

Wirkungen und Wirksamkeit der externen Schulevaluation

Überblick zum Stand der Forschung

Vera Husfeldt

Zusammenfassung: Schulinspektionen erlangen in verschiedenen europäischen Ländern einen immer höheren Stellenwert. Gleichzeitig steigen damit auch die Erwartungen in ihre Wirksamkeit, wobei vor allem Wirksamkeit in Hinblick auf nachhaltige Schulentwicklung gemeint ist. Ausgehend von einer Darstellung der Funktionen von Schulinspektionen gibt der Artikel eine Übersicht über Wirksamkeitsmodelle und Forschungsarbeiten. Teilweise wird in den Forschungsarbeiten davon ausgegangen, dass die Wirksamkeit von Inspektionen hinsichtlich der Schulentwicklung direkt an den Schülerleistungen messbar ist. Allerdings bleibt unklar, wie die schulinternen Verarbeitungsprozesse als Bindeglied zwischen Schulinspektion und verbesserter Schülerleistung gestaltet sind. Andere Studien betrachten die Einstellungen oder Reaktionen der Akteure auf die Inspektion, wobei teils Wirksamkeitserwartungen und tatsächliche Schulentwicklung konfundiert sind. Ausgehend von der Rekonstruktion der unterschiedlichen Forschungsansätze zeigt der Artikel den Bedarf für Modelle auf, die die schulischen Verarbeitungsprozesse im Anschluss an Inspektionen stärker in den Blick nehmen.

Schlüsselwörter: Schulinspektion · Schulentwicklung · Wirksamkeit

The impact of school inspection – Does it really work? State of research

Abstract: School inspection is growing more and more important in many European countries. At the same time, the expectations in its effectiveness in terms of school development are rising. Based on a general description of functions of school inspection the article gives an overview of research in this field. One line of research is led by the assumption that the impact of inspection on school development can be measured in terms of student performance. Yet the presupposed causal chain which links inspection to student performance remains rather nebulous. A second line of research is focused on attitudinal changes of teachers and principals in the wake of school inspections—which leads sometimes to the confounding of efficacy beliefs and school effectiveness. Based on the reconstruction of research lines the author indicates the requirements of

Online publiziert: 15.06.2011

© VS Verlag für Sozialwissenschaften 2011

Prof. Dr. V. Husfeldt (✉)

Institut Forschung und Entwicklung, Fachhochschule Nordwestschweiz,

Igelweid 22,5000 Aarau, Schweiz

E-Mail: vera.husfeldt@fhnw.ch

heuristic modeling in order to provide new perspectives for research on school effectiveness and school development.

Keywords: Efficacy · School development · School inspection

1 Wirkungen und Wirksamkeit

Schulinspektionsverfahren nehmen in jüngerer Zeit im europäischen Bildungswesen einen immer größeren Stellenwert ein. Auch in den föderalen Systemen Deutschlands und der Schweiz wurde in den letzten Jahren eine Vielzahl unterschiedlicher Inspektionsverfahren implementiert, die unter Bezeichnungen wie externe Schulevaluation, Schulvisitation, Schulinspektion etc. firmieren, wobei die zugrundeliegenden Konzeptionen unabhängig von diesen Bezeichnungen oft sehr ähnlich sind.

Während in Großbritannien und in den Niederlanden bereits auf eine relativ lange Tradition der datengestützten Schulinspektion zurückgeblickt werden kann, sind die externen Schulevaluationen im deutschsprachigen Raum noch eher im Aufbau begriffen. Dementsprechend gibt es erst wenige Überlegungen zu Wirkungen und Wirksamkeit dieser neu eingesetzten Instrumente. Es gibt ebenfalls bis jetzt nur sehr wenige brauchbare Wirkungsmodelle, an denen das Design von Metaevaluationen abgestützt werden könnte.

Eine Untersuchung der Wirksamkeit bedingt, dass vorgängig Ziele formuliert worden sind, auf die sich die Wirksamkeit beziehen und an deren Erreichung sie beurteilt werden kann. Wenn beispielsweise die Entwicklung von Schülerleistungen das Ziel einer Maßnahme ist, dann kann anhand der Messung dieser Entwicklung beurteilt werden, ob eine Steigerung der Leistungen stattgefunden hat und somit die Maßnahme erfolgreich war. Wirksam ist eine Maßnahme – in diesem Fall die externe Schulevaluation – dann, wenn sie einen Prozess so beeinflusst, dass gewünschte Ziele erreicht werden können. An der Zielerreichung ist der Grad der Wirksamkeit ablesbar.

Die Wirkung einer Maßnahme ist im Gegensatz zur Wirksamkeit auch unabhängig von expliziten Zielen zu betrachten. Es ist durchaus denkbar, dass eine Maßnahme eine Wirkung hat, die im Vorfeld nicht als Ziel der Maßnahme definiert wurde bzw. erst eine Zwischenstation zur Erreichung des Zieles ist. Wenn beispielsweise mit einer Maßnahme das Ziel der Steigerung von Schülerleistungen verbunden ist und sich in einer ersten Phase ein höheres Interesse der Schülerinnen und Schüler gegenüber einer bestimmten Materie zeigt, so kann dies bereits als Wirkung angesehen werden, obwohl der Nachweis der Wirksamkeit in Hinblick auf die Leistungssteigerung noch aussteht. Ebenso können bei jeder Maßnahme auch unerwünschte Wirkungen auftreten, die den geplanten Prozess auf dem Weg zur Zielerreichung behindern können. Die erwünschten und unerwünschten Wirkungen manifestieren sich in Einstellungen und Handlungen, die im unmittelbaren Zusammenhang mit der Evaluation, mit ihrer Vorbereitung oder mit den Ergebnissen stehen. Wenn ein Kollegium im Vorfeld und zur Vorbereitung einer externen Evaluation gemeinsame Überlegungen zur Qualität von Schule und Unterricht anstellt, so wäre dies ein Beispiel für eine erwünschte Wirkung. Wenn hingegen die Ankündigung der Evaluation abwehrende und lähmende Befürchtungen auslöst, könnte dies als unerwünschte

Wirkung betrachtet werden. In diesem Sinne ist auch die Wirkung immer im Hinblick auf einen förderlichen oder hinderlichen Einfluss auf die Zielerreichung zu betrachten.

Die gesetzten Ziele stellen somit den Kern der Frage nach Wirkungen und Wirksamkeit der externen Schulevaluation dar. Während die Wirksamkeit an explizit formulierte Ziele gebunden ist, kann eine Wirkung auch ohne diese Ziele eintreten. Der vorliegende Artikel ist als Überblick zu verstehen, der – ausgehend von den Zielen der externen Schulevaluation – zum einen die Begriffe Wirkung und Wirksamkeit verdeutlichen und zum anderen den Forschungsstand zu diesem Thema darstellen möchte. Es werden dazu zwei relevante Wirkungsmodelle betrachtet und anschließend empirische Studien vorgestellt, die zu diesem Thema aus verschiedenen Ländern vorliegen. Abgeschlossen wird der Artikel durch eine Diskussion, die verdeutlichen möchte, in welchen Bereichen die Fragen zu Wirkungen und Wirksamkeit externer Schulevaluation noch unbeantwortet sind.

2 Ziele externer Schulevaluation

Die Frage nach den Zielen externer Schulevaluation oder -inspektion ist häufig nicht eindeutig zu beantworten und teils ist von einem Funktionenmix die Rede (Maritzen 2006). Es gibt neben den offiziellen Deklarationen der Evaluationsagenturen eine Reihe von eher diffusen Erwartungen aufseiten der verschiedenen Akteure. Vanhoof und van Petegem (2007) unterscheiden zwischen schulinternen und -externen Erwartungen, wobei sich die schulexternen Erwartungen noch in solche aufteilen lassen, die gesetzlich verankert sind und solche, die keine rechtliche Verankerung haben. Trotz erheblicher nationaler oder föderaler Unterschiede bei der Planung und Durchführung externer Evaluationen sind die dargestellten Zielsetzungen ähnlich. Gemeinsam ist ihnen auch die Schwierigkeit der Vereinbarkeit zwischen Entwicklung und Kontrolle. Bei allem Bemühen um eine deutliche und transparente Trennung sind die beiden Funktionen häufig konfundiert.

Brägger et al. (2005) unterscheiden zwischen einer Rechenschafts- und einer Entwicklungsorientierung als Hauptfunktionen eines ganzheitlichen Qualitätsmanagements, zwischen denen sich auch die externe Schulevaluation positionieren müsse. Sie betonen aber auch die Notwendigkeit zu klären, welches Gewicht jede der Funktionen im Evaluationsprozess ausfüllen soll. Unterschiedliche Schwerpunktsetzungen können nach Ansicht der Autoren zu unterschiedlichem Verhalten der Akteure in den Schulen und letztlich darüber auch zu anderen Ergebnissen führen. „Für Rechenschaftslegung gilt die Devise, sich von der besten Seite zu zeigen; für die Entwicklungsfunktion dagegen ist die möglichst ungeschminkte Offenlegung der Schwachstellen unabdingbar“ (Brägger et al. 2005, S. 21).

Unter der Zielsetzung der Entwicklungsorientierung ist die externe Schulevaluation als Instrument der strategischen Schulentwicklungsberatung zu verstehen. Aufgrund der Rückmeldungen durch die externe Schulevaluation können interne Maßnahmen geplant und umgesetzt werden, die die Qualitätsdefizite vermindern und positive Aspekte stärken können. Unter der Zielsetzung der Rechenschaftslegung steht hingegen die Aufsicht und Kontrolle im Vordergrund. Obwohl auch über die Funktion der Rechenschaftslegung letztlich die Sicherung und Entwicklung der Schulqualität erreicht werden soll, sind die Prozesse dabei ganz anders gesteuert. Geht es um Rechenschaftslegung und Kontrolle, werden die vorgesetzten Instanzen über die festgestellte Qualität der Schule und

über eventuelle Qualitätsdefizite informiert und sorgen sich ihrerseits um die Entwicklung oder Durchsetzung geeigneter Maßnahmen zur Qualitätssicherung. Dies geschieht in der Schweiz z. B. im kantonalen Auftrag zur Sicherstellung der Gleichwertigkeit der Bildungsangebote. In allen Kantonen, die die externe Schulevaluation bisher eingeführt haben, geht es zum einen um die Qualitätssicherung und Schulentwicklung und zum anderen um die Rechenschaftslegung, wobei die Akzentuierung praktisch überall auf der Entwicklungsfunktion liegt (Brägger et al. 2005). Beispielsweise wird im Qualitätskonzept des Kantons Appenzell Ausserrhoden die Zielsetzung folgendermaßen definiert: „Die Externe Evaluation hat in erster Linie entwicklungsorientierten Charakter. Sie dient in zweiter Linie jedoch auch der Rechenschaftslegung gegenüber der Gemeinde und der Erziehungsdirektion“ (Erziehungsdirektion Appenzell Ausserrhoden, Fachstelle Schulevaluation 2004, S. 10).

