

Editorial zum Schwerpunktthema: Schule und Evaluation

Editorial to the Focus Topic: School and Evaluation

Manchmal erweist es sich als besonders schwierig, sich über Selbstverständlichkeiten zu verständigen. In Bezug auf das Thema „Schule und Evaluation“ scheint gerade dies zuzutreffen: Vermittelt über die unterschiedlichsten Reforminstrumente, ausgehend von Schulprogrammarbeit über Schulinspektion bis hin zu Vergleichsarbeiten auf der Systemebene, hat ein „evaluationsorientiertes Denken“ in der Schulorganisation in den letzten Jahren eine solche Bedeutung gewonnen, dass es fast schon naiv anmutet, die schlichte Frage zu stellen, was denn eigentlich Evaluation sei.

Während diese Frage vor zwanzig Jahren im Schulkontext noch als legitim gegolten hätte, erweckt der oder die Fragende heutzutage beim Gegenüber den Eindruck, dass die letzten zwei Dekaden Schulentwicklungsdiskurs an ihm oder ihr vorübergegangen sein müssen. Gleichwohl – so unsere Überlegungen zum vorliegenden Themenschwerpunkt – erscheint uns die Klärung oder Reflexion dieser Frage nicht nur legitim, sondern notwendig.

Womöglich verhält es sich mit dem Evaluationsbegriff und unserem Denken über Evaluation nicht sehr viel anders als mit dem Qualitätsbegriff: Für diesen gilt in besonderem Maße, dass er nur als relationaler Begriff überhaupt Sinn macht; zugleich wird er aber von vielen Akteuren mit einer so großen Selbstverständlichkeit verwendet, dass vollkommen aus dem Blick gerät, dass eigentlich in gar keiner Weise geklärt ist, was es denn mit dem Begriff der „Qualität von Schule“ auf sich hat. Möglicherweise ist es auch nicht zufällig, dass dieses Schicksal auf beide „Modebegriffe“, den der „Evaluation“ und den der „Qualität“, zutrifft, die aus dem gemeinsamen semantischen Feld stammen, das insgesamt die Schulentwicklungsdiskussion in den letzten Jahren so stark geprägt hat.

Dies aber ist genau der entscheidende Punkt, denn die Tatsache, dass sich auf den verschiedenen Ebenen der Handlungskoordination im Mehrebenensystem „Schule“ eine hohe Ausdifferenzierung der Praxen und Praktiken der Evaluation und Qualitätssicherung findet und diese kaum noch infrage gestellt werden, erzeugt eine Pfadabhängigkeit im Denken und Handeln, die es kritisch zu reflektieren gilt.

Ein aktuelles Beispiel für den Hochschulbereich ist das Urteil des Bundesverfassungsgerichts vom 18.03.2016, demzufolge die Akkreditierung von Studiengängen in Nordrhein-Westfalen durch externe Qualitätssicherungsagenturen in der bislang praktizierten Form für unzureichend erklärt wurde. Begreift man dies, d.h. die Akkreditierung von Studiengängen, als ein zentrales Merkmal der universitären Qualitätssicherung und die Begehungen durch die Peers als Evaluationsmoment, dann wird deutlich, wie viel Ungeklärtes in unserem Denken über Evaluation mitschwingt – bis hin zu der besorgten Nachfrage, wer denn nunmehr an den Universitäten für die Qualitätssicherung zuständig bzw. wie nach einer solch harschen Kritik Qualitätssicherung neu zu denken sei – und ob es hier durch externe Agenturen tatsächlich notwendig zu einer Qualitätssteigerung kommt. Aus dem Blick geraten dabei beispielsweise die Praktiken der Qualitätssicherung durch die Wissenschaftlerinnen und Wissenschaftler selbst, die in den Jahren vor der Etablierung der Akkreditierungsagenturen gepflegt wurden und nunmehr durch den vermehrten Ruf nach Systemakkreditierung wieder aufleben. Das Forschen und Lehren an deutschen Universitäten wurde ja nicht erst durch Bologna qualitativ voll. Überträgt man diese Irritationen aus dem Hochschulsektor auf das Schulsystem, so wird deutlich, dass wir uns immer wieder einmal zentraler Grundbegriffe der Schulentwicklung kritisch-reflexiv vergewissern müssen, um hier nicht blind zu werden, wo es doch gerade eine der entscheidenden Aufgaben von Evaluationen selbst sein sollte, blinde Flecken aufzudecken.

Im einführenden Beitrag von *Wolfgang Böttcher* und *Jan Hense* wird dementsprechend die Erfolgsgeschichte der Evaluation kritisch in den Blick genommen, um sich zentrale Elemente wie die Bedeutung formativer und summativer Evaluation oder auch die Implikationen der Differenzen zwischen Selbst- und Fremdevaluation zu vergegenwärtigen sowie die Bedeutung von Evaluation im Feld und die damit notwendig werdende Qualitätssicherung für Evaluation selbst, die in den letzten Jahren in der öffentlichen Diskussion über Standards der Evaluation ihren Niederschlag gefunden hat.

Welche unterschiedlichen Vorstellungen von Evaluation von den verschiedenen Akteuren im Feld vertreten werden, dokumentiert der Beitrag von *Jutta Wolff*. Zugleich konzentriert sie sich hierbei auf ein bedeutsames Missverständnis im Kontext von Evaluationen: die überhöhten Hoffnungen („Goldstandard“) auf Programmevaluationen im Sinne von Wirksamkeitsstudien, die mit (quasi-)experimentellen Designs den Eindruck erzeugen, Kausalbeziehungen im Feld nachweisen zu können. Bedeutsam ist dieses Missverständnis in zweierlei Hinsicht: Erstens zeigt es die übersteigerten Ansprüche, die Auftraggeber an Evaluationen erheben, ohne dass diese forschungsmethodisch und forschungspraktisch erfüllt werden könnten, und zweitens die starke Fokussierung auf Wirksamkeitsstudien, die mit ihren Kausalimplikationen Technologisierbarkeitsphantasien schüren und somit unerfüllbare Hoffnungen gegenüber dem sozialwissenschaftlichen Instrument Evaluation auslösen.

