

Rezension zu:

Morell, Jonathan A.: *Evaluation in the Face of Uncertainty. Anticipating Surprise and Responding to the Inevitable*. New York/London: The Guilford Press, 2010. 306 Seiten, 60,00 \$ (Hardcover), ISBN 978-1-60623-858-5

*Daniela Keller*¹

Jonathan A. Morell greift ein wichtiges Thema der Evaluationsforschung auf, zu dem bisher kaum veröffentlicht wurde: Unsicherheiten, unerwartete Ereignisse und Veränderungen, die eine Evaluation vor Schwierigkeiten stellen. *Evaluation in the Face of Uncertainty* fügt sich ein in aktuell diskutierte Themen der Evaluationsforschung: systemisches Denken, Komplexität und Programmkontext. Michael Quinn Patton schreibt über sein Modell der *Developmental Evaluation* (2011), dass diese den von Morell beschriebenen „agilen Evaluator“ benötigt. Morell hat die oft theoretische Diskussion über komplexe Systeme und unerwartete Veränderungen um einen mutigen Schritt erweitert: Welchen konkreten Herausforderungen und Misserfolgen stehen Evaluationen bei ihrer Durchführung aufgrund von ungeplanten Ereignissen gegenüber und wie können „agile Evaluationen“ konstruiert werden, die sich an diese Ereignisse anpassen? Die theoretischen Ausführungen und praktischen Ratschläge sind zum Teil hilfreich, v.a. aber dann auch kritisch zu sehen, wenn zugunsten einer Evaluation, die wenig Durchführungsrisiken beinhaltet, auf risikoreichere Evaluationsdesigns, deren erfolgreiche Implementation aber einen weitaus größeren Informationsgehalt bereitstellt, verzichtet werden soll.

Morell wollte seine Analysen an konkreten Fällen vollziehen und hat zwei Jahre lang in Evaluationsplattformen der American Evaluation Association und bei seinen eigenen Kontakten nach Evaluationen gefragt, bei denen die Evaluation oder das zu evaluierende Programm nicht

wie geplant verliefen. In zwei Jahren erhielt er 18 von Evaluationskolleg(inn)en bereitgestellte Fälle. Die relativ geringe Zahl erscheint nicht weiter verwunderlich, da es meist kein gutes Licht auf Evaluator(inn)en oder Auftraggeber(innen) wirft, wenn die Evaluation oder das Programm nicht wie geplant verlaufen. Umso gewagter ist Morells Fokussierung auf dieses Thema, wenngleich man auch berücksichtigen sollte, dass er sich in den USA befindet, wo das Verständnis für die Notwendigkeit von Evaluationsdurchführungen bereits ziemlich gefestigt ist.

Die Monographie, die ein Konzept der „agilen Evaluation“ entwickelt, überlässt es konsequenterweise auch bezüglich ihres Aufbaus der Flexibilität der Leser(innen), wie sie das Buch lesen möchten. Das erste Kapitel beschreibt, auf welche unterschiedlichen Arten man beim Lesen vorgehen kann, welche Kapitel man z.B. mit/ohne das Lesen welcher anderen Kapitel verstehen kann. Die gesammelten 18 Evaluationsfälle werden nach Morells Schlusskapitel auf jeweils drei bis sieben Seiten von den Evaluator(inn)en, die den Fall bereitgestellt haben, selbst beschrieben. Morell nimmt auf diese Fälle in seinen Ausführungen immer wieder Bezug. Er erklärt den jeweiligen Fall dann stets kurz; den ganzen Fall zu lesen, erleichