gebildet werden, scheint noch völlig offen zu sein.
- (2) Ebene Schulverwaltung und Schulaufsicht: Die Schulbehörde bzw. ihr zugeordnete Behörden haben in allen Bundesländern in unterschiedlicher Konzeption die Organisation des Evaluations- und Qualitätsmanagements übernommen. Ihr Aufgabenfeld hat sich damit erheblich verändert. Eine besondere Herausforderung liegt in der qualitativen

Bewertung des Unterrichts. Dabei kommt der leicht zugänglichen Information, der offensiv angebotenen Transparenz des gesamten Qualitätsmanagements, der professionellen Kooperation mit den Schulen und der kontinuierlichen Verbesserung der Instrumentarien und des Verfahrens eine besondere Bedeutung zu (vgl. Beitrag von Märkl S. 283ff.). Dies ist Voraussetzung, um bei Schulleitungen und Lehrkräften Bereitschaft für Veränderungsprozesse zu erzeugen.

- (3) Ebene der Einzelschule: Die Qualität und Arbeit an der Einzelschule, bereits vor der Evaluation ist in hohem Maße entscheidend für erfolgreich implementierte Konsequenzen aus Evaluationsergebnissen: Eine qualitätsfördernde Arbeit der Schulleitung (z. B. zielbezogene Führung, vgl. Beitrag von Bensen), hohe Schulzufriedenheit und präventiver Umgang mit Belastungen (vgl. Beitrag von Kieschke), intensive Kooperation im Kollegium (z. B. Fachkonferenzen zur Diskussion über Aufgaben, Entwicklung gemeinsamer mentaler Modelle; Verhältnis Gruppe zur Organisation; vgl. Beiträge von Häcker, Rosenstiel, Miller, Bauer), Offenheit gegenüber Innovationen sind wesentliche Aspekte qualitätsvoller Schulentwicklungsarbeit.
- (4) Mit Blick auf Rückmeldeformate: Für Lehrkräfte verständliche, zeitnahe, didaktisch aufbereitete und komfortable Rückmeldung erhöht die Wahrscheinlichkeit, mit diesen weiter zu arbeiten. Innerhalb der einzelnen Bundesländer sind Evaluationsverfahren und Rückmeldeformate deutlich verbesserbar (vgl. Beiträge von Kuper/Schneewind; Maier; Ballasch). In diesem technischen und handwerklichen Bereich besteht noch hohes Entwicklungspotential in Deutschland.
- (5) Mit Blick auf vermittelbares Wissen: Kenntnis, Wissen und Handlungsmöglichkeiten von angehenden, praktizierenden und berufserfahrenen Lehrkräften sind gleichermaßen entwicklungsfähig. Hilfreich für das Verstehen von Evaluationsprozessen ist Wissen über den Umgang mit Daten, methodisches Wissen, Diagnosewissen, Wertwissen, Wissen über Problemlöseprozesse (vgl. Beitrag von Kiper S. 13ff.)
- (6) Mit Blick auf Lehrerfortbildung: Bisher ist die Wirksamkeit der Lehrerfortbildung begrenzt. Daher ist im Kontext von Qualitätssicherungsmaßnahmen darüber nachzudenken, in welcher Weise unterrichtsnahe, alltagsstabile Kompetenzen für Lehrerinnen und Lehrer

vermittelt werden können. (vgl. Beitrag von Landert S. 253ff.). Fortbildungen differenziert anzulegen. Je nach Situation und Konstellation in einem Lehrerkollegium kann es sinnvoll sein, dass einzelne Lehrkräfte eine externe Fortbildung besuchen oder gesamte Kollegium oder einzelne Gruppen im Kollegium eine Fortbildung erhalten. Insbesondere bei der konzeptionellen Veränderung von Unterricht sind individualisierte und begleitende Fortbildungsmaßnahmen vermutlich unumgänglich.