Ähnlich wie Brägger et al. (2005) identifiziert van Bruggen (2001, 2006) drei Hauptfunktionen der Inspektion:

1. Öffentliche Rechenschaftslegung über die Qualität von Schule,
2. Garantie über die Einhaltung der Vorschriften und
3. Bereitstellung eines obligatorischen Angebots für den Qualitätsentwicklungsprozess.

Der Schwerpunkt der externen Schulevaluationen liegt den Aussagen unterschiedlicher Autoren zufolge jedoch deutlich auf der Funktion der Rechenschaftslegung (Janssens 2007; Janssens und van Amelsvoort 2008; Kyriakides und Campbell 2004; van Amelsvoort et al. 2006), auch wenn weitgehend anerkannt ist, dass die externe Schulevaluation ein Katalysator für Schulentwicklung sein kann (Janssens und van Amelsvoort 2008).

In England wurde 1992 das Office for Standards in Education (Ofsted) eingesetzt, um die Inspektionen in allen Schulen zu reorganisieren. Ursprünglich war als Ziel festgelegt, über unabhängige Inspektoren öffentliche Berichte zu erarbeiten, die in der Kurzfassung den Eltern als Entscheidungshilfe für die Schulwahl zur Verfügung gestellt werden sollten. Außerdem sollte mit einem jährlichen Bericht an das Parlament und dem Beratungsangebot für Ministerien die Rechenschaftsfunktion erfüllt werden. Schulentwicklung war ursprünglich kein festgeschriebenes Ziel des Ofsted (Matthews und Sammons 2004). Trotzdem zeigt sich deutlich, wie unten noch weiter ausgeführt wird, dass die öffentlichen Erwartungen und die Erwartungen der Akteure in den Schulen auch die Schulentwicklung wesentlich mit einbeziehen. Zu den deklarierten Zielen und Prinzipien von Ofsted gehört heute explizit der Beitrag der Inspektionsergebnisse zur Schulentwicklung (Ofsted 2010). Die angestrebte Schulentwicklung soll über verschiedene Wege verlaufen:

1. Kriterien und Deskriptoren im Inspektionsrahmenprogramm sollen eine Richtlinie geben.
2. Die Schulen entwickeln die Fähigkeit und das Vertrauen, sich kritisch mit Schulqualität und Inspektionsergebnissen auseinanderzusetzen.
3. Die Inspektion schlägt Maßnahmen zur Entwicklung vor.
4. Der Dialog zwischen den Inspektoren und den schulischen Akteuren wird gestärkt.
5. Durch die Ergänzung und Bestätigung der Selbstevaluation wird die Fähigkeit zur Entwicklung gefördert.

Die Erwartungen an die englische Inspektion richten sich offenbar nicht ausschließlich auf die Funktion der Rechenschaftslegung. Ouston et al. (1997) berichten über eine Studie im Auftrag der British Educational Management and Administration Society (BEMAS), in der Schulleitende Angaben über die Bedeutung der Inspektion für die Schulentwicklung machten. Als wichtigste Punkte wurden die Vorbereitung auf die Inspektion, die Information aus den Evaluationen (speziell darüber, in welchen Bereichen Entwicklung angezeigt ist) und die Auflage, einen Aktionsplan zu entwickeln, genannt.

Ein Blick auf die aktuelle Forschungsliteratur zur Wirksamkeit externer Schulevaluation, die sich weitestgehend auf Analysen zur Schulentwicklung im Zusammenhang mit Inspektion beschränkt, lässt sogar die Vermutung entstehen, Schulentwicklung sei das alleinige Ziel der englischen Inspektion. Trotzdem sind nach einer Untersuchung von Chapman (2001) 70% der englischen Lehrpersonen der Auffassung, das Hauptziel der Ofsted-Inspektionen sei es, die Schulen für ihre Handlungen rechenschaftspflichtig zu machen. Nur 12% stimmten dieser Aussage nicht zu. Chapman fand jedoch ebenfalls, dass die Hälfte der in seiner Studie befragten Lehrpersonen glaubten, dass Ofsted mit seinen Inspektionen Veränderungen im Klassenzimmer herbeiführen könne, wobei nur ca. 20% im Sinne hatten, Veränderungen im eigenen Umfeld anzustoßen (Chapman 2001).

In den Niederlanden verfolgt die externe Evaluation ebenfalls die Ziele der Qualitätsentwicklung und der Rechenschaftslegung. Es wird der Ansatz der sogenannten proportionalen Supervision verfolgt. Das bedeutet erstens, dass die Häufigkeit und die Form der externen Evaluation von der Schulqualität abhängen. Schulen, die in der Vergangenheit positiv bewertet wurden, werden später und seltener neu evaluiert als Schulen, bei denen Defizite festgestellt werden. Zweitens basiert die externe Schulevaluation auf den Ergebnissen von internen Evaluationen. Es besteht in den Niederlanden traditionell eine freie Schulwahl. Eltern können, von einzelnen regionalen Ausnahmen abgesehen, die Schulen für ihre Kinder frei und unabhängig vom Wohnort wählen. Deshalb sind die Schulen dazu verpflichtet, regelmäßige Selbstevaluationen durchzuführen und einen „Schulführer“ herauszugeben, in dem sie über ihre Ziele und Errungenschaften berichten. Dem Dutch Supervision Act von 2002 folgend werden die Aspekte, die in der internen Evaluation ausreichend analysiert werden, nicht mehr extern erhoben.

In den Niederlanden bestehen durch diesen Ansatz unterschiedliche Formen von Rechenschaftslegung durch Schulevaluation. Neben der vertikalen (hierarchischen) externen Rechenschaftslegung unterscheiden Janssens und van Amelsvoort (2008) zwischen horizontaler Rechenschaftslegung, im Rahmen derer die Schulen ihre Schulgemeinschaft und die Akteure über Prozesse, Entscheidungen und Ergebnisse unterrichten und einer vertikalen Rechenschaftslegung, die auf die Aufsichtsgremien gerichtet ist und diesen Einblicke in die Qualität des Managements, der Strategie und der Führung liefert.

Die Berichte der externen Evaluationen werden im Internet veröffentlicht. Dadurch liegt ein Schwerpunkt der externen Schulevaluation auf der Rechenschaftslegung (Janssens 2007; Janssens und van Amelsvoort 2008).

Durch die erweiterte Schulautonomie, die den Schulen große Gestaltungsfreiheit bei der Organisation der Schule und des Unterrichts zugesteht, sollen die Ergebnisse jedoch ebenfalls dazu dienen, einen einheitlichen Standard hinsichtlich vorgegebener Qualitätsaspekte zu gewährleisten. Die Inspektion und vor allem die Selbstevaluation werden dabei als Vehikel angesehen. Janssens und van Amelsvoort (2008) stellen fest, dass

Selbstevaluation, wenn sie als Prozess aufgefasst ist, klar als Funktion oder Aspekt von Schulentwicklung angesehen wird. Wenn sie als Produkt für die Weiterverwendung in der externen Schulevaluation aufgefasst ist, wird sie in der Regel der Rechenschaftslegung zugeordnet.

In Deutschland scheint der Begriff der Rechenschaftslegung in den Darstellungen der Evaluationsagenturen eher vermieden zu werden. Er wird allenfalls im Zusammenhang mit und als Instrument der Qualitätsentwicklung verwendet. In der Erklärung des Sächsischen Bildungsinstituts auf der Webseite des Staatsministeriums für Kultus und Sport (<http://www.sachsen-macht-schule.de>) wird beispielsweise darauf verwiesen, dass die externe Schulevaluation dazu beitrage, die staatlich garantierten Bildungschancen zu sichern und den Schulen die Möglichkeit biete, auf der Basis einheitlicher Kriterien und einer Außenansicht, die Bewertung des eigenen Entwicklungsstandes einzuholen. Wie die Bildungschancen konkret gesichert werden sollen, wird nicht explizit ausgeführt. Es wird jedoch angedeutet, dass die Schulen auf der Grundlage des Evaluationsberichtes eine Qualitätsentwicklung von innen her anstoßen sollen und dafür Zielvereinbarungen mit der Schulaufsicht abschließen. Die Rolle der Schulaufsicht wird als eine beratende Rolle im Entwicklungsprozess beschrieben. Die Rechenschaftslegung scheint zumindest in den öffentlichen Erklärungen eine nachgeordnete Funktion zu haben, obwohl die Direktive zur gemeinsam mit der Schulaufsicht zu schließenden Zielvereinbarung deutlich auf eine Konfundierung der Entwicklungs- und Kontrollfunktion hinweist. Böttger-Beer und Koch (2008) nennen für Sachsen neben dem Ziel der Qualitätsentwicklung noch Ziele, die sich zwischen Erkenntnis als Dienstleistung, Legitimation/Transparenz und Kontrolle verorten lassen (vgl. dazu auch Kotthoff und Böttcher 2010).

Das Ziel der Hessischen Schulinspektion wird von Dietrich (2007) als Bereitstellung einer kriteriengestützten Rückmeldung an die Schule beschrieben, die Informationen zum aktuellen Stand der Qualitätsentwicklung, zu den Ergebnissen, Wirkungen und Stärken und zum Veränderungsbedarf liefert. In dieser Zielbeschreibung findet sich zunächst kein Hinweis auf die Intention, der Schulinspektion eine Rechenschaftslegungsfunktion zuschreiben zu wollen. Etwas deutlicher in diese Richtung formuliert Creutzburg (2006). Neben der Entwicklungsfunktion nennt sie auch das Ziel der Inspektion, den eigenverantwortlicher werdenden Schulen Rechenschaftslegung gegenüber Staat und Gesellschaft zu ermöglichen. Was hier nach einer eher freiwillig durch die Schulen initiierten Form der Rechenschaftslegung gegenüber nicht näher genannten Überinstanzen aufgefasst werden kann, konkretisiert sich bei der genaueren Betrachtung des Ablaufs und der Folgeprozesse der Inspektion. Den Evaluationsbericht erhält nämlich nicht nur die Einzelschule, sondern auch das Dezernat der Schulaufsicht. In einer Auswertungskonferenz wird der Bericht mit den Vertretern der Schule und der Schulaufsicht diskutiert. Daraufhin muss die Schule die Ergebnisse auswerten und auf der Grundlage eigener Reflexionen eine Zielvereinbarung mit der Schulaufsicht abschließen. Außerdem muss sie die neuen Schwerpunkte und Ziele in ihr Schulprogramm aufnehmen (Dietrich 2007).

