

Mandl, H. & Hense, J. (2007). Lässt sich der Unterricht durch Evaluation verbessern? In W. Schönig (Hrsg.), "Spuren der Schulevaluation. Zur Bedeutung und Wirksamkeit von Evaluationskonzepten im Schulalltag" (S. 85-99). Bad Heilbrunn: Klinkhardt.

Heinz Mandl und Jan Hense

Lässt sich der Unterricht durch Evaluation verbessern?

Abstract:

Obwohl unterschiedliche Verfahren der Evaluation von Schule und Unterricht zunehmend Akzeptanz finden, lautet eine bisher wenig beachtete Frage, wie Evaluationsergebnisse optimal zu einer Verbesserung der Praxis genutzt werden können. Ausgehend von einem Überblick über einige zentrale Fragen bei der Evaluation von Schule und Unterricht, werden in diesem Beitrag die Faktoren und Prozesse diskutiert, die bei einer effektiven Nutzung von Evaluationen eine Rolle spielen. Zwei Beispielfälle werden vorgestellt, die aufzeigen wie eine effektive Evaluationsnutzung in verschiedenen Umsetzungskonstellationen realisiert wurde. Abschließend wird für die Einhaltung von Standards der Evaluation und für eine verstärkte empirische Forschung über Bedingungen erfolgreicher Evaluationsnutzung plädiert.

1 Evaluation auf dem Weg zur Selbstverständlichkeit?

In den vergangenen Jahren wurde eine Vielzahl von Hinweisen dafür erkennbar, dass Ansätze und Verfahren der Evaluation sich zunehmend in der Schule etabliert und Akzeptanz gefunden haben. Auch wenn nach wie vor viele Erfahrungen von Skepsis und offenen oder verdeckten Widerständen geprägt sind, so scheinen die Zeiten der Fundamentalopposition überwunden.

Als Beispiele für diese Entwicklung lassen sich zunächst die öffentlich breit rezipierten internationalen Schulleistungsstudien wie TIMSS, PISA oder PIRLS/IGLU nennen (vgl. Weinert, 2001). Neben diesen großangelegten Evaluationsstudien gibt es aber auch eine Vielzahl lokaler und regionaler Initiativen zur Evaluation von Schule und Unterricht (z. B. Altrichter, Buhren, Burkard & Eikenbusch, 2001). Dieser Entwicklungsstand veranlasste Burkard und Eikenbusch (2000) zu einem eher optimistischen Zwischenfazit: „Nach einer Phase der konzeptionellen und bildungspolitischen Absicherung rückt nun stärker in den Vordergrund, wie Evaluation (...) zur alltäglichen Praxis werden kann“ (S. 7).

Die allgemeine Hoffnung, von der solche Bestrebungen nach einer Verankerung von Evaluation in der alltäglichen pädagogischen Praxis angetrieben werden, lautet, dass Evaluation einen Beitrag zur Qualitätssteigerung von Schule und Unterricht leisten kann. Eine im allgemeinen Evaluationsboom oft vernachlässigte Frage ist aber, ob und inwiefern sich diese Hoffnung überhaupt bewahrheitet und falls ja, unter welchen Bedingungen dies der Fall ist (vgl. Stamm, 2003). Erfüllt also Evaluation von Schule und Unterricht die in sie gesetzten Erwartungen?

Wenn man diese Frage nach den Folgen von Evaluation stellt, so sind Wirkungen auf unterschiedlichen Ebenen zu berücksichtigen. Am deutlichsten und einfachsten sind derzeit Konsequenzen auf der *Ebene des Gesamtsystems* zu identifizieren, die also das Bildungssystem als Ganzes betreffen (vgl. Baumert, 2001). Hier haben insbesondere die angesprochenen internationalen Vergleichsstudien zu einer breiten Diskussion geführt, die weit über das Bildungssystem hinausgingen. Bereits in der Folge von TIMSS sind eine Reihe von Modellversuchen und Reforminitiativen initiiert worden, die den aufgedeckten Defiziten entgegen wirken sollten.

Schwieriger ist die Frage zu beantworten, welche Konsequenzen Evaluationen auf der *Ebene von Einzelschulen* sowie auf *Unterrichtsebene* nach sich ziehen (Kotthoff, 2003). Erst relativ wenige systematische empirische Studien haben sich mit dieser Fragestellung beschäftigt. Sie fanden vor allem als Begleituntersuchungen zu vergleichenden Schulleistungsstudien statt, bei denen Ergebnisse der Leistungstests auf Schul- oder Klassenebene rückgemeldet wurden (vgl. Kohler & Schrader, 2004). Trotz eines allgemein optimistischen Untertons lassen die Ergebnisse dieser Begleituntersuchungen jedoch nicht den Schluss zu, dass Evaluation immer die in sie gesetzten Hoffnungen erfüllt.

So fand etwa Schwippert (2004), dass nur ca. 50% der Lehrkräfte überhaupt Gebrauch von der Möglichkeit einer gut verständlich aufbereiteten Evaluationsrückmeldung machten. Konsequenzen aus Evaluationsergebnissen können aber natürlich erst eintreten, wenn diese zur Kenntnis genommen und im didaktischen Handeln berücksichtigt werden. Und selbst für den Fall, dass dies geschieht, stellen Ditton und Arnoldt ernüchtert fest: „Die Erwartung, dass Feedback per se zu Verbesserungen führt, ist auf dem Hintergrund vorliegender Forschungsergebnisse jedenfalls kaum begründet“ (Ditton & Arnoldt, 2004, S. 115). Insgesamt scheint sich hier ein Befund widerzuspiegeln, der in der allgemeinen internationalen Evaluationsforschung seit der sogenannten „Nutzungskrise“ der Evaluation bekannt ist (Weiss, 1972; Patton, 1997; Leviton, 2003): Das Vorliegen von Evaluationsergebnissen allein führen nicht automatisch zu ihrer Verwendung und damit zu pädagogischen Verbesserungen.

Die Frage, der wir daher im Folgenden nachgehen, lautet: Wie kann Evaluation in der Schule so gestaltet werden, dass ihre Ergebnisse möglichst umfassend zu einer Verbesserung der Praxis bei-

tragen? Dazu ist zunächst auf den allgemeinen Begriff der Evaluation von Schule und Unterricht einzugehen, bevor dann die Prozesse diskutiert werden, die bei einer effektiven Nutzung von Evaluationen eine Rolle spielen. Schließlich zeigen wir an Hand zweier Beispiele, wie effektive Evaluationsnutzung in verschiedenen Umsetzungskonstellationen realisiert werden kann.

2 Evaluation von Schule und Unterricht

Evaluation kann definiert werden als die systematische Untersuchung von Nutzen oder Wert einer Sache (Joint Committee on Standards for Education, 1994, S. 3). Drei wesentliche Elemente machen diese Definition aus: Systematische Untersuchung, Nutzen bzw. Wert sowie der Evaluationsgegenstand („Sache“).

In Bezug auf das Element der *systematischen Untersuchung* besteht meist Konsens darüber, dass es sich dabei um ein wissenschaftlich begründbares Vorgehen handeln muss und wissenschaftliche Methoden zum Einsatz kommen müssen, um zu einem Urteil über Wert bzw. Nutzen einer Sache zu kommen. Dieser Anspruch kommt besonders deutlich im Begriff der *Evaluationsforschung* zum Ausdruck, der ansonsten aber synonym zu Evaluation zu verstehen ist.

Die Differenzierung von *Wert und Nutzen* ist folgendermaßen zu verstehen: *Wert* meint den „inneren“ Wert oder die Qualität eines Gegenstands unabhängig von einem bestimmten Verwendungskontext. *Nutzen* oder *Nützlichkeit* einer Sache ergibt sich dagegen erst angesichts konkreter Anforderungen und sind damit immer vom Verwendungskontext abhängig (vgl. Wittmann, 1985). So kann etwa eine Unterrichtsstunde von hohem didaktischem Wert sein, wenn sie aber nicht auf die Lernausgangslage der Schulklasse abgestimmt ist, in der sie gehalten wird, bleibt ihr Nutzen gering.

Gegenstand einer Evaluation können je nach Kontext und Praxisfeld ganz unterschiedliche Sachverhalte sein. Im Bereich Schule gehört die Evaluation von Schülern bzw. deren Lernleistungen ebenso wie die Evaluation von Lehrkräften in Form von dienstlichen Beurteilungen seit Langem zum alltäglichen Geschäft, wobei das Vorgehen in der Regel wenig wissenschaftlich ist. Immer noch relativ neu dagegen ist, dass systematische und vergleichende Evaluationen des Unterrichts und von Schulen als Gesamtorganisationen durchgeführt werden. Um die Vielfalt unterschiedlicher Konstellationen und Vorgehensweisen zu differenzieren, die dabei möglich sind, bietet es sich an, eine Reihe von „Grundfragen der Evaluation“ zu betrachten, wie es in den folgenden Abschnitten getan wird (vgl. Wottawa, 2001).

Warum wird evaluiert?

Die Frage nach dem Warum ist eine Schlüsselfrage jeder Evaluation, denn schließlich soll Evaluation nicht zum Selbstzweck verkommen, sondern einen Mehrwert für die Qualität im Bildungswesen erbringen. Üblicherweise werden eine ganze Reihe von Hoffnungen und Erwartungen mit Evaluati-

onen verknüpft. Bringt man diese auf einen gemeinsamen Nenner, so lassen sich vier Grundfunktionen der Evaluation unterscheiden, die mit den möglichen Ziel- oder Schwerpunktsetzungen von Evaluationsprojekten korrespondieren (vgl. Hense & Mandl, 2003):

1. *Entwicklung und Verbesserung*: Am häufigsten sollen Evaluationen Informationen liefern, die dabei helfen, die pädagogische Praxis zu verbessern. Durch den Vergleich von erwarteten Ergebnissen mit tatsächlichen Ergebnissen sollen etwa Schwachstellen im Unterricht sichtbar gemacht werden um anschließend gezielt intervenieren und fördern zu können.
2. *Entscheidungsstützung*: Oft sollen Evaluationsergebnisse auch das notwendige Wissen zur Verfügung stellen, um in Entscheidungssituationen Alternativen gegeneinander abzuwägen. Gewöhnlich handelt es sich dabei um Entscheidungen über die Einstellung, Fortführung oder Implementation von evaluierten Maßnahmen wie es beispielsweise bei Modellversuchen der Fall ist.
3. *Rechenschaft und Legitimation*: Evaluation kann auch dazu dienen, zu überwachen bzw. darüber Rechenschaft abzulegen, ob sich weisungsgebundene Personen oder Instanzen im vorgesehenen Sinne verhalten. Evaluation durch die Schulaufsicht etwa ist in Deutschland immer noch stark durch diese Funktion geprägt, auch wenn der Trend hin zu einer Ergänzung ihrer Rolle um Beratungsfunktionen geht.
4. *Lernen*: Hier konzentriert sich der Anspruch an die Evaluation darauf, mehr über Wirksamkeit und Funktionsweise einer untersuchten Maßnahme zu erfahren. Die Erkenntnisse sollen nicht direkt verwendet werden, sondern als allgemeines und generalisiertes Wissen in die pädagogische Arbeit einfließen.