Wie man demgegenüber reflektierter mit der Idee von Programmevaluation umgehen kann, illustriert *Susanne Giel* anhand eines fiktiven Beispiels von Willkommensklassen. Sie verdeutlicht, wie gleichzeitig wirkungsorientiert und praxisorientiert operiert werden kann, so dass vermittelt über eine konkrete Programmtheorie keine unerfüllbaren Ansprüche gegenüber Evaluationsmaßnahmen formuliert werden.

Eine besondere Form des Praxisbezugs stellen *Sylvia Rahn*, *Sabine Gruehn*, *Miriam Keune* und *Christoph Fuhrmann* vor. Sie nehmen eben jene Akteursgruppe in den Blick, die sonst oftmals vergessen wird: die Schülerinnen und Schüler. Hierbei ist die Zwispältigkeit zu bearbeiten, dass Schulqualität sich nicht nur durch Zufriedenheitsabfragen bei Schülerinnen und Schülern erheben lässt, gleichwohl aber umgekehrt deren Perspektive für eine „professionelle Feedbackkultur“ in der Schule unentbehrlich ist.

Wie komplex das Zusammenspiel unterschiedlicher Daten hierbei – auch auf Systemebene – gedacht werden muss, illustrieren *Klaudia Schulte*, *Detlef Fickermann* und *Markus Lücken* in ihrem Bericht zum Hamburger Prozessmodell zur datengestützten Schulentwicklung, in dem sie aufzeigen, wie Daten unterschiedlicher Aggregationsebenen in ein Gesamtsystem sinnvoll integriert werden müssen, um Effekte für die Schulentwicklung auslösen zu können.

Dass allerdings die Orchestrierung eines Datenmanagements auf Systemebene für Schulentwicklung zwar eine wahrscheinlich notwendige, aber noch nicht hinreichende Bedingung ist, zeigen *Wolfgang Beywl* und *Lars Balzer* in ihrem Beitrag zum „Aufbau von Evaluationskompetenzen für interne Schulevaluation durch projektbezogene Fortbildung“. Die Autoren entwerfen praxisnah ein „Kompetenzportfolio für schulinterne Evaluationen“, um auch auf schulischer Seite die Voraussetzungen für positive Effekte evaluativen Denkens und Handelns zu schaffen.

Der Durchgang durch Strukturprobleme und Anwendungsbezüge von Evaluation im schulischen Feld zeigt das hohe Maß an Ausdifferenziertheit des Evaluationsdiskurses, das sich – weltweit – nicht zuletzt in der Existenz der zahlreichen Fachgesellschaften für Evaluation dokumentiert. Hier werden Standards für Evaluationen formuliert, die auch für alle anderen sozialen Felder jenseits der Schule gelten, wie bspw. das Gesundheits- oder das Sozialwesen. Für den deutschsprachigen Raum existieren hier die Standards der DeGEval (vgl. dazu ausführlich den Beitrag von *Wolfgang Böttcher* und *Jan Hense* im vorliegenden Heft).

Demgegenüber hat sich in der schul(-system-)bezogenen empirischen Bildungsforschung ein spezifisches Verständnis von „Evaluation“ durchgesetzt, das sich nicht zuletzt durch die bildungspolitischen Legitimationszwänge des Monitorings der öffentlichen Institution Schule ergibt und damit immer auch in der Gefahr steht, eine stärkere Nähe zum Kontroll- und weniger zum Entwicklungsparadigma zu haben.

Dieses spezifische Evaluationsverständnis der schul(-system-)bezogenen empirischen Bildungsforschung wird im abschließenden Bericht des Themenschwerpunkts entfaltet. In seinem Beitrag nimmt *Hans Merkens* dementsprechend Begriffsbestimmungen vor, systematisiert historische Theorielinien und entwirft einen Orientierungsrahmen für die unterschiedlichen Bezugspunkte von Evaluation in ihren Kontexten der Neuen Steuerung, der Bildungsstandards und anderer Bereiche des Schulsystems.

Die Beiträge dieses Themenheftes dokumentieren damit die mehrfachen Interessenslagen und Bedürfnisse gegenüber dem Evaluationsbegriff und seinen Anwendungen im Schulbereich: sowohl die kritische Reflexion von Evaluation in ihren Anwendungskontexten als auch konkrete Varianten der Operationalisierung, des Datenmanagements und der Beteiligung unterschiedlichster Akteursgruppen sowie letztendlich wiederum das Bedürfnis nach einer orientierenden Begriffsklärung – auch wenn deutlich geworden sein dürfte, dass solche Vereindeutigungen nie für alle paradigmatisch divergierenden Diskurse Geltung beanspruchen können. Die schulbezogene Evaluationsforschung stellt eben nur einen Teil des gesamtgesellschaftlich bedeutsamen Evaluationsdiskurses dar. Die Beiträge in diesem Heft belegen, dass es Sinn macht, über die vermeintliche Selbstverständlichkeit von Evaluation im Schulkontext nochmals nachzudenken, oder wie *Hans Merkens* in seinem Schlusssatz resümierend zum Evaluationsbegriff festhält: „Die Selbstverständlichkeit, mit der der Begriff verwendet wird, trägt.“

Kathrin Dederling/Martin Heinrich