- (7) Mit Blick auf Aufgabenformate: Eine beiderseitige Annäherung von Testaufgaben und didaktischen Aufgaben, ohne deren jeweilige Charakteristika zu leugnen, ist hilfreich um Befunde der Bildungsforschung mit dem unterrichtlichen Alltag zu verbinden. Dabei kommt der Unterstützung bei der Vermittlung zwischen wissenschaftlich konzipierten Testaufgaben und (fach-)didaktischen Aufgaben eine entscheidende Rolle zu.
- (8) Mit Blick auf den beruflichen Alltag von Lehrkräften: In mehreren Beiträgen (z. B. Bonsen S.9xx; Kieschke S 1) wird deutlich, dass der berufliche Alltag von Lehrkräften innerhalb eines bestimmten Kollegiums in einer bestimmten Institution in zeitlich-organisatorischer Hinsicht reformbedürftig ist. So sind etwa die Zeiten für Schulentwicklungsaktivitäten nicht definiert, der Unterrichtsaufwand für deutsche Schulleitungen ist im internationalen Vergleich sehr hoch. Derzeitige, an 45-Minuten-Stunden orientierte Arbeitszeitmodelle sind angesichts der komplexer werdenden und ausdifferenzierten Aufgaben von Lehrkräften überholt.
- (9) Mit Blick auf Wissenschaft: Seit Veröffentlichung der ersten PISA-Studie 2001 hat die empirische Bildungsforschung einen immensen Aufwind erfahren, dokumentiert beispielsweise in einer Vielzahl von Professuren für empirische Bildungsforschung (Krüger u. a. 2008, S. 106ff). Die Zahl der inzwischen veröffentlichten nationalen und internationalen Vergleichsstudien mit jeweils angeschlossenen Vertiefungsbereichen ist kaum noch überschaubar und kaum noch intensiv und in Ruhe lesbar. Obschon nicht wenige Befunde in Fachkreisen kaum überraschen und immer wieder forschungsmethodische Kritik geäußert wird, ist der Erkenntnisgewinn der vergangenen Jahre groß. Bildung hat als politisches Thema einen enormen Stellenwert erhalten. Empiri-

sche Befunde dominieren die erziehungswissenschaftliche Diskussion und haben vordergründig zu einer Versachlichung der Diskussion beigetragen. Gleichwohl gehen problematische Entwicklungen einher (vgl. auch Tilmann 2008; Tilmann u. a. 2008):

- Bildungspolitik, Schulverwaltung und Landesinstitute für Schulentwicklung/Qualitätssicherung sind inspiriert von den Befunden der empirischen Bildungsforschung und forcieren selbst Evaluationsverfahren, Vergleichsarbeiten, eigene wissenschaftliche Studien etc. Die genaue Verbindung zu den Forschungsbefunden bleibt jedoch unklar und entzieht sich letztlich wissenschaftlicher Gründlichkeit. Widersprüchliche Ergebnisinterpretationen, zuweilen auch von Wissenschaftlichen selbst, kommen hinzu. So kommt es nach wie vor Systemsteuerungslogiken, die kaum nachvollziehbar sind, z. B. wenn in Baden-Württemberg Vergleichsarbeiten benotet und als Klassenarbeiten verrechnet werden können.
- Über die Stärkung der empirischen Bildungsforschung wird bisher weitgehend bildungswissenschaftliches Grundlagenwissen unterstützt. Das Ungleichgewicht zwischen der Menge empirisch ermittelter Erkenntnisse einerseits und der geringen Anschlussfähigkeit dieses Wissens an den innovationsresistenten Unterrichts- und Schulalltag ist nicht zu übersehen.

- (10) Mit Blick auf die Theoriebildung in Schulpädagogik und Allgemeiner Didaktik: Die neuen Überlegungen in der empirischen Bildungsforschung machen den Diskurs zwischen Schul- und Unterrichtsforschung und den hier verwendeten Testformaten und der Allgemeinen Didaktik, Unterrichtstheorie und Fachdidaktik auch mit Blick auf die Erweiterung des theoretischen und konzeptionellen Denkens über Schule und Unterricht notwendig.

Insgesamt besteht – dies wird in nahezu allen Beiträgen in unterschiedlicher Charakteristik und Ausprägung deutlich, ein eklatantes Ungleichgewicht zwischen Ressourcen, die für Verfahren der Datenerhebung (Evaluationen, Vergleichsstudien, Diagnosearbeiten etc.) eingesetzt werden und Ressourcen, die Konsequenzen aus Evaluationen unterstützen: Rückmeldeformate, Unterstützung, Beratung, Fortbildung etc.

## Literatur

- Ackeren, Isabell, v. (2003): Evaluation, Rückmeldung und Schulentwicklung. Erfahrungen mit zentralen Tests, Prüfungen und Inspektionen in England, Frankreich und den Niederlanden. Münster: Waxmann.
- Altrichter, H./ Wiesinger, S. (2005): Implementation von Schulinnovationen – aktuelle Hoffnungen und Forschungswissen. *Journal für Schulentwicklung*, 9. Jg./Heft 4, 28-36.
- Altrichter, H./Posch, P. (Hrsg.) (1996): Mikropolitik der Schulentwicklung. Förderliche und hemmende Bedingungen für Innovationen in der Schule. Innsbruck, Wien: Studien Verlag.
- Apel, H. J./Sacher, W. (2002): Schulpädagogik als Wissenschaft. In: Apel, H. J./Sacher, W. (Hrsg.) *Studienbuch Schulpädagogik*. Bad Heilbrunn: Klinkhardt, S. 7-34.
- Baker, E. L. (2004): Aligning Curriculum, Standards, and Assessments: Fulfilling the Promise of School Reform. CSE Report 645. Center for the Study of Evaluation. National Center of Research on Evaluation, Standards and Student Testing. University of California. URL: <http://cresst96.cse.ucla.edu/index4.htm> [Datum der Recherche: 07.07.2005]
- Bauer, K.-O. (1997): Professionelles Handeln in pädagogischen Feldern. Weinheim und München: Juventa.
- Bauer, K.-O. (1997): Unterrichtsentwicklung und professionelles Selbst der Lehrerinnen und Lehrer. In: Rolff, H.-G./Bos, W./Klemm, K./Pfeiffer, H./Schulz-Zander, R. (Hrsg.): *Jahrbuch der Schulentwicklung*. Band 11. Weinheim und München: Juventa. S. 201-231.
- Bohl, T. (2000): Unterrichtsmethoden in der Realschule. Bad Heilbrunn: Klinkhardt.
- Büeler, S./Buholzer, A. (2005): Innovation durch Schulentwicklung. Bilanz einer Längsschnittstudie. In: *Journal für Schulentwicklung* 9. Jg./Heft 4, S. 63-68.
- Ditton, H./Arnoldt, B. (2004): Schülerbefragungen zum Fachunterricht – Feedback an Lehrkräfte. In: *Empirische Pädagogik*. 18. Jg./Heft 1, S. 115-139.
- Fend, H. (1980): *Theorie der Schule*. München: Urban & Schwarzenberg.
- Fend, H. (2006): *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen*. Wiesbaden: Verlag für Sozialwissenschaften.
- Bonsen, M./Gathan, J. v. d. (2004): Schulentwicklung und Testdaten. In: Holtappels, H. G./Klemm, K./Pfeiffer, H./Rolff, H.-G./Schulz-Zander, R. (Hrsg.): *In: Jahrbuch der Schulentwicklung*. Band 13. Daten, Beispiele und Perspektiven. Weinheim und München: Juventa, S. 225-252.
- Gathan, J. v. d. (2006): Die innerschulische Rezeption von Leistungsrückmeldungen aus Large-Scale-Assessments – Grundlagen und Ziele von Fallstudien. In: Kuper, H./Schneewind, J. (Hrsg.): *Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungsbereich*. Münster: Waxmann, S. 77-88.
- Groß Ophoff, J./Hosenfeld, I./Koch, U. (2007): Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. In: *Empirische Pädagogik*. 21. Jg./Heft 4, S. 411-427.
- Gruber, H./Mandl, H./Renkl, A. (2000): Was lernen wir in Schule und Hochschule? Träges Wissen? In: Mandl, H./Gerstenmaier, J. (Hrsg.): *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze*. Göttingen: Hogrefe, S. 139-157.
- Helmke, A./Jäger, R. (Hrsg.) (2002): *Das Projekt MARKUS: Mathematikgesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau: *Empirische Pädagogik*.

- Jorgenson, O./ Vanosdall, R. (2002): The Death of Science? What We Risk in Our Rush Toward Standardized Testing and the Three R's. In: Kappan Professional Journal; Last updated 11 April 2002; URL: <http://www.pdkintl.org/kappan/k0204jor.htm>; Copyright 2002 Phi Delta Kappa International [Datum der Recherche: 03.02.2008]
- Klug, D./Reh, S. (2000): Was fangen die Schulen mit den Ergebnissen an? Die Hamburger Leistungsvergleichsstudie aus der Sicht ‚beforschter‘ Schulen. In: Pädagogik. 52. Jg./Heft 12, S. 16-21.
- Kohler, B. (2004): Zur Rezeption externer Evaluation durch Lehrkräfte, Eltern sowie Beamte der Schulaufsicht. In: Empirische Pädagogik. 18. Jg./Heft 1, S. 18-39.
- Krüger, H. H./Schnoor, O./Weishaupt, H. (2008): Personal. In: Tillmann, K.-J./Rauschenbach, Th./Tippelt, R./Weishaupt, H. (Hrsg.): Datenreport Erziehungswissenschaft 2008, Opladen: Barbara Budrich, S. 87-112.
- Kuper, H. (2006): Rückmeldung und Rezeption – zwei Seiten der Verwendung wissenschaftlichen Wissens im Bildungssystem. In: Kuper, H./Schneewind, J. (Hrsg.): Rückmeldung und Rezeption von Forschungsergebnissen. Münster: Waxmann, S. 7-16.
- Landesinstitut für Schulentwicklung (2008): URL: <http://www.schule-bw.de/entwicklung/qualieval/qualiab/sevstart/eisneu/basis/bifragen> [Datum der Recherche: 18.01.2008]
- Nachtigall, C. (Hrsg.) (2005): Landesbericht – Thüringer Kompetenztest 2005. Friedrich-Schiller-Universität Jena. URL: [http://www.kompetenztest.de/download/kt05/Landesbericht\\_KT2005.pdf](http://www.kompetenztest.de/download/kt05/Landesbericht_KT2005.pdf) [Datum der Recherche: 03.02.2008]
- Peek, R. (2003): Klassenbezogene Rückmeldungen aus Schulleistungsstudien und ihre Rezeption in beteiligten Schulen – QuASUM 2 in Brandenburg. Potsdam.
- Reh, S. (2004): Welches Wissen benötigt die ‚pädagogische Praxis‘? Wissen über Schulentwicklungsprozesse. In: Popp, U./Reh, S. (Hrsg.): Schule forschend entwickeln. Schul- und Unterrichtsentwicklung zwischen Systemzwang und Reformansprüchen. Weinheim: Juventa, S. 75-88.
- Reh, S./Schelle, C. (2006): Biographieforschung in der Schulpädagogik. Aspekte biographisch orientierter Lehrerforschung. In: Krüger, H.-H./Marotzki, W. (Hrsg.): Handbuch erziehungswissenschaftliche Biographieforschung. Wiesbaden, VS Verlag für Sozialwissenschaften, S. 391-411, 2. überarb. Aufl.
- Rolff, H.-G. (1998): Entwicklung von Einzelschulen: Viel Praxis, wenig Theorie und kaum Forschung - ein Versuch, Schulentwicklung zu systematisieren. In: Rolff, H. et al. (Hrsg.): Jahrbuch der Schulentwicklung. Band 10. Daten, Beispiele und Perspektiven. Weinheim, München: Juventa, S. 295-326.
- Rolff, H.-G. (2001): Was bringt die vergleichende Leistungsmessung für die pädagogische Arbeit an Schulen? In: Weinert, F. E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim und Basel: Beltz, S. 337-365.
- Schaarschmidt, U. (Hrsg.) (2005): Halbtagsjobber? Psychische Gesundheit im Lehrerberuf – Analyse eines veränderungsbedürftigen Zustandes. Weinheim und Basel: Beltz, 2. Aufl.
- Schaarschmidt, U./Kieschke, U. (2007): Gerüstet für den Schulalltag. Psychologische Unterstützungsangebote für Lehrerinnen und Lehrer. Weinheim und Basel: Beltz.
- Schrader, F.-W./Helmke, A. (2004): Von der Evaluation zur Innovation? Die Rezeptionsstudie WALZER: Ergebnisse der Lehrerbefragung. In: Empirische Pädagogik. 18. Jg./Heft 1, S. 140-161.
- Stahlberg, D./Frey, D. (1992): Stroebe, W./Miles, H./Codol, J.-P./Stephenson, G. M. (Hrsg.): Sozialpsychologie. Eine Einführung. Teil II: Die Konstruktion der sozialen Welt. Berlin u. a.: Springer-Verlag, S. 144-170, 2. Aufl.

## 26 Thorsten Bohl

- Stamm, M. (2003): Evaluation und ihre Folgen für die Bildung. Eine unterschätzte pädagogische Herausforderung. Münster: Waxmann.
- Stroebe, W./Jonas, K. (1992): Einstellungen I: Struktur, Messung und Funktionen. In: Stroebe, W./Miles, H./Codol, J.-P./Stephenson, G. M. (Hrsg.): Sozialpsychologie. Eine Einführung. Teil II: Die Konstruktion der sozialen Welt. Berlin u. a.: Springer-Verlag, S. 144-170, 2. Aufl.
- Tillmann, K. J./Dederich, K./Kneuper, D./Kuhlmann, C./Nessel, I. (2008): PISA als bildungspolitisches Ereignis. Fallstudien in vier Bundesländern. Wiesbaden: VS-Verlag.
- Tilman, K. J. (2008): Das Verhältnis von Bildungsforschung und Bildungspolitik am Beispiel der PISA-Studie. URL: <http://www.dji.de/cgi-bin/projekte/output.php?projekt=859&Jump1=LINKS&Jump2=17> [11.08.2008]
- Wimmer, R. (1999): Wider den Veränderungsoptimismus. Zu den Möglichkeiten und Grenzen einer radikalen Transformation von Organisationen. In: Soziale Systeme. 5. Jg./Heft 2, S. 159-190.