Grundsätzlich schließen sich die vier genannten Funktionen nicht gegenseitig aus, das heißt, die Evaluation von Schule und Unterricht kann durchaus mehrere oder sogar alle dieser Funktionen verfolgen. Zu berücksichtigen ist dabei, dass unterschiedliche Akteure oder Beteiligte (z. B. Lehrer, Eltern, Schüler, Schulaufsicht oder Wissenschaft) bei einer Evaluation durchaus unterschiedliche Interessen vertreten können. Außerdem hat die Zielsetzung einer Evaluation gewöhnlich auch einen starken Einfluss auf das konkrete Vorgehen und die gewählten Methoden (s. u.).

Neben den vier genannten „regulären“ Funktionen lassen sich in der Praxis noch verschiedene sogenannte „Pseudo-Funktionen“ von Evaluation beobachten. Diese liegen etwa vor, wenn Evaluationen in Auftrag gegeben werden, um eine Entscheidung hinauszuzögern (vgl. Suchman, 1970) oder eine bereits gefallene Entscheidung nachträglich zu legitimieren (vgl. Legge, 1984) oder wenn sie nur pro forma und dem Augenschein nach durchgeführt werden. Solche Motive für Evaluation hat Alkin (1985) unter dem Begriff „symbolic use“ zusammengefasst. Sie entspringen der Tatsache

und verdeutlichen sie zugleich, dass Evaluation oft in komplexe soziale Zusammenhänge eingebettet ist, die durch widersprüchliche Interessen und oft auch Machtfragen geprägt sind.

Was wird evaluiert?

Diese Frage betrifft den Evaluationsgegenstand. Wie oben bereits angedeutet, wird in Schulen auf mehreren Ebenen evaluiert. Auf der Ebene des Unterrichts können der Unterricht selbst, dessen Wirkungen in Form von Schülerleistungen sowie das Lehrerhandeln Gegenstand von Evaluationen sein. Auf der Ebene der Einzelschule rücken organisationale Aspekte in den Vordergrund, wie etwa das Schulklima oder innerschulische Arbeits- und Kommunikationsstrukturen. Auf der Systemebene schließlich bewegen sich die großangelegten internationalen Studien, die nationale Bildungssysteme vergleichen, ebenso wie Studien, die verschiedene Schulsysteme innerhalb eines Landes miteinander vergleichen (z. B. Fend, 1982).

Wer evaluiert?

Das herkömmliche Evaluationsmodell der Fremdevaluation sieht vor, dass externe Instanzen oder Personen die Evaluation einer Schule oder des Unterrichts vornehmen. Diese Rolle kann je nach Konstellation etwa von der Schulaufsicht oder von Wissenschaftlern wahrgenommen werden. Im Kontext der allgemeinen Schulentwicklungsbewegung haben sich aber auch verschiedene Formen der internen oder Selbstevaluation etabliert (z. B. Buhren, Killus & Müller, 1999; vgl. Hense, in Druck). Hier liegt die Verantwortung für die Evaluation in den Schulen selbst. Oft werden dazu Qualitätsgruppen aus dem Kollegium gebildet, teils auch unter Einbeziehung von Vertretern von Eltern und Schülern.

Wie wird evaluiert?

Bei Evaluationen kann prinzipiell die ganze Bandbreite an Methoden der empirischen Sozialforschung zum Einsatz kommen. Die Wahl hängt dabei üblicherweise eng mit den oben bereits diskutierten Fragen zusammen. Während die großen Vergleichsstudien etwa aufwändige Verfahren der Testkonstruktion, der Stichprobenziehung und Datenauswertung umfassen, kann eine schulinterne Evaluation zum Thema Hausaufgabenbetreuung auch in Form einfacher ad-hoc Fragebögen durchgeführt werden, die direkt den interessierenden Sachverhalt abfragen. Im Bereich der Selbstevaluation werden oft auch alternative Methoden vorgeschlagen, die über das klassische Methodenspektrum der Sozialforschung hinausgehen und stark projektiv oder expressiv angelegt sind, wie etwa Collagen oder Fotodokumentationen (z. B. Burkard & Eikenbusch, 2000; MacBeath, Schratz, Meuret & Jakobsen, 2000). Begründen lässt sich die Methodenwahl in unterschiedlichen Evaluationskontexten durch den jeweils vorliegenden Erkenntnisanspruch, der natürlich bei einer Studie wie PISA ein anderer ist als bei einer schulinternen Evaluation, etwa zur Gestaltung des Pausenhofs.

3 Wie kommt es zur Unterrichtsverbesserung durch Evaluation?

Um zu untersuchen, wie und unter welchen Bedingungen Evaluation zu Verbesserungen auf der Unterrichtsebene führt, haben Helmke und Schrader (2001; vgl. Schrader & Helmke, 2003, 2004) im Rahmen des Projekts „Wirkungsanalyse der Leistungsevaluation: Zielerreichung, Ertrag für die Bildungsqualität der Schule und die Rückmeldung von Evaluationsergebnissen“ (WALZER) ein Rahmenmodell entwickelt, das die dabei relevanten Einflussfaktoren zusammenfassen soll (Abbildung 1).