In Niedersachsen wird es den Schulen in der Regel freigestellt, wie sie mit den Ergebnissen umgehen, wenn eine Schule jedoch unter Standard evaluiert wird, ist ein Auswertungsgespräch mit der Schulaufsicht obligatorisch, sodass auch hier die Kontrollfunktion deutlich zutage tritt. Ähnliche Ansätze gibt es auch in anderen deutschen Bundesländern. Immer steht die Entwicklungsfunktion im Vordergrund, wobei die Rechenschaftslegung

mitgedacht, aber wenig konkret dargestellt wird. Es gibt kaum Hinweise darauf, wie eine in Hinsicht auf Rechenschaftslegung erfolgreiche Evaluation aussehen könnte.

Auf der Grundlage dieser Zusammenstellung der Erwartungen und Ziele externer Schulevaluation und mit Bezug auf die oben dargestellte Definition von Wirksamkeit und Wirkung kann festgehalten werden, dass eine externe Schulevaluation dann als wirksam angesehen werden kann, wenn tatsächlich Schulentwicklung stattfindet oder wenn – was möglicherweise dem ersten Ziel entgegensteht – die zusammengetragenen Informationen dazu dienen, die Aufsicht und Kontrolle über die Schule zu ermöglichen bzw. zu verbessern. Sehr konkret sind diese Hinweise auf die Zielvorgaben jedoch nicht. Was ist genau unter Schulentwicklung zu verstehen? Wie soll sie ausgestaltet sein? Wer ist an ihr beteiligt? Zu was soll sie führen? Welche Art von Aufsicht und Kontrolle wird angestrebt? Welche Informationen werden dafür benötigt? Dies sind Beispiele für Fragen, deren Beantwortung notwendig ist, wenn man mit dem Blick auf feste Ziele beurteilen möchte, ob eine externe Evaluation wirksam ist und nicht lediglich unbestimmte zielungerichtete Wirkungen zeigt.

3 Forschungsstand zu Wirkungen und Wirksamkeit der externen Schulevaluation

In der Schweiz bilden bisher die Erfahrungen aus der Interaktion zwischen Evaluatoren und Akteuren in der Schule eine zentrale Grundlage für die Weiterentwicklungen der Methoden externer Schulevaluation. Dabei stehen in erster Linie Wirksamkeitsvermutungen im Vordergrund, weil bislang noch keine empirisch erhärteten Befunde zur tatsächlichen Wirksamkeit im Hinblick auf die oben genannten Ziele vorliegen.

Einige, hauptsächlich aus England stammende Studien gehen gezielt auf die Wirksamkeit in Bezug auf die Entwicklung der Schülerleistung in Schulen (quasilängsschnittliche Untersuchungen) ein; die meisten Studien jedoch beschränken sich auf Einschätzungen zur Wirksamkeit bei verschiedenen Akteuren. Insgesamt stammen die Untersuchungen, die sich mit Wirksamkeit externer Schulevaluation auseinandersetzen, mehrheitlich aus England oder aus den Niederlanden. Im deutschsprachigen Raum gibt es zur Zeit nur sehr wenige Ansätze empirischer Überprüfung von Wirksamkeit, wobei der Schwerpunkt auf Befragungen zu Einstellungen gegenüber der Schulevaluation liegt.

Das Fehlen von Befunden hängt auch damit zusammen, dass bis vor Kurzem noch kaum differenzierte Wirkungsmodelle für den Bereich der externen Schulevaluation zugrunde gelegt werden konnten, die der komplexen Anlage der Evaluationsverfahren und deren Einbettung in das mehrdimensionale Steuerungssystem der Bildungsinstitutionen einigermaßen gerecht zu werden vermochten. Eine Struktur zur Verwendung und Auswirkung von Evaluationsergebnissen im Allgemeinen liefert das metatheoretische Modell von Stamm (2003), das allerdings für die hier dargestellten Ziele zu breit gefasst ist und zu wenig auf spezifische Details eines Schulevaluationsprozesses eingehen kann. Während konkrete Wirkungsmodelle für eine erfolgreiche Rechenschaftslegung noch nicht bestehen, gibt es inzwischen einzelne Modelle, die die Wirksamkeit externer Schulevaluation für die Schulentwicklung beschreiben. Allgemeine Schulentwicklungsmodelle können einen guten Überblick über die Wirkmechanismen im Zusammenhang mit der Entwicklungsfunktion liefern. Das von Reezigt und Creemers (2005) vorgestellte ESI-

Modell wurde auf der Basis unterschiedlicher Modelle und Theorien in einer Zusammenarbeit verschiedener Experten entwickelt. Es kann wie ein zusammenfassendes Modell angesehen werden, in das die Ideen unterschiedlicher Entwicklungsmodelle und Theorien eingegangen sind. Exemplarisch wird es deshalb im Rahmen dieses Forschungsüberblicks dargestellt. Das zweite Modell von Ehren und Visscher (2006), das in diesem Zusammenhang angeführt werden soll, ist das bisher einzige Modell, das speziell auf die Wirkungen externer Schulevaluation eingeht. Beide Modelle sind beschreibender Art und nicht in ihrem ganzen Umfang, sondern nur in Teilen empirisch überprüft.

3.1 Wirkungsmodelle

3.1.1 Wirkungsmodell von Reezigt und Creemers

2005 stellten Reezigt und Creemers ein Rahmenmodell vor, das von Experten unterschiedlicher Länder in einem Projekt zur effektiven Schulentwicklung (Effective school improvement ESI) auf der Grundlage von Theorien und ergänzenden Fallanalysen entwickelt wurde. Das Rahmenmodell beschreibt Bedingungen auf unterschiedlichen Ebenen, die erfüllt sein müssen, um effektive Schulentwicklung zu erreichen. Es ist nicht ausschließlich auf externe Schulevaluation beschränkt, beinhaltet diese jedoch. Das ESI-Modell (Reezigt und Creemers 2005) unterscheidet zwischen externen und internen Bedingungen. Interessant ist, dass aus diesen Bedingungen letztendlich auch eine Definition für effektive Schulentwicklung abgelesen werden kann, die zur Erforschung der Wirksamkeit von Evaluation hinsichtlich dieses Aspektes anwendbar ist.

Als externe Bedingungen werden drei Aspekte genannt:

- Druck zur Entwicklung
 - durch den Markt,
 - durch externe Evaluationen,
 - durch externe Berater oder gesellschaftliche Ansprüche.
- Ressourcen zur Entwicklung
 - Autonomie,
 - finanzielle Ressourcen und Arbeitsbedingungen,
 - lokaler Support, z. B. durch die Eltern.
- Bildungsziele
 - z. B. vorgegebene Standards, die in einem Land gelten.

Weitere drei Aspekte, Entwicklungskultur, Entwicklungsprozesse und Entwicklungsergebnisse (outcomes) beschreiben die internen Bedingungen. Sie werden in der Tab. 1 weiter ausgeführt. Gemeinsam beschreiben sie einen idealen Schulentwicklungsprozess in seinen Voraussetzungen, Prozessen und Ergebnissen.

Die Wirksamkeit einer externen Evaluation in Hinblick auf Schulentwicklung kann anhand der Überprüfung der Ausprägungen dieser Variablen beurteilt werden. Unter dem Punkt Entwicklungskultur sind einige Voraussetzungen für Schulentwicklung genannt.

Tab. 1: Interne Bedingungen der effektiven Schulentwicklung nach Reezigt und Creemers (2005)

Improvement culture	Improvement processes	Improvement outcomes
<ul style="list-style-type: none"> • Internal pressure to improve • Autonomy used by schools • Shared vision • Willingness to become a learning organisation/a reflective practitioner • Training and collegial collaboration • Improvement history • Ownership of improvement, commitment, and motivation • Leadership • Staff stability • Time for improvement 	<ul style="list-style-type: none"> • Assessment of improvement needs • Diagnosis of improvement needs • Phrasing of detailed improvement goals • Planning of improvement activities • Implementation of improvement plans • Evaluation • Reflection 	<ul style="list-style-type: none"> • Changes in the quality of the school • Changes in the quality of the teachers • Changes in the quality of Student outcomes (knowledge, skills, and attitudes)

Interessanter ist jedoch die Beschreibung eines idealen Entwicklungsprozesses, der letztlich zu den beschriebenen Entwicklungsergebnissen führen soll.

Bevor eine Schule den Entwicklungsprozess beginnt, muss Klarheit über die Bedürfnisse und ihre Begründungen bestehen. Gehen wir von dem Fokus auf Wirkungen und Wirksamkeit externer Schulevaluation aus, sind dies die Erkenntnisse, die aus dem Evaluationsbericht folgen. In der nächsten Phase werden die Grobziele heruntergebrochen und konkreter beschrieben. Es braucht eine Einigung über die Formulierung der Feinziele. Auch Indikatoren, die Aussagen über den Erfolg der Entwicklung möglich machen, müssen festgelegt werden. Danach folgt die Planung der Aktivitäten im Detail. In der Implementationsphase liegt der Fokus darauf, die in den vorangegangenen Schritten von der Schule definierten Ziele zu erreichen. Der letzte Schritt, bevor sich der Kreis schließt, ist die Überprüfung der Zielerreichung anhand der Indikatoren. Dazu können wiederum erneut Selbstevaluationen oder externe Evaluationen dienen. Mit dieser Überprüfung wird eine Beurteilung der an Zielen orientierten Wirksamkeit möglich. Deutlich betonen Reezigt und Creemers (2005) auch, dass Schulentwicklung in allen Phasen eine Koordination auf der Schulebene benötigt.

Sie unterscheiden schließlich zwei unterschiedliche Zielvarianten im Entwicklungsprozess. Zum einen sind dies Ziele, die explizit die Schüleroutcomes betreffen. Das kann eine große Spanne von Aspekten sein, die von Wissen über Fähigkeiten und Einstellungen bis hin zu Verhalten reichen. Nicht notwendigerweise muss eine Einschränkung auf kognitive Fähigkeiten erfolgen. Zum anderen werden Ziele genannt, die auf Veränderungen

hinweisen. Dies können Ziele sein, die auf die Schulorganisation, das Lehrerverhalten oder eingesetzte Materialien fokussieren. Letztendlich ist auch hier der Schüleroutcome der entscheidende Punkt, aber dennoch kann die Veränderung als Indikator für einen zukünftigen Schüleroutcome gewertet werden.

Da das Ziel des Entwicklungsprozesses hier deutlich über die Schüleroutcomes definiert ist, kann letztlich daran auch die Wirksamkeit des Prozesses abgelesen werden. Wenn die Qualität der Schule, der Lehrpersonen und des Schüleroutcomes steigt, dann wird nach diesem Modell von einer Wirksamkeit der Entwicklung ausgegangen. Die Phasen des Prozesses und die Voraussetzungen für einen erfolgreichen Prozess sind ebenfalls im Modell beschrieben. Allerdings fehlt eine qualitative Betrachtung der Phasen des Entwicklungsprozesses. Das Modell bietet keine Informationen über die nähere Ausgestaltung der Phasen im Entwicklungsprozess und klammert damit letztlich die konkreten Wirkmechanismen aus, die zur Zielerreichung führen. Fragen danach, wer welche Aktivitäten in welcher Weise planen sollte, werden beispielsweise von diesem Modell ausgespart. Das speziell auf die externe Schulevaluation ausgerichtete und im Folgenden beschriebene Modell von Ehren und Visscher zeigt eine ähnliche Unsicherheit bei der konkreten Formulierung der Wirkmechanismen.

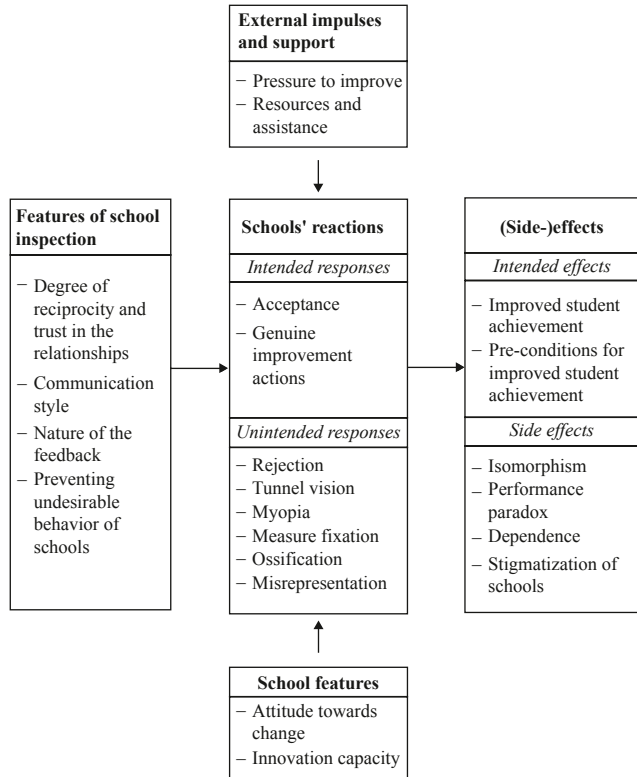
3.1.2 Wirkungsmodell von Ehren und Visscher

Ehren und Visscher (2006) gehen im Gegensatz zu Reezigt und Creemers (2005) nicht von einem Idealmodell für effektive Schulentwicklung aus, sondern versuchen auch unerwünschte Wirkungen einer externen Schulevaluation mit zu modellieren (s. Abb. 1). Als Eingangsbedingungen definieren sie die Merkmale des Schulinspektionsprozesses, auf den die Schule in unterschiedlicher Weise reagieren kann. Die Reaktionen der Schule wiederum hängen zum einen von innerschulischen Merkmalen und zum anderen von externen Impulsen und Unterstützungsmaßnahmen ab. Sie führen letztlich zu den resultierenden erwünschten bzw. unerwünschten Wirkungen.

Der Schulinspektionsprozess kann in dem Modell von Ehren und Visscher (2006) infolge der Ausprägung verschiedener Variablen unterschiedliche Wirkungen auslösen. Zu diesen Variablen gehört zum einen das wechselseitige Vertrauen im Evaluationsprozess, das einen offenen Informationsaustausch in beide Richtungen ermöglicht. Der Kommunikationsstil zwischen den verschiedenen Akteuren aufseiten des Evaluationsteams und aufseiten der Schule spielen ebenfalls eine Rolle. Die Art des Feedbacks an die Schulen wird ebenfalls als charakteristisches Merkmal des Inspektionsprozesses genannt. Schließlich berücksichtigt das Modell auch die Art der Organisation der Inspektion in einem Sinne, dass unerwünschten Effekten oder Reaktionen aus der Schule vorgebeugt werden kann. Der Einbezug des Schulpersonals in den Prozess, könnte beispielsweise eine solche vorbeugende Wirkung erzielen.

Wie im Modell von Reezigt und Creemers finden sich auch in diesem Modell externe und interne Aspekte, die die Reaktion der Schule auf die Inspektion beeinflussen können. Von externer Seite sind ebenfalls der externe Druck sowie die Ressourcen und Unterstützungsleistungen zur Entwicklung genannt. Die Aspekte Druck und Unterstützung als Motoren für die Schulentwicklung lassen sich beispielsweise auch bei Fullan (1999) finden. Die schulinternen Faktoren werden etwas knapper gefasst als im Modell von Ree-

Abb. 1: Wirkungsmodell externer Schulevaluation nach Ehren und Visscher (2006)



zigt und Creemers. Ehren und Visscher (2006) beschränken sich auf die beiden Aspekte der Einstellungen gegenüber Veränderungen und der Innovationsfähigkeit vonseiten der Schule bzw. der Schulgemeinschaft.

Bei den Reaktionen aus der Schule auf die Inspektion unterscheidet das Modell von Ehren und Visscher (2006) zwischen erwünschten und unerwünschten Reaktionen. Zu den erwünschten Reaktionen zählen sie Akzeptanz des Inspektionsprozesses und der Ergebnisse sowie die Entwicklung von Maßnahmen zur Qualitätssicherung oder Qualitätssteigerung. Anders als bei Reezigt und Creemers werden jedoch auch deutlich die möglichen unerwünschten Reaktionen herausgestellt. Dazu gehören die Zurückweisung der Ergebnisse, die Beschränkung auf Entwicklungsmaßnahmen mit kurzfristigen Zielen auf Kosten derer mit längerfristigen Zielen und das Abarbeiten von Empfehlungen ohne eine innere Akzeptanz der Problemdiagnose. Eine ähnliche Darstellung von Nebeneffekten bei Inspektionen findet sich auch bei de Wolf und Janssens (2007).

Auch die Wirkungen der Inspektion werden im Modell von Ehren und Visscher (2006) zweigeteilt betrachtet. Als erwünschte Effekte werden zum einen Leistungssteigerungen bei den Schülerinnen und Schülern genannt und zum anderen Bedingungen, die zu Leistungssteigerungen führen können. Diese Aspekte entsprechen den im Modell von Reezigt und Creemers dargestellten Wirkungen. Zusätzlich werden jedoch auch unerwünschte Wirkungen aufgeführt. Dazu gehören Entmutigung, Verringerung der Identifikation, Stig-

matisierung der Schule, negative Auswirkungen auf die Personalführung und die Konzentration auf weniger relevante Aspekte der Schulentwicklung.

Wie oben im ESI-Modell (Reezigt und Creemers 2005) zeigt sich auch in diesem Modell die Schwierigkeit der konkreten Bestimmung von Wirkmechanismen. Es ist zwar auch aus diesem Modell klar, dass eine externe Schulevaluation, die zu besserem Schüleroutcome führt, als wirksam betrachtet werden kann; wie und unter welchen konkreten Bedingungen dies im Verarbeitungsprozess der Schule jedoch möglich ist, wird nicht weiter ausgeführt. Es wird beispielsweise nicht klar, was zwischen der Akzeptanz der externen Schulevaluation sowie ihren Ergebnissen durch die innerschulischen Akteure und den entwickelten Maßnahmen steht. Auch der Zusammenhang zwischen den ausgelösten Maßnahmen und der Schülerleistung ist unklar. Dem Modell zufolge scheinen diese Maßnahmen automatisch ohne weitere Einflüsse aus der Akzeptanz hervorzugehen.

Ehren und Visscher (2006) beschreiben vier Schritte, die üblicherweise durchlaufen werden müssen, um einen, wie sie es nennen, positiven Effekt zu erreichen: 1) diagnosis, 2) initiation, mobilisation and adoption, 3) implementation or initial use und 4) continuation, incorporation and institutionalisation (Ehren und Visscher 2006, S. 61 f.). Während die Diagnose noch als Ergebnis der externen Schulevaluation gelten kann, sind die nächsten drei Schritte in Orientierung an dieser Diagnose selbstständig von der Schule zu bewältigen. Ein erfolgreicher schulinterner Verarbeitungsprozess kann jedoch nicht als selbstverständlich angenommen werden, sondern verlangt besondere Kompetenzen (van Petegem et al. 2005). Weder das Modell noch die dazu gehörende Beschreibung geben klaren Aufschluss darüber, wie eine Schule sinnvoll mit den Evaluationsergebnissen weiterarbeiten kann. Nach Huberman und Miles (1984) oder Fullan (1999) sind Partizipation, Identifikation, Unterstützung und andere Aspekte dabei wichtig; genauere Angaben über den Ablauf des Prozesses sind jedoch ausgespart. In Teilen kann diese Lücke über den Einbezug des ESI-Modells geschlossen werden, in dem zumindest die Phasen beschrieben werden, die im schulinternen Verarbeitungsprozess auf dem Weg zur erfolgreichen Schulentwicklung durchlaufen werden. Konkrete Angaben über die Ausgestaltung dieser Phasen fehlen aber auch dort.

Abschließend ist zu betonen, dass beide Modelle sich auf die Wirksamkeit der Inspektion in Bezug auf die Schulentwicklung beschränken. Die Funktion der Rechenschaftslegung ist nicht einbezogen. Das Fehlen eines Modells zur Wirksamkeit in Bezug auf Rechenschaftslegung spiegelt sich ebenfalls in den gegenwärtig vorhandenen Metaevaluationen der Schulinspektion wider. Fast ausschließlich wird darin auf die Funktion der Schulentwicklung fokussiert. Einzelne Ausnahmen, bei denen der Nutzen von Inspektionen für die Eltern untersucht wird, bilden beispielsweise die Studien von Karsten und Visscher (2001) und die von Meijer (2004). Diesen Studien zufolge nutzen Eltern in den Niederlanden nur in seltenen Fällen die Ergebnisse der Inspektion und oft haben sie nicht einmal Informationen darüber.

3.2 Empirische Studien

Wie weiter oben angesprochen, sind die Befunde aus empirischen Studien zu Wirkungen und Wirksamkeit externer Schulevaluation bisher noch eher rar. Als Grund kann zum einen der oben angedeutete Mangel an passenden Wirkungsmodellen angesehen wer-

den, zum anderen wird aber aus der Betrachtung der Modelle ebenfalls deutlich, dass sie in bestimmten Bereichen noch Lücken aufweisen. Diese Lücken scheinen sich im Forschungsstand widerzuspiegeln. Es gibt Studien, die die Einstellungen und Wirksamkeitserwartungen verschiedener Akteure in den Blick nehmen; es gibt Studien, die sich eher quantitativ mit den auf eine Evaluation folgenden Maßnahmen beschäftigen und es gibt schließlich Studien, die versuchen, den Zusammenhang zwischen externer Evaluation und den Leistungen der Schülerinnen und Schüler festzustellen. Es ist wohl kein Zufall, dass die vorhandenen Studien die Bereiche, die in den Modellen aufgeführt sind, im Wesentlichen abdecken, aber z. B. Analysen zur schulinternen Verarbeitung der Evaluationsergebnisse weitgehend aussparen. Zumindest qualitative Betrachtungen, der Verarbeitungsprozesse, die über das Auszählen von Maßnahmen und die Wirksamkeitseinschätzung zu diesen Maßnahmen hinausgehen, fehlen bis jetzt. Wie auch in den Modellen finden sich die Schwerpunkte der Studien zum einen im Bereich der Beschreibung der Bedingungen und der Entwicklungskultur der Schule sowie ihrer Akteure (z. B. Einstellungen der Akteure zur Schulevaluation und ihren Wirkungsmöglichkeiten), zum anderen in den Reaktionen der Schule (Maßnahmen und Einschätzung ihrer Wirksamkeit) und schließlich in den Schüleroutcomes.

Tatsächlich lassen sich die Studien zu Wirkungen und Wirksamkeit externer Schulevaluation grob in drei Typen charakterisieren. Es gibt eine sehr große Anzahl von eher quantitativ deskriptiven Studien, die die Einstellungen der Akteure zur externen Evaluation abfragen und Wirksamkeitserwartungen oder -vermutungen thematisieren. Dahinter steht die Idee, dass die Wirksamkeit von Inspektionen maßgeblich auch von der Einstellung der Schulakteure gegenüber Veränderungen abhängt (vgl. Standaert 2001). Ein zweiter Typ von ebenfalls hauptsächlich deskriptiven Studien beschäftigt sich mit der Betrachtung der aus den externen Evaluationen folgenden Reaktionen und den entwickelten Maßnahmen. Schließlich gibt es drittens einige Studien, zumeist aus dem angelsächsischen Raum, die die Wirksamkeit der Inspektion im Hinblick auf die Leistungen bzw. auf Leistungssteigerungen betrachten. Dabei werden Zusammenhänge zwischen Leistung und Inspektion untersucht, indem Schülerleistungen aus inspizierten Schulen mit Schülerleistungen aus Schulen verglichen werden, in denen im betrachteten Zeitraum keine Inspektionen durchgeführt wurden. Das Ofsted bietet mit dem umfangreichen Datenmaterial zu Leistungswerten dabei eine zuverlässige Quelle. Zusammen mit der längeren Tradition der instrumentenbasierten Schulinspektion scheint dies der Grund dafür zu sein, dass die meisten Studien, die gegenwärtig vorliegen, aus England stammen.

Die hier dargestellte Charakterisierung der drei Typen entspricht in etwa auch der Aufteilung der Studien bei de Wolf und Janssens (2007). Diese bilden jedoch statt drei vier Typen ab, wobei bei den Reaktionen und Folgemaßnahmen unterschieden wird zwischen dem Verhalten des Schulpersonals und der Entwicklung der Schulpolitik.

Im Folgenden ist ein Überblick über Studien zu Wirkungen und Wirksamkeit in Hinblick auf Schulentwicklung aufgeführt, der in erster Linie eine Ordnung der vorhandenen Forschungsansätze zum Ziel hat und keinen Anspruch auf Vollständigkeit der Darstellung erhebt. Es sind im Wesentlichen Studien aus der Zeit seit Mitte der 1990er-Jahre aus England und den Niederlanden aufgeführt, Ländern also mit einer längeren Evaluations-tradition. Weiter werden Studien aus Deutschland genannt, wo die Evaluation erst in jüngerer Zeit eine stärkere Bedeutung erlangt. Über Literaturdatenbanken wurde zunächst

nach Studien zu Inspektion, externer Schulevaluation, Schulvisitation etc. gesucht, die in Fachzeitschriften oder Sammelbänden veröffentlicht sind. Ausgehend von diesen Fachartikeln und ihren Literaturangaben wurden weitere Recherchen vorgenommen.

3.2.1 Forschungen zu Einstellungen und Wirksamkeitserwartungen

Bereits Mitte der 1990er-Jahre beginnen Forschende im angelsächsischen Raum die Reaktionen der Akteure in den Schulen auf die Schulinspektion zu untersuchen. Fidler und Kollegen (1994, 1995) untersuchen die Einschätzungen von Lehrpersonen zur Inspektion im Hinblick auf ihre Wirksamkeit für die Schulentwicklung und stellen fest, dass Lehrpersonen in der Regel sowohl das mündliche als auch das schriftliche Feedback durch die Inspektion als wichtigen Ausgangspunkt für die Schulentwicklung ansehen. Verschiedene Studien (z.B. Wilcox und Gray 1996; Brimblecombe et al. 1996; Kogan und Maden 1999; McCrone et al. 2006) fokussieren auch auf die Beziehung zwischen den Inspektoren und den Akteuren in der Schule, die den Angaben der Autoren zufolge in den meisten Fällen als positiv und respektvoll wahrgenommen wird (Wilcox und Gray 1996). Brimblecombe et al. (1996) zeigen, dass Lehrpersonen einem hohen Druck während der Inspektionen ausgesetzt sind. Sie vermuten, dass die Vorstellungen über die Inspektion im Vergleich mit der Inspektion selbst häufig dramatisierend sind.

Kogan und Maden (1999) führen eine Evaluation der Ofsted-Evaluation durch, bei der sie sowohl Aspekte der Schulentwicklung als auch der Rechenschaftslegung betrachten. Untersucht werden die Erfahrungen und der Umgang der Schulen mit der Inspektion, Reflexionen der Schulen über das Ofsted-System einige Zeit nach der Inspektion und die Kosten, die für die Inspektion aufgebracht werden. Die Ergebnisse zeigen, dass die Vorstellungen über die Wirksamkeit von Ofsted divergent sind. Eltern haben das größte Vertrauen in die Wirksamkeit in Hinblick auf die Bildungsqualität, die Standards und das Finanzmanagement. Schulleitende glauben hingegen am wenigsten daran. Die Aussagen der Schulleitenden zeigen, dass diese in der Regel keine Verbesserungen erwarten. Politische Akteure sind in dieser Hinsicht positiver eingestellt. Sie sprechen der Inspektion eher eine Wirkung in Hinblick auf Schülerleistungen zu als die Schulleitenden. Während die meisten politischen Akteure glauben, dass die Rechenschaftslegung durch die Inspektion verbessert wurde, kommen nahezu keine Schulleitenden oder Lehrpersonen zu diesem Schluss.

Cuckle und Broadhead (1999) untersuchen unter Anwendung unterschiedlicher Methoden die Effekte, die Ofsted auf die Schulentwicklung und die Moral der Lehrpersonen hat. Dazu wurden Daten von Schulleitenden einerseits schriftlich erhoben und andererseits wurde eine Gruppe von Schulleitenden über den eingeschätzten Nutzen der Inspektionen befragt. Zusätzlich wurde Einsicht in die Aktionspläne der Schulen genommen. Wie auch in anderen Studien stellten die Schulleitungen einzelne positive Aspekte der Inspektion dar, waren sich jedoch weitgehend einig darin, dass diese nicht der Mühe und Kosten wert seien, die mit einer Inspektion verbunden sind. Cuckle und Broadhead (1999) stellen fest, dass Schulleitende dann die Inspektion als für ihre eigene Schule nützlich ansehen, wenn sie die Qualität der Inspektion und die Kompetenzen der Inspektoren anerkennen.

Eine ähnliche Studie legen auch Fitz-Gibbon und Stephenson-Forster (1999) vor. Sie befragen Schulleitende nach der Einschätzung der Nützlichkeit der Ofsted-Inspektionen.

Hinsichtlich der Informationen, die sie für eine erfolgreiche Schulentwicklung einsetzen können, waren die Schulleitungen relativ positiv eingestellt. Etwa 40% gaben eine recht moderate Antwort zwischen „nicht viel gelernt“ und „recht viel gelernt“. 70% der Schulleitenden bezeichnen das Rahmenprogramm von Ofsted als hilfreich für das Schulmanagement und die Organisation. Wenn jedoch die Fragen auf eine geldwerte Einschätzung des Nutzens kommen, ist die Zustimmung der Schulleitenden sehr verhalten. Nur 13% meinen, die Inspektion sei ihre Kosten wert, während 54% dies verneinen.

In einer Studie von Chapman (2001) wird stärker auf die Veränderungen bzw. die intendierten Veränderungen im Schulzimmer fokussiert. Chapman ist davon überzeugt, dass Schulentwicklung nur über Veränderungen in den Klassenzimmern zu erreichen ist. Er stellt deshalb die Frage, ob Lehrpersonen ihren Unterricht nach einer Inspektion verändern. Obwohl die befragten Lehrpersonen mehrheitlich die Schulentwicklung als Ziel der Inspektion erkennen, sind nur wenige von ihnen bereit, in ihrem Unterricht etwas zu verändern. Auch Brimblecombe et al. (1995) betrachten die Intentionen der Lehrpersonen hinsichtlich Veränderungen des Unterrichts. Sie untersuchen den Zusammenhang von Ofsted-Inspektionen und Veränderung auf dem Niveau der Klasse sowie die Bereitschaft der Lehrpersonen zur Veränderung ihrer Praxis. Etwas über ein Drittel der befragten Lehrpersonen waren laut dieser Studie bereit, etwas an ihrer eigenen Praxis zu ändern.

In Nordirland wurde 1999 die erste Befragung zu Einstellungen zur Inspektion von Gray und Gardner durchgeführt. Der Hauptbefund besteht darin, dass Inspektionen häufig dazu geeignet sind, die Lehrpersonen zu verängstigen, was sich durch physische oder emotionale Probleme manifestieren kann. Diese Ängste treten vor allen im Vorfeld und während der Befragung, aber teilweise auch noch danach auf. Den Autoren zufolge sind diese Ängste zum Teil dem Auftreten des Inspektionsteams zu schulden. Gray und Gardner (1999) liefern zusammen mit Cuckle und Broadhead (1999) die Vorlage für eine deutsche Studie von Gärtner et al. (2009), die ein ähnliches Untersuchungsdesign gewählt haben.

Gärtner et al. (2009) befragen Schulleitende von Brandenburger Schulen zum Inspektionsprozess. Für Deutschland ist dies der erste Vorstoß zur Analyse der Wirksamkeit von Inspektion. Die Studie zielt zum einen konkret auf die Einstellungen der Schulleitungen und zum anderen – ähnlich wie bei Ehren und Visscher (2008), deren Studie weiter unten beschrieben wird – auf die im Anschluss an die Inspektion geplanten Schulentwicklungsmaßnahmen. Die Ergebnisse deuten darauf hin, dass sich die Schulleitenden im Vorfeld der Inspektion bemühen, den normalen Arbeitsalltag zu zeigen, nur wenige Schulleitende berichten von „window dressing“ und auch relativ selten wird berichtet, dass im Vorfeld Befürchtungen bestanden. In England scheinen diese Befürchtungen wesentlich stärker ausgeprägt zu sein, was vermutlich auch mit den schärferen Konsequenzen eines negativen Inspektionsergebnisses zusammenhängt. Die Qualität der Inspektion wurde in Brandenburg in der Regel hoch geschätzt. Bei Schulen jedoch, die gut bewertet wurden, waren die Einstellungen zur Qualität der Inspektion noch einmal positiver.