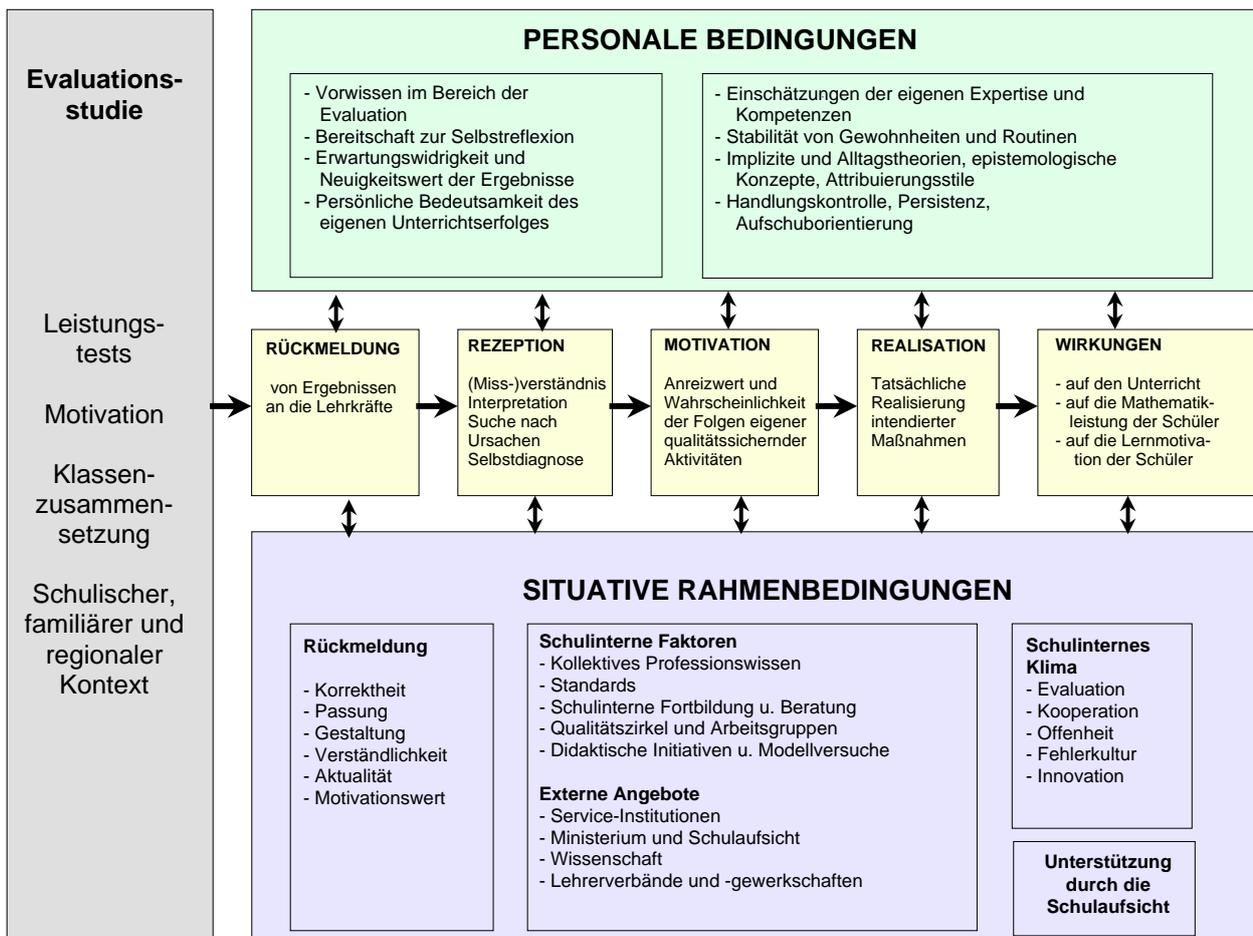


Abbildung 1: Rahmenmodell der Wirkung von Evaluationsrückmeldungen (Helmke & Schrader, 2001; Quelle der Graphik: <http://www.uni-landau.de/helmke/projekte.htm>)

Im Zentrum des Modells stehen eine Reihe von Schritten, die vollzogen werden müssen, damit die Ergebnisse einer Evaluation sich in beobachtbaren Wirkungen niederschlagen: Ausgehend von der *Rückmeldung* von Evaluationsergebnissen müssen diese zunächst einmal von relevanten Akteuren angemessen verstanden und *rezipiert* werden. Falls daraus die *Motivation* zu einer Verhaltensänderung entsteht, kommt es zur Planung und *Realisation* von Maßnahmen, die aus den Evaluationsergebnissen resultieren. Erst wenn diese Kette von Einzelschritten durchlaufen ist, können nach dem

Modell von Helmke und Schrader (2001) tatsächlich *Wirkungen* in Folge von Evaluationen eintreten.

Neben dieser Schrittfolge nennt das Modell personale und situative Rahmenbedingungen, die einen Einfluss auf die oben skizzierten Schritte nehmen. Zu den *personalen Rahmenbedingungen* gehören vor allem motivationale, aber auch verschiedene kognitive Merkmale jener Personen, die Evaluationsergebnisse aufgreifen sollen. Im Normalfall handelt es sich dabei um Lehrkräfte. Dem Rahmenmodell zufolge kommt es etwa auf deren Selbstreflexionsfähigkeit an und auf den Stellenwert, den sie dem eigenen Unterrichtserfolg beimessen. Zu den *situativen Rahmenbedingungen* der Wirkung von Evaluationsrückmeldungen gehören einerseits Faktoren, die die Gestaltung der Rückmeldung betreffen, wie etwa deren Verständlichkeit. Andererseits sind es vor allem Faktoren auf Ebene der Einzelschule, die hier eine Rolle spielen. Zu ihnen gehören das kollektive Professionswissen oder Aspekte der Schulkultur wie die allgemeine Innovationsbereitschaft. Externe Angebote zur Unterstützung der Umsetzung von Konsequenzen aus Evaluationsergebnissen vervollständigen den Bereich der situativen Rahmenbedingungen.

Ursprünglich bezieht sich das Modell von Helmke und Schrader (2001) nur auf die Frage der Wirksamkeit von Rückmeldungen im Rahmen punktueller Evaluationsvorhaben: Nach Eintreten der Wirkungen endet der im Modell skizzierte Prozess. Im Prinzip ist das Modell aber auch auf die Wirksamkeit kontinuierlich konzipierter Evaluationsbemühungen anwendbar, wie sie beispielsweise in Ansätzen des schulischen Qualitätsmanagements vorkommen. Notwendig ist dazu seine Einbettung in einen zyklischen Veränderungsprozess, bei dem die Wirkungen von Evaluation in konkreten Maßnahmen der Schul- und Unterrichtsentwicklung münden, die wiederum den Ausgangspunkt für erneute Evaluationsmaßnahmen bilden.