3.2.2 Forschungen zu Reaktionen auf Inspektionen und Folgemaßnahmen

Studien zu Reaktionen auf Inspektionen wurden erst in jüngerer Zeit und wesentlich seltener durchgeführt. Auch in diesem Bereich sind die englischen Studien die Vorreiter,

wobei inzwischen ebenfalls Erkenntnisse aus niederländischen und deutschen Studien vorliegen.

Lowe (1998) berichtet aus einer Fallstudie mit sieben Schulen in England, dass nur eine von ihnen die von der Inspektion vorgeschlagenen Maßnahmen zum Lehren und Lernen umgesetzt hat. Die Lehrpersonen bezögen sich dabei auf ihr Recht, die Ziele des Lehrens und Lernens bestimmen zu können. Die Untersuchung verweist auf einen relativ geringen Einfluss der Inspektion auf Veränderungen im Klassenzimmer.

In einer Studie der British Educational Management and Administration Society (BEMAS) wurden alle Schulleitenden von Sekundarschulen befragt, die 1993 (n=284) und 1994 (n=388) in England inspiziert wurden (Ouston et al. 1997). Nach Ablauf eines Jahres folgten jeweils Follow-up-Befragungen. Schulleitende schienen insgesamt zufrieden mit der Inspektion. Wichtig für die Entwicklung war aus ihrer Sicht zum einen die Vorbereitung auf die Inspektion, zum andern die Information aus den Evaluationen speziell darüber, in welchen Bereichen Entwicklung angezeigt ist und schließlich die Aufgabe, einen Aktionsplan zu entwickeln. Die Autoren schließen aus der Studie, dass die Entwicklung vor allem von drei Faktoren abzuhängen scheint:

1. die Einstellung der Schulleiterin oder des Schulleiters zur Inspektion,
2. die Art der Durchführung der Inspektion und
3. die Struktur und der Rahmen der Inspektion, wobei aufgefallen war, dass in den Inspektionen teilweise der Schwerpunkt stärker auf Unterrichtsprozesse als auf das Management gelegt wurde.

In den Niederlanden führten Ehren und Visscher (2008) eine Fallstudie mit insgesamt zehn Schulen durch. Die Schulen wurden so gewählt, dass das Innovationspotenzial zwischen ihnen variierte. Eine ähnliche Variation wurde bei der Auswahl der Inspektoren angestrebt, die sich durch unterschiedliche Inspektionsstile, direktiv bis reserviert, auszeichnen sollten. Während die Variabilität der Schulen sich tatsächlich auch im Inspektionsprozess zeigte, war das Verhalten der Inspektoren über die zehn Schulen vergleichbar.

Um die Effekte der Schulinspektion zu beurteilen, wurden in dieser Studie die Entwicklungsaktivitäten über einen Zeitraum von sechs Monaten nach der Inspektion beobachtet. Danach wurden die Schulleitenden erneut mit den Inspektionsergebnissen konfrontiert. Sie mussten angeben, welche Aspekte aus dem Inspektionsbericht sie noch erinnerten, ob sie mit dem Feedback einverstanden waren und ob dieses Feedback schließlich zu neuen Entwicklungsplänen geführt hat. Alle untersuchten Schulen hatten die Inspektionsergebnisse genutzt, um daraus Maßnahmen abzuleiten, wobei es vor allem die einfacher umzusetzenden Anregungen waren, die realisiert wurden. Ehren und Visscher (2008) nutzen offenbar die Anzahl der Entwicklungsmaßnahmen als Indikator für Schulentwicklung. Dabei stellen sie fest, dass kein Zusammenhang zwischen dem Innovationspotenzial einer Schule und den ergriffenen Maßnahmen besteht. Auch die Art oder der Umfang des Feedbacks scheint nach dieser Studie keinen Einfluss auf die Anzahl der Initiativen zu haben. Ob jedoch tatsächlich das Niveau der Schulentwicklung über die Anzahl von ergriffenen Maßnahmen zu messen ist, bleibt zumindest fraglich.

Ein Teil der bereits oben angeführten Untersuchung von Gärtner et al. (2009) widmete sich ebenfalls der Frage nach den Maßnahmen, die einer Inspektion folgen. Zunächst betrachten die Autoren aber auch, wie in den Schulen mit den Inspektionsergebnissen

konkret umgegangen wird. Es zeigt sich, dass die Information über den Bericht sehr unterschiedlich gehandhabt wird. Die Spanne reicht hier vom einfachen Auslegen des Berichts für Interessierte bis hin zu aktiver Einforderung von Rückmeldungen von allen Betroffenen. Häufig gibt es Diskussionen in der Lehrerkonferenz und teilweise in der Schulkonferenz, wo auch darüber abgestimmt wird, ob ein Bericht veröffentlicht werden soll. Dies ist aber fast ausschließlich nur bei Schulen der Fall, die gut bewertet wurden. Wie bei Ehren und Visscher (2008) werden auch bei Gärtner et al. (2009) die ergriffenen Maßnahmen ausgezählt, wobei bei 3,8 durchschnittlichen Maßnahmen der Schwerpunkt bei der Schulprogrammarbeit und der schulinternen Evaluation liegt. Der Begriff schulinterne Evaluation wird aber offenbar von den Schulleitenden in unterschiedlicher Weise verstanden. Die offenen Antworten dazu zeigen, dass beispielsweise Maßnahmen zur Fortbildung im Bereich Binnendifferenzierung dazu gezählt werden. Nur selten werden hingegen in diesem Zusammenhang interne Evaluationsmaßnahmen genannt, die Befragungen von Lehrpersonen, Schülern oder anderen Betroffenen beinhalten. Häufig beziehen sich die Maßnahmen auf den Unterricht; beispielsweise wird oft der Vorsatz genannt, mehr Binnendifferenzierung einzuführen. Zwischen der Anzahl der Maßnahmen und der Bewertung der Schule durch die Inspektion lässt sich auch in dieser Studie – wie bei Ehren und Visscher (2008) – kein Zusammenhang feststellen.

Im Bericht der Befragung zur Wirkung von Schulvisitation in Brandenburg (Gärtner und Wurster 2009) wird außerdem die Einschätzung der Schulleitungen zum Zusammenhang zwischen den Maßnahmen und dem Visitationsbericht aufgegriffen. Demnach stehen in über 50% der Fälle, die meisten oder alle Ziele der im Anschluss an die Visitationen ausgearbeiteten Zielvereinbarungen im Zusammenhang mit den Empfehlungen des Visitationsberichts. Noch deutlichere Zustimmung der Schulleitungen berichtet Huber (2006) aus einer Befragung zur externen Evaluation in Bayern. Hier sind es 98% der Schulleitenden, die angeben, dass die Zielvereinbarungen die wichtigsten Empfehlungen aus dem Evaluationsbericht aufgreifen. Auch stellt Huber fest, dass der Katalog der abgeleiteten Maßnahmen die Arbeitsfelder abdeckt, die sich auch als Handlungsfelder aus einer Auswertung der Evaluationsberichte in der Pilotphase ergeben haben. Daraus schließt er, dass die Empfehlungen der Evaluationsberichte in der Regel bei der Planung der Maßnahmen berücksichtigt werden.

Trotz dieser positiven Wirksamkeitseinschätzung seitens der Schulleitungen ist durchaus auch Vorsicht bei der Interpretation angebracht. Es kann vermutet werden, dass soziale Erwünschtheit bei den Antworten der Schulleitenden eine Rolle spielt, zumal die Schulleitenden zur Teilnahme an dieser Befragung von offizieller Seite verpflichtet wurden. Da Schulleitende offenbar generell vor allem sich selbst, nicht andere Akteure der Schule, als Motor der Schulentwicklung wahrnehmen (Gärtner et al. 2009), ist ein teilweise erheblicher Druck zu sozial erwünschten Antworten denkbar.

3.2.3 *Forschungen zu Schülerleistungen*

Die meisten der Studien zu Schülerleistungen basieren, wie weiter unten aufgeführt wird, auf Analysen von Testscores. Scanlon (1999) verfolgt einen anderen Ansatz. Sie untersucht den Effekt auf Schülerleistungen über die Einschätzungen von Lehrpersonen und Schulleitenden. Vor allem in schwächer bewerteten Schulen hatten die Schulleitungen

und Lehrpersonen mehrheitlich die Hoffnung, dass die Inspektion positiv zur Leistungssteigerung der Schülerinnen und Schüler beitragen könne. In anderen Schulen war diese Erwartung eher verhalten, nur ein Drittel der Schulleitungen und ein Viertel der Lehrpersonen waren in den besser bewerteten Schulen der Meinung, Inspektion trage positiv zur Leistungsentwicklung bei.

Eine der ersten Untersuchungen zur Entwicklung von Schülerleistungen im Zusammenhang mit der Inspektion auf Grundlage von Testdaten stammt von Cullingford und Daniels (1999). Für diese Untersuchung nutzten sie umfangreiches Datenmaterial von ca. 50.000 Schülerinnen und Schülern und leiteten anhand einer logistischen Regressionsanalyse her, welchen Einfluss unterschiedliche Bedingungen auf das erfolgreiche Abschneiden im General Certificate in Education (GCSE) haben. Es zeigt sich anhand der Studie, dass es generell einen Anstieg der positiven A*-C-Bewertungen im GCSE gab, dass aber die Inspektionen einen eher negativen Einfluss auf die Schülerleistungen ausüben. Besonders stark war dieser negative Effekt bei Inspektionen, die in den Monaten Mai und Juni stattfanden. Die September/Oktober-Inspektionen hatten einen weniger negativen Einfluss.

Ebenfalls anhand einer logistischen Mehrebenenanalyse mit Daten aus über 3.000 englischen Schulen konnten Shaw et al. (2003) einige Jahre später die Ergebnisse von Cullingford und Daniels (1999) bestätigen. In den meisten der Schulen konnte kein positiver Effekt der Ofsted-Inspektionen auf die Schülerleistungen im GCSE festgestellt werden. Tendenziell war die Wirkung der Inspektion eher negativ. Nur in spezielleren, eher privilegierten Schulen, wie den girls' schools, den selective schools und den grant-maintained schools, konnten leichte positive Effekte ausgemacht werden.

Eine ähnliche Analyse von Rosenthal (2004) mit Daten aus ebenfalls über 2.000 Schulen über einen Zeitraum von vier Jahren von 1994–1998 bestätigt diese Ergebnisse erneut. Während Cullingford und Daniels (1999) noch nicht sicher waren, ob die negativen Effekte evtl. dadurch bedingt sein könnten, dass zunächst schwächere Schulen inspektioniert wurden, konnte Rosenthal nachweisen, dass es keinen Zusammenhang zwischen dem Ofsted-Besuch und den Schülerleistungen im Vorjahr gab. Das Hauptergebnis ihrer Studie ist die Feststellung, dass die Inspektionen einen hindernden Effekt auf die Entwicklung der Schülerleistungen haben, wobei grundsätzlich und unabhängig von der Inspektion die Schülerleistungen stetig steigen.

Trotz der Ergebnisse dieser Studien bleibt die Hoffnung auf Wirksamkeit auch hinsichtlich der Schülerleistungen bestehen. Matthews und Sammons (2004) kommen in ihrer Studie, die sie im Auftrag von Ofsted durchführen, zu Ergebnissen, die sie zu gegensätzlichen Interpretationen veranlassen. Sie stellen, wie auch Cullingford und Daniels (1999) sowie Rosenthal (2004) fest, dass die Leistungen der Schülerinnen und Schüler in der Zeit seit 1996 stetig gestiegen sind. Dies schreiben sie jedoch der Einführung der Inspektion zu und nicht einer allgemein zu verzeichnenden Leistungssteigerung. Die Unterschiede im Vergleich zur Studie von Cullingford und Daniels erklären sie damit, dass in den ersten Jahren der Inspektion von 1993–1997 zunächst die Schulen zur Inspektion ausgewählt wurden, die eher Anlass zur Besorgnis boten. Berücksichtigt man heute jedoch die Ergebnisse der Studie von Rosenthal (2004), die möglicherweise zum Zeitpunkt der Veröffentlichung der hier beschriebenen Untersuchung noch nicht vorlagen, ist diese Deutung nicht mehr haltbar.

Für die USA können Hanushek und Raymond (2005) sowie Jacob (2005) zeigen, dass die Einführung eines neuen Rechenschaftssystems im Rahmen des Programms „No child left behind“ dazu geführt hat, dass die Ergebnisse von Leistungstests besser ausfallen. Zumindest gilt dies nach Hanushek und Raymond (2005) für die weiße und hispanische Bevölkerung der USA, sodass neben dem Effekt der Leistungssteigerung auch der Effekt zu einer höheren Disparität zu nennen ist.

Auch in den Niederlanden kann eine Studie von Luginbuhl et al. (2007) zeigen, dass die Inspektionen generell zu besseren Leistungen im CITO-Test geführt haben, wobei intensivere Inspektionsprozeduren zu noch besseren Leistungen führen als weniger intensive Prozeduren.

Auffällig ist bei allen Studien zu Schülerleistungen, dass grundsätzlich und möglicherweise unabhängig von den externen Evaluationen die Leistungen bei nationalen Vergleichstests in den untersuchten Ländern stetig steigen. Die Leistungsentwicklungen bei PISA können diesen Trend hingegen nicht bestätigen. Im Vergleich der in 2000 und 2009 gemessenen mittleren Leseleistung fällt beispielsweise ein enormer Leistungseinbruch bei Schülerinnen und Schülern aus dem Vereinigten Königreich auf. In den USA und den Niederlanden¹ ist zwar dieser massive Leistungsrückgang nicht zu verzeichnen, aber Leistungssteigerungen, wie sie die nationalen Testergebnisse suggerieren, sind ebenfalls nicht festzustellen. Daniel Koretz (2010) weist darauf hin, dass eine stark testbasierte Rechenschaftslegung dazu führen kann, dass die Testergebnisse in die Höhe getrieben werden und dadurch Bildungsprozesse irreleiten können (vgl. auch Koretz 2011 und Tymms 2011). So sind also auch Studien, die steigende Testleistungen mit externen Evaluationen in Zusammenhang bringen wollen, durchaus mit Vorsicht zu genießen.

Alle Studien, die die Wirksamkeit der Inspektion in Hinblick auf Schülerleistungen analysieren, untersuchen Kurzzeiteffekte. Fraglich bleibt dabei, ob eine auf den Ergebnissen einer Inspektion aufbauende nachhaltige Schulentwicklung nicht eine längere Zeitspanne benötigt, bevor sie sich in den Schülerleistungen niederschlagen kann. De Wolf und Janssens (2007) plädieren deshalb für eine Unterscheidung zwischen Kurzzeiteffekten und Langzeiteffekten.

3.2.4 Zusammenfassende Betrachtung der empirischen Studien

Wie oben bereits angedeutet wird aus der Zusammenstellung des Forschungsstandes deutlich, dass viele der Studien sich auf eine deskriptive Analyse des Inspektionsprozesses, der Einschätzungen und Wirksamkeitserwartungen der Akteure, der geplanten Folgemaßnahmen und der eingeschätzten Wirksamkeit solcher Maßnahmen konzentrieren. Teilweise handelt es sich auch um Einzelfallstudien mit beschränkter Repräsentativität. Lediglich die Studien zu den Schülerleistungen sind in der Regel breiter angelegt und greifen die Wirksamkeit in einer Weise auf, die über die Einschätzung der internen Akteure hinausgehend ein objektiv gemessenes Kriterium, nämlich die Testleistung der Schülerinnen und Schüler, mit einbezieht. Dabei zeigt sich meist, dass die Leistung im Zusammenhang mit der externen Evaluation kurz- bis mittelfristig eher leidet. Langfristige Effekte sind jedoch – wohl nicht zuletzt wegen der Vielzahl von möglichen leistungsbeeinflussenden Faktoren – nicht untersucht. Der schulinterne Verarbeitungsprozess, der in sehr unterschiedlicher Weise auf die Ergebnisse der externen Schulevaluation folgt

und als Ausdruck der ersten unmittelbaren Wirkung der Evaluation angesehen werden kann, ist durch die bisherigen Forschungsstudien relativ schwach beleuchtet. Zwar gibt es Untersuchungen zu den geplanten und realisierten Maßnahmen, jedoch wird dabei in der Regel der Fokus auf die Quantität und nicht auf die Qualität der Maßnahmen gerichtet. Die Maßnahmen werden als Wirkung der externen Schulevaluation angesehen, ohne dass überprüft wird, ob sie hinsichtlich der von der Schule aufgrund des Evaluationsberichts formulierten Zielvereinbarungen auch wirksam sind bzw. sein können. Studien, die die Qualität der Maßnahmen erfassen fehlen bislang. Insgesamt liegt der Schwerpunkt der Studien eher auf der Erfassung einer zielunabhängigen Wirkung bzw. der Wirksamkeitseinschätzung durch die Akteure als auf einer an konkret formulierten Zielen abgeglichenen Wirksamkeit.

4 Diskussion und Ausblick

Die breite Vielfalt der Studien, die durch diese Darstellung verdeutlicht werden sollte, weist eindrücklich auf die Komplexität des Feldes hin. Wirkungen externer Schulevaluation werden anhand der Ausleuchtung der unterschiedlichsten Facetten sinnvoll untersucht. Es lohnt sich jedoch, die Ziele, die mit der externen Evaluation verbunden sind, deutlich zu definieren und die Analysen mit einem Wirkungsmodell zu verknüpfen, das an Theorien und bisherige empirische Befunde anschließt. Nur auf diese Weise können auch Studien angebunden werden, die spezieller auf die an Zielen orientierte Wirksamkeit der externen Schulevaluation eingehen. Die Modelle von Reezigt und Creemers (2005) sowie Ehren und Visscher (2006) können dafür wichtige Impulse liefern. Sie sind speziell für die Qualitätsentwicklung an den Schulen ausgelegt und orientieren sich an dem Ziel, bessere Schüleroutcomes zu erreichen. In beiden dargestellten Modellen spielen Aspekte der Schulkultur und des Evaluations- und Entwicklungsprozesses eine Rolle. Die konkrete Ausgestaltung des innerschulischen Verarbeitungsprozesses ist jedoch sehr unscharf modelliert.

Die dargestellten empirischen Studien gehen mehrheitlich auf die Schulkultur und die mit der Evaluation verbundenen Entwicklungsprozesse ein, indem sie Einstellungen und Reaktionen der Akteure abfragen und teilweise auch die ergriffenen Maßnahmen auszählen oder von den Akteuren bewerten lassen. Das Hauptproblem dieser Herangehensweise ist die Einschränkung der Ergebnisse auf den engen Bereich der Akzeptanz und der Einstellungen der möglicherweise stark durch den Druck zur sozialen Erwünschtheit beeinflussten Akteure. Um die Wirksamkeit der externen Schulevaluation in Bezug auf das Ziel der Qualitätsentwicklung beurteilen zu können, ist ein möglichst unverfälschter Einblick in die Prozeduren notwendig, die in den Schulen vor, während und nach der Evaluation ablaufen. Dies ist zu einzelnen Aspekten in einigen der dargestellten Studien bereits gut realisiert worden. Die „improvement processes“ aus dem Modell von Reezigt und Creemers (2005) können für weitere Studien einen wertvollen Ansatzpunkt bieten, müssen aber noch näher beschrieben werden.

Insbesondere ist im Zusammenhang mit dem Entwicklungsprozess auch zu fragen, wer dafür die Verantwortung übernimmt. Betrachtete man die Schule allein, wäre dies wohl zu kurz gegriffen. Die Rahmenbedingungen, unter denen eine Schule in den Ent-

wicklungsprozess einsteigt und die Unterstützungsleistungen, die sie dabei beispielsweise durch die Schulaufsicht oder durch staatliche bzw. föderale Finanzierungen erhält, spielt möglicherweise eine entscheidende Rolle bei der Frage nach der Wirksamkeit des Prozesses. Auch die Verfahren der externen Schulevaluation unterscheiden sich teilweise sehr stark in ihrer Organisation und wahrscheinlich in der Qualität der Durchführung, Erkenntnisgewinnung und der Validität der Empfehlungen. Auch hier ist zu vermuten, dass diese Aspekte entscheidend zur Wirksamkeit der externen Evaluation beitragen.

Während jedoch die Rahmenbedingungen für einen wirksamen Entwicklungsprozess in beiden dargestellten Modellen angeführt sind und die Charakteristika der externen Schulevaluation zumindest in Ansätzen im Modell von Ehren und Visscher (2006) aufgenommen sind, ist der schulinterne Verarbeitungsprozess sowohl in den Modellen als auch in den Forschungsstudien zur Zeit noch weitgehend vernachlässigt.

Es wird die Aufgabe zukünftiger Studien in diesem Bereich sein, die Lücken vor allem im Zusammenhang mit dem innerschulischen Verarbeitungsprozess aufzufüllen und die Schulentwicklung in einen breiteren Kontext zu stellen, der Rahmenbedingungen und unterschiedliche Inspektionsverfahren mit berücksichtigt.