4 Evaluationen mit Verbesserungsfokus: Zwei Beispiele

In den folgenden Abschnitten präsentieren wir zwei Beispiele konkreter Evaluationsprojekte, die im Sinne der obigen Diskussion die Verbesserung von Schule und Unterricht zum Ziel hatten. Das erste Beispiel aus dem Projekt „Tatfunk“ folgt dem klassischen Modell der Fremdevaluation durch externe Evaluatoren, während das zweite Beispiel die Selbstevaluation im BLK-Modellversuchsprogramm SEMIK zum Inhalt hat.

4.1 Externe Evaluation im Projekt „Tatfunk“

Das Projekt „Tatfunk“ ist ein innovatives Schulprojekt, das die Förderung des unternehmerischen Denkens und Handelns in der Schule zum Ziel hat. Kern des Projekts ist ein einjähriger Wahlpflichtkurs, der in der Regel im 12. Schuljahr (Kollegstufe) an Gymnasien durchgeführt wird. Aufgabenstellung im Kurs ist die gemeinsame und weitgehend selbstständige Produktion und Vermark-

tung einer Radiosendung. Die Schüler werden dabei nach Bedarf durch einen professionellen Mediencoach und durch den kursleitenden Lehrer unterstützt.

Nach einem ersten Testlauf an einer einzelnen Münchner Schule (2002/03) und einer zweiten Erprobung an vorwiegend bayerischen Gymnasien (2003/04) wurde das Projekt Tatfunk im Schuljahr 2004/05 erstmals bundesweit auf 20 Schulen ausgedehnt. Zusätzlich wurde eine projektunterstützende virtuelle Lernplattform implementiert. Tatfunk ist ein kooperatives Vorhaben unter der übergreifenden Projektleitung der Eberhard von Kuenheim Stiftung. Projektpartner sind die BMW Group, der Bayerische Rundfunk, die Bayerische Landeszentrale für neue Medien und die Stiftung Zuhören.

Das wichtigste Lernziel des Kurses „Tatfunk“ ist der Erwerb von Kompetenzen des unternehmerischen Denkens und Handelns, das aus einer spezifischen Kombination von Teilkompetenzen auf kognitiver, motivationaler, sozialer und organisationaler Ebene besteht. Um den Erwerb dieser Kompetenzen zu ermöglichen sollte der Kurs ausgehend von den Ergebnissen der aktuellen Lehr-Lern-Forschung als selbstgesteuerter, kooperativer und problemorientierter Projektunterricht realisiert werden. Dies erfordert, dass die Schüler an authentischen und realitätsnahen Problemen arbeiten, dass Themen und Inhalte aus verschiedenen Perspektiven und in unterschiedlichen Kontexten betrachtet werden, und dass bei der selbstständigen und kooperativen Arbeit in erforderlichem Maße instruktionale Unterstützung zur Verfügung gestellt wird (vgl. Reinmann-Rothmeier & Mandl, 2001).

Zu Beginn des Kurses erfolgt nach einer Phase der gemeinsamen Themenfindung die Projekt- und Budgetplanung für den Rest des Schuljahrs durch die Schüler. Auf dieser Basis wird eine ca. einstündige Radiosendung konzipiert und realisiert. Dabei arbeiten die Schüler weitgehend eigenständig in Form von „Redaktionskonferenzen“ und können dabei nach Bedarf auf das medientechnische Know-how externer Experten („Mediencoaches“) zurückgreifen. Am Ende des Schuljahres sollte ein Abnehmer gefunden werden, an den die Sendung „vermarktet“ wird, etwa eine lokale Rundfunkstation.

Die Evaluation des Projekts in den Schuljahren 2003/04 und 2004/05 hatte das Ziel, die Umsetzung des Kurses und seine Wirkungen zu untersuchen und auf dieser Basis Hinweise zur Optimierung des Projekts zu geben (vgl. Hense & Mandl, 2004, 2005; Mandl, Hense & Schratzenstaller, 2005). Als Methoden wurden Interviews und Fragebogendaten herangezogen, die wiederholt im Kursverlauf erhoben wurden.

Insgesamt zeigten die Befunde der Evaluation, dass die Umsetzung des Projektkonzepts in den beteiligten Kursen weitgehend gelungen ist und dass zentrale Ziele des Projekts erreicht wurden. Für den Gesamterfolg des Projekts erwiesen sich eine Vielzahl von Faktoren gemeinsam verant-

wortlich. Dazu gehören in erster Linie die gemeinsame Arbeit der Schülerinnen und Schüler, das Vorliegen wichtiger organisationaler Rahmenbedingungen, die Unterstützung des Projektträgers und der Mediencoaches, die Arbeitsweise im Kurs und nicht zuletzt die Kursleitung.

Welche Konsequenzen und Wirkungen hatten die Evaluationsmaßnahmen im Projekt Tatfunk? Drei Aspekte sollen hier beispielhaft genannt werden:

1. In der Evaluation zeigte sich, dass gewisse *Mindestvoraussetzungen* für eine möglichst reibungslose Projektdurchführung wünschenswert waren. Dies betraf sowohl Merkmale der teilnehmenden Schülerinnen und Schülern (wie etwa Teamfähigkeit und Einsatzbereitschaft) als auch Merkmale der Kursleiter (etwa die Bereitschaft, Kontrolle an die Lernenden abzugeben) und der projektdurchführenden Schulen (etwa aktive Unterstützung der Schulleitung). Auf Basis dieser Ergebnisse wurden die Richtlinien zur Auswahl von Schulen, Kursleitern und Schülern für die Durchführung des Projekts überarbeitet.

2. Als ein wesentlicher Faktor für den Lernerfolg der Schülerinnen und Schüler im Bereich des unternehmerischen Denkens und Handelns, der ja ein zentrales Ziel des Projekts ist, erwies sich in der Evaluation die Umsetzung eines problemorientierten Unterrichts. Als Konsequenz für das Projekt ergab sich daraus die Notwendigkeit, ein Angebot zur spezifischen Fortbildung von Lehrkräften für das Projekt zu schaffen.

3. Unter anderem zur Umsetzung der oben genannten Änderungen im Kurs wurde im Schuljahr 2004/05 eine projektbegleitende virtuelle Lernplattform im Internet eingerichtet. Neben allgemeinen Vor- und Hintergrundinformationen zum Projekt enthält sie Inhalte für die Schüler im Projekt zu den Themen Projektmanagement, Radiojournalismus und Radiotechnik. Für die Kursleiter sind darüber hinaus vertiefte Angaben und Hinweise zur didaktischen Umsetzung des Kurses enthalten.

4.2 Selbstevaluation im Modellversuchsprogramm SEMIK

Selbstevaluation ist ein Ansatz der Evaluation, bei der die Verantwortung und zu wesentlichen Teilen auch die Durchführung einer Evaluation in den Händen jener liegt, deren pädagogische Arbeit beurteilt werden soll. Üblicherweise ist die schulische Selbstevaluation in einen zyklischen Prozess der Qualitätsentwicklung eingebettet (Abbildung 2). Dieser beginnt mit der Auswahl eines allgemeinen Evaluationsbereichs und mündet am Ende in die Planung konkreter Maßnahmen. Der letzte Schritt schafft dann eine neue Grundlage für den nächsten Einstieg in den Selbstevaluationszyklus.

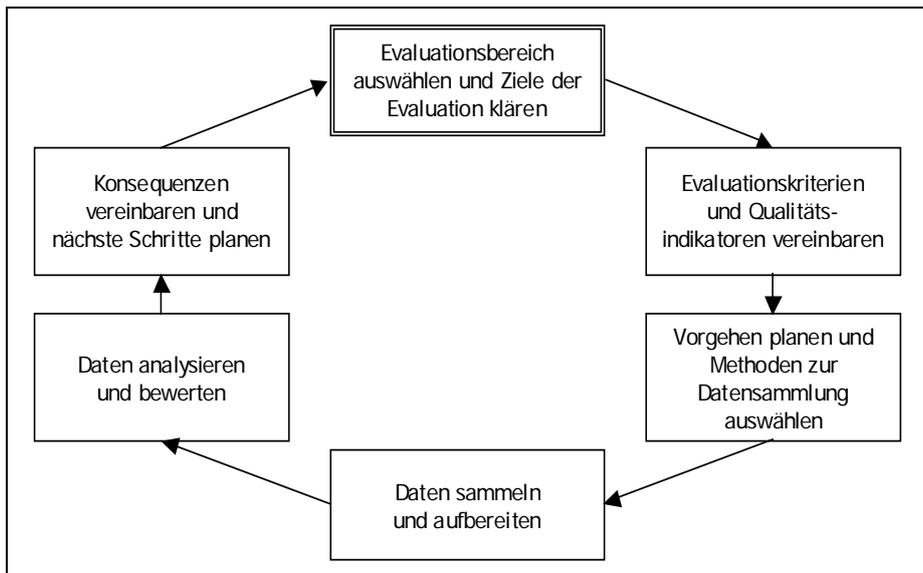


Abbildung 2: Schritte einer Selbstevaluation (nach Burkard & Eikenbusch, 2000)

Begründet werden solche Formen der selbstverantwortlichen Evaluation einerseits pragmatisch (etwa durch Kostenersparnis), vor allem aber evaluationstheoretisch. Das Argument dabei lautet, dass Evaluationen, die eigenverantwortlich von Einzelschulen durchgeführt werden, eher den spezifischen Informationsbedürfnissen der Schule entgegenkommen, mehr Akzeptanz finden und damit auch eher zu Konsequenzen führen.

Allerdings kann man auch bei dieser Form der Evaluation nicht von vorne herein davon ausgehen, dass sie automatisch zu einer Verbesserung von Schule und Unterricht führt. Aus diesem Grund untersuchten wir im Kontext des Modellversuchsprogramms SEMIK („Systematische Einbeziehung von Medien, Informations- und Kommunikationstechnologien in Lehr- und Lernprozesse“), welche Faktoren den Erfolg einer Selbstevaluation ausmachen (Hense & Mandl, 2003).

Das Ziel von SEMIK war es, Medien und insbesondere die neuen Medien im Unterricht aller Schularten und -stufen auf breiter Basis zu implementieren und dies mit der Umsetzung innovativer Lehr- und Lernformen zu verbinden (Mandl, Hense & Kruppa, 2003). Finanziert wurde das Programm im Zeitraum 1998 bis 2003 von der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK). In SEMIK waren insgesamt 25 Einzelprojekte aus allen 16 Bundesländern vertreten. Neben einer zentralen externen Evaluation, die hauptverantwortlich für die Qualitätssicherung auf Programmebene war, fanden in den Einzelprojekten verschiedene Maßnahmen zur Selbstevaluation statt. Die wissenschaftlich begleitete Selbstevaluation in vier der Einzelprojekte war Gegenstand einer quantitativen Studie auf Basis von Teilnehmerbefragungen.

Bei deren Auswertung zeigte sich, dass drei Faktoren zentral sind für eine erfolgreiche Selbstevaluation, wobei der Erfolg nach den drei Ebenen individueller Nutzen, projektbezogener Nutzen und Verhältnis von Nutzen zu investiertem Aufwand differenziert wurde (Hense, in Druck):

1. Der Aspekt der *Selbstbestimmung* der selbstevaluierenden Lehrer spielte eine herausragende Rolle für den Erfolg der Selbstevaluation. Dies galt für den individuellen Nutzen der beteiligten Lehrkräfte ebenso wie für den projektbezogener Nutzen und das Verhältnis von investiertem Aufwand zum erreichten Nutzen.
2. Positive *Erwartungen* und in weniger starkem Maße auch die Abwesenheit von *Befürchtungen* der beteiligten Lehrkräfte zu Beginn einer Selbstevaluation waren ebenfalls wichtige Erfolgsfaktoren, die einen deutlichen Zusammenhang mit Nutzeneinschätzungen zeigten. Während positive *Erwartungen* die Einschätzungen des individuellen und projektbezogenen Nutzens steigerten, machten sich *Befürchtungen* bei der Einschätzung des Kosten-Aufwands-Verhältnisses negativ bemerkbar.
3. Ebenfalls sehr wichtig war die Rolle der *Evaluationsexpertise*, die deutliche Zusammenhänge mit dem Kosten-Aufwand-Verhältnis und dem individuellen Nutzen zeigte.

Insgesamt unterstreichen die Ergebnisse, dass Selbstevaluation umso eher einen Nutzen erbringt, je mehr die beteiligten Personen bestimmte Einstellungen und Werthaltungen mitbringen. Auch wenn in der Studie nur der spezifische Ansatz der Selbstevaluation untersucht wurde, so liegt doch der Schluss nahe, dass diese Faktoren auch bei der Verwendung von Ergebnissen der externen Evaluation eine wichtige Rolle spielen dürften.

5 Ausblick

Die Frage der Verwendung von Evaluationsergebnissen ist keine neue. Ihre Diskussion im Kontext der umfangreichen internationalen Forschung über Evaluation und ihre Wirkungen hat im Zuge der Professionalisierung der Evaluation (vgl. Rossi, 1984) Einfluss auf die Formulierung der sogenannten „Standards der Evaluation“ genommen (Joint committee, 1994; deutschsprachige Fassung: Deutsche Gesellschaft für Evaluation [DeGEval], 2002). Die Standards der Evaluation sollen garantieren, dass bei der Evaluation bestimmte Mindestanforderungen eingehalten werden und damit ihre Akzeptanz und die Akzeptanz ihrer Ergebnisse steigern. Sie umfassen vier zentrale Bereiche:

1. *Nützlichkeit*: Evaluation sollte sich an den Bedürfnissen der vorgesehenen Evaluationenutzer ausrichten. Dazu ist im schulischen Bereich erforderlich, dass die Informationsbedürfnisse von Lehrkräften ermittelt und angemessene Informationsquellen herangezogen werden, oder dass vollständig, verständlich und zeitnah berichtet und die Nutzung von Ergebnissen aktiv durch das Evaluationsteam gefördert wird.
2. *Durchführbarkeit*: Evaluationen sollen realistisch konzipiert, diplomatisch durchgeführt und kostenbewusst realisiert werden. Für den schulischen Bereich, wo immer noch we-

nig Vertrautheit mit Evaluation besteht, spielt wohl vor allem ein diplomatisches Vorgehen eine besondere Rolle.

3. *Ethische Korrektheit*: Die Korrektheitsstandards des Joint Committee (1994) verlangen, dass eine Evaluation rechtlich und ethisch korrekt durchgeführt wird. Dabei wird besonders betont, dass Beteiligte und Betroffene mit wenigen oder ohne Einflussmöglichkeiten auf die Evaluation angemessen berücksichtigt werden müssen und dass entsprechende Verfahrensregeln zu vereinbaren und einzuhalten sind. In der Schule bedeutet das etwa, dass nicht nur die Belange von Schulaufsicht und Schulleitung bei Evaluationen zu berücksichtigen sind, sondern dass idealerweise auch die Interessen von Lehrkräften, Eltern und nicht zuletzt der Schüler in einem möglichst starken Umfang einbezogen werden sollten.
4. *Genauigkeit*: Evaluation sollte garantieren, dass sachlich richtige und fachlich angemessene Informationen geliefert werden. Dazu gehört nicht nur, die allgemeinen Gütekriterien empirischer Forschung und wissenschaftlichen Arbeitens einzuhalten, sondern auch eine umfassende Berücksichtigung des Umfelds. So sollte Evaluation von Schule und Unterricht immer die vorliegenden Rahmenbedingungen und Ausgangslagen berücksichtigen und nicht alleine global vorgegebene Standards überprüfen.

Dass Schule Unterricht in Zukunft nicht ohne Evaluation auskommen werden, steht inzwischen wohl außer Frage. Auch Evaluation muss sich aber die Frage gefallen lassen, ob ihre Erträge ihren Aufwand rechtfertigen. Obwohl in den vergangenen Jahren eine Reihe von Forschungsbemühungen zu beobachten sind, die die Wirkungen und die Wirksamkeit von Evaluation zum Gegenstand haben (z. B. Kohler & Schrader, 2004; Kotthoff, 2003; Stamm, 2003), sind viele Fragen in diesem Bereich nach wie vor ungeklärt.

Für die Wissenschaft erscheint es also lohnend, unterschiedliche Formen der Schul- und Unterrichtsevaluation systematisch im Hinblick auf ihre positiven und auch negativen Wirkungen hin vergleichend zu untersuchen. Für die Praxis können Standards für die Evaluation wie die des Joint Committee oder der DeGEval einen Weg aufzeigen, wie die Akzeptanz von Evaluationsmaßnahmen zu steigern ist.

Literatur

- Alkin, M. C. (1985). *A guide for evaluation decision makers*. London: Sage.
- Altrichter, H., Buhren, C. G., Burkard, Ch. & Eikenbusch, G. (Hrsg.). (2001). *Praxishilfen Evaluation* (Themenheft). *Pädagogik*, 53 (11).
- Baumert, J. (2001). *Evaluationsmaßnahmen im Bildungsbereich. Eine kritische Sicht auf mögliche*

Zugänge. *ÖFEB-Newsletter*, 2, 3-20.

- Buhren, C. G., Killus, D. & Müller, S. (1999). *Wege und Methoden der Selbstevaluation. Ein praktischer Leitfaden für Schulen (2. Aufl.)*. Dortmund: IFS.
- Burkard, C. & Eikenbusch, G. (2000). *Praxishandbuch Evaluation in der Schule*. Berlin: Cornelsen.
- Deutsche Gesellschaft für Evaluation (DeGEval). (2002). *Standards für Evaluation*. Köln: Deutsche Gesellschaft für Evaluation.
- Ditton, H. & Arnoldt, B. (2004). Schülerbefragungen zum Fachunterricht. Feedback an Lehrkräfte. *Empirische Pädagogik*, 18, 115-139.
- Fend, H. (1982). *Gesamtschule im Vergleich. Bilanz der Ergebnisse des Gesamtschulversuchs*. Weinheim: Beltz.
- Helmke, A. & Schrader, F. W. (2001). Von der Leistungsevaluation zur Unterrichtsentwicklung. Neue Herausforderungen für die Pädagogische Psychologie. In R. K. Silbereisen & M. Reitzle (Hrsg.), *Psychologie 2000* (S. 594-606). Lengerich: Pabst Science Publishers.
- Hense, J. (in Druck). *Selbstevaluation. Erfolgsfaktoren und Wirkungen eines Ansatzes zur selbstbestimmten Qualitätsentwicklung im schulischen Bereich*. Frankfurt a.M.: Lang.
- Hense, J. & Mandl, H. (2003). Selbstevaluation. Ein Ansatz zur Qualitätsverbesserung pädagogischer Praxis und seine Umsetzung am Beispiel des Modellversuchprogramms SEMIK. In *Forschungsbericht Nr. 162*. München: Ludwig-Maximilians-Universität, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie.
- Hense, J. & Mandl, H. (2004). *Projekt „Tatfunk“. Zwischenbericht zur Evaluation im Schuljahr 2003/04*. München: Ludwig-Maximilians-Universität, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie.
- Hense, J. & Mandl, H. (2005). *Projekt Tatfunk: Abschlussbericht zur Evaluation im Schuljahr 2003/04*. In *Forschungsbericht Nr. 178*. München: Ludwig-Maximilians-Universität, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie.
- Joint committee on standards for educational evaluation. (1994). *The program evaluation standards. How to assess evaluations of educational programs*. Thousand Oaks: Sage.
- Kohler, B. & Schrader, F.-W. (2004). Ergebnisrückmeldung und Rezeption. Von der externen Evaluation zur Entwicklung von Schule und Unterricht. *Empirische Pädagogik*, 18, 3-17.
- Kotthoff, H. G. (2003). *Bessere Schulen durch Evaluation? Internationale Erfahrungen*. Münster: Waxmann.
- Legge, K. (1984). *Evaluating planned organizational change*. London: Academic Press.
- Leviton, L. (2003). Evaluation use: Advances, challenges and applications. *American Journal of Evaluation*, 24, 525-535.

- MacBeath, J., Schratz, M., Meuret, D. & Jakobsen, L. (2000). *Self-evaluation in European schools. A story of change*. London: Routledge Palmer.
- Mandl, H., Hense, J. & Kruppa, K. (2003). *SEMIK - Abschlussbericht der wissenschaftlichen Programmbegleitung und zentralen Evaluation*. Grünwald: FWU Institut für Film und Bild in Wissenschaft und Unterricht.
- Mandl, H., Hense, J. U. & Schratzenstaller, A. (2005). *Tatfunk - Schüler unternehmen. Schule lernt. Abschlussbericht zur wissenschaftlichen Evaluation*. München: Eberhard von Kuenheim Stiftung, Stiftung der BMW AG.
- Patton, M. Q. (1997). *Utilization-focused evaluation. The new century text*. Thousand Oaks: Sage.
- Reinmann-Rothmeier, G. & Mandl, H. (2001). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (S. 601-646). Weinheim: Beltz.
- Rossi, P. H. (1984). Professionalisierung der Evaluationsforschung? Beobachtungen zu den Entwicklungen in den USA. In G.-M. Hellstern & H. Wollmann (Hrsg.), *Handbuch zur Evaluationsforschung* (Bd. 1, S. 654-673). Opladen: Westdeutscher Verlag.
- Schrader, F. W. & Helmke, A. (2003). Evaluation und was danach? Ergebnisse der Schulleiterbefragung im Rahmen der Rezeptionsstudie WALZER. *Schweizerische Zeitschrift für Bildungswissenschaften*, 25 (1), 79-110.
- Schrader, W. & Helmke, A. (2004). Von der Evaluation zur Innovation? Die Rezeptionsstudie WALZER. Ergebnisse der Lehrerbefragung. *Empirische Pädagogik*, 18, 140-161.
- Schwippert, K. (2004). Leistungsrückmeldungen an Grundschulen im Rahmen der Internationalen Grundschul-Lese-Untersuchung (IGLU). *Empirische Pädagogik*, 18, 62-81.
- Stamm, M. (2003). *Evaluation und ihre Folgen für die Bildung. Eine unterschätzte pädagogische Herausforderung*. Münster: Waxmann.
- Suchman, E. A. (1970). Action for what? A critique of evaluative research. In R. O'Toole (Hrsg.), *The organization, management, and tactics of social research*. Cambridge, Mass.: Schenkman.
- Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. In C. H. Weiss (Hrsg.), *Evaluating action programs: Readings in social action and education* (S. 318-326). Boston: Allyn and Bacon.
- Weinert, F. E. (Hrsg.). (2001). *Leistungsmessungen in Schulen*. Weinheim: Beltz.
- Wittmann, W. W. (1985). *Evaluationsforschung: Aufgaben, Probleme und Anwendungen*. Berlin: Springer.
- Wottawa, H. (2001). Evaluation. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 649-674). Weinheim: Beltz.