Die hier angesprochenen Forschungsdesiderata beziehen sich auf Untersuchungen im Zusammenhang mit dem Evaluationsziel der Qualitätsentwicklung. Für den Bereich der Rechenschaftslegung fehlen bis anhin brauchbare Wirkungsmodelle. Diese müssten neben der schulischen und unterrichtlichen Ebene auch die administrative Ebene mit einbeziehen (Dedering und Müller 2008). Ein erster Schritt zur Bearbeitung dieser Evaluationszielrichtung wäre die Entwicklung eines entsprechenden Wirkungsmodells.

Anmerkung

- 1 Da die Niederlande in PISA 2000 die notwendigen Beteiligungsquoten nicht erreicht haben, wurden an dieser Stelle die Ergebnissen aus PISA 2003 und PISA 2009 verglichen.

Literatur

- Amelvoort, G. H. W. C. H. van, Bos, K. T., Janssens, F. J. G., Klaver, L., Lelyveld, J., & Pol, M. (2006). *Proportional supervision and school improvement from an international perspective*. Utrecht: Inspectie van het Onderwijs.
- Böttger-Ber, M., & Koch, E. (2008). Externe Schulevaluation in Sachsen – ein Dialog zwischen Wissenschaft und Praxis. In W. Böttcher, W. Bos, H. Döbert, & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive* (S. 253–264). Münster: Waxmann.
- Brägger, G., Bucher, B., & Landwehr, N. (2005). *Schlüsselfragen zur externen Schulevaluation*. Bern: h.e.p.
- Brimblecombe, N., Ormston, M., & Shaw, M. (1995). Teachers' perceptions of school inspection: A stressful experience. *Cambridge Journal of Education*, 25(1), 53.
- Brimblecombe, N., Ormston, M., & Shaw, M. (1996). Teachers' perceptions of inspections. In J. Ouston, P. Earley, & B. Fidler (Hrsg.), *Ofsted-Inspections: the early experience*. London: Fulton.

- Bruggen, J. C. van (2001). *Functions of inspectorates in Europe*. Vortrag: International Inspection Academy, Berlin.
- Bruggen, J. C. van (2006). *Schulinspektion im internationalen Vergleich*. Vortrag: Bilanztagung der Hessischen Schulinspektion, Frankfurt a. M.
- Chapman, C. (2001). Changing classrooms through inspection. *School Leadership Management*, 2(1), 59–73.
- Creutzburg, C. (2006). Die externe Evaluation der Schulen in Hessen: Konzept. In Institut für Qualitätsentwicklung (Hrsg.), *Qualitätsentwicklung durch externe Evaluation – Konzepte – Strategien – Erfahrungen: Dokumentation der Fachtagung vom 30. Juni – 1. Juli 2005 in Wiesbaden* (S. 99–114). Wiesbaden: Institut für Qualitätsentwicklung.
- Cuckle, P., & Broadhead, P. (1999). Effects of Ofsted inspection on school development and staff morale. In C. Cullingford (Hrsg.), *An inspector calls – Ofsted and its effect on school standards* (S. 176–187). London: Kogan.
- Cullingford, C., & Daniels, S. (1999). Effects of Ofsted inspections on school performance. In C. Cullingford (Hrsg.), *An inspector calls – Ofsted and its effect on school standards* (S. 59–69). London: Kogan.
- Dedering, K., & Müller, S. (2008). Schulinspektion als Gegenstand wissenschaftlicher Forschung. In S. Müller, K. Dedering, & W. Bos (Hrsg.), *Schulische Qualitätsanalyse in Nordrhein-Westfalen. Konzepte, erste Erfahrungen, Perspektiven* (S. 337–347). Köln: LinkLuchterhand.
- Dietrich, J. (2007). Qualitätsentwicklung von Schule und Unterricht in Hessen. In W. Böttcher & H. Kotthoff (Hrsg.), *Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung* (S. 199–221). Münster: Waxmann.
- Ehren, M. C. M., & Visscher, A. J. (2006). Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, 54(1), 51–72.
- Ehren, M. C. M., & Visscher, C. J. (2008). The relationships between school inspections, school characteristics and school improvement. *British Journal of Educational Studies*, 56(2), 205–227.
- Erziehungsdirektion Appenzell Ausserrhoden, Fachstelle Schulevaluation. (2004). *Qualitätskonzept für die Volksschule im Kanton Appenzell Ausserrhoden*. Herisau: Appenzeller Medienhaus.
- Fidler, B., Ouston, J., & Earley, P. (1994). *Ofsted inspections and their contribution to school development*. Vortrag: Annual Conference of the British Educational Research Association, Oxford University, Sept.
- Fidler, B., Earley, P., & Ouston, J. (1995). *Ofsted school improvement through inspection*. Vortrag: European Conference on Educational Research, Bath University, Sept.
- Fitz-Gibbon, C. T., & Stephenson-Forster, N. J. (1999). Is Ofsted helpful? In C. Cullingford (Hrsg.), *An inspector calls – Ofsted and its effect on school standards* (S. 97–118). London: Kogan.
- Fullan, M. (1999). *Die Schule als lernendes Unternehmen*. Stuttgart: Klett-Cotta.
- Gärtner, H., & Wurster, S. (2009). *Befragung zur Wirkung von Schulvisitation in Brandenburg. Ergebnisbericht*. Berlin: Institut für Schulqualität der Länder Berlin und Brandenburg.
- Gärtner, H., Hüsemann, D., & Pant, H. A. (2009). Wirkungen von Schulinspektion aus Sicht betroffener Schulleitungen. Die Brandenburger Schulleiterbefragung. *Empirische Pädagogik*, 23(1), 1–18.
- Gray, C., & Gardner, J. (1999). The impact of school inspections. *Oxford Review of Education*, 25(4), 455–468.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297–327.
- Huber, F. (2006). *Konsequenzen aus der externen Evaluation an Bayerns Schulen: Auswertung einer Befragung von Schulleiterinnen und Schulleitern zu den Entwicklungen an ihren Schulen*. München: ISB.
- Huberman, A. M., & Miles, M. B. (1984). *Innovation up close. How school improvement works*. New York: Plenum.

- Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics*, 89(5–6), 761–796.
- Janssens, F. J. (2007). Supervising the quality of education. In W. Böttcher & H. Kotthoff (Hrsg.), *Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung* (S. 113–132). Münster: Waxmann.
- Janssens, F. J., & Amelvoort, G. H. van (2008). School self-evaluations and school inspections in Europe: An exploratory study. *Studies in Educational Evaluation*, 34, 15–23.
- Karsten, S., & Visscher, A. J. (2001). Ervaringen met het openbaar maken van schoolprestaties in Engeland en Frankrijk. In A. B. Dijkstra, S. Karsten, R. Veenstra, & A. J. Visscher (Hrsg.), *Het oog der natie: scholen op rapport; standaarden voor de publicatie van schoolprestaties* (S. 36–53). Assen: Koninklijke Van Gorcum BV.
- Kogan, M., & Maden, M. (1999). An evaluation of evaluators: The Ofsted system of school inspection. In C. Cullingford (Hrsg.), *An inspector calls – Ofsted and its effect on school standards* (S. 9–31). London: Kogan.
- Koretz, D. (2010). *Some implications of current policy for educational measurement*. <http://www.k12center.org/publications.html>. Zugegriffen: 1. März 2011.
- Koretz, D. (2011). Lessons from test-based education reform in the U.S. In I. Gogolin, J. Baumert, & A. Scheunpflug (Hrsg.), „*Transforming education*“. *Large-scale reform projects and their effects – German and international experience* (ZfE Sonderh. 13, S. 9–23). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kotthoff, H. G., & Böttcher, W. (2010). Neue Formen der „Schulinspektion“: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 295–325). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kyriakides, L., & Campbell, R. J. (2004). School self-evaluation and school improvement: a critique of values and procedures. *Studies in Educational Evaluation*, 30(1), 23–36.
- Lowe, G. (1998). Inspection and change in the classroom: Rhetoric and reality? In P. Earley (Hrsg.), *School improvement after inspection – School and LEA Responses*. London: Paul Chapman Publishing.
- Luginbuhl, R., Webbink, D., & Wolf, I. de (2007). *Do school inspections improve primary school performance?* Den Haag: CPB Netherlands Bureau for Economic Policy Analysis.
- Maritzen, N. (2006). Schulinspektion in Deutschland. In H. Buchen, L. Horster, & H.-G. Rolff (Hrsg.), *Schulinspektion und Schulleitung* (S. 7–26). Stuttgart: Raabe.
- Matthews, P., & Sammons, P. (2004). *Improvement through inspection*. London: Ofsted.
- McCrone, T., Rudd, P., Blenkinsop, S., & Wade, P. (2006). *Impact of section 5 inspections: maintained schools in England*. Slough: National Foundation for Educational Research.
- Meijer, A. (2004). *Vreemde ogen dwingen: de betekenis van internet voormaatschappelijke controle in de publieke sector*. Den Haag: Boom Juridische uitgevers.
- Ofsted. (2010). *The framework for school inspection: the framework for school inspection in England under section 5 of the Education Act 2005, from September 2009*. London: Ofsted.
- Ouston, J., Fidler, B., & Earley, P. (1997). What do schools do after Ofsted school inspections or before? *School Leadership and Management*, 17(1), 95–104.
- Petegem, P. van, Verhoeven, J. C., Buvens, I., & Vanhoof, J. (2005). *Zelfevaluatie en beleidseffectiviteit in Vlaamse scholen. Het gelijke onderwijskansenbeleid als casus*. Gent: Academia Press.
- Reezigt, G. J., & Creemers, B. P. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement*, 16(4), 407–424.
- Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review*, 23(2), 143–151.
- Scanlon, M. (1999). *The impact of Ofsted inspections*. Slough: National Foundation for Educational Research.

- Shaw, I., Newton, D. P., Aitkin, M., & Darnell, R. (2003). Do Ofsted inspections of secondary schools make a difference to GCSE results? *British Educational Research Journal*, 29(1), 63–75.
- Stamm, M. (2003). *Evaluation*. Münster: Waxmann.
- Standaert, R. (2001). *Inspectorates of education in Europe: a critical analysis*. Leuven: Acco.
- Tymms, P. (2011). Evidence? The impact of large-scale reform in England. In I. Gogolin, J. Baumert, & A. Scheunpflug (Hrsg.), „Transforming education“. Large-scale reform projects and their effects – German and international experience (ZfE Sonderh. 13, S. 105–115). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Vanhoof, J., & Petegem, P. van (2007). Matching internal and external evaluation in an era of accountability and school development: lessons from a flemish perspective. *Studies in Educational Evaluation*, 33, 101–119.
- Wilcox, B., & Gray, J. (1996). *Inspecting schools: holding schools to account and helping them to improve*. Buckingham: Open university press.
- Wolf, I. F. de, & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education*, 33(3), 379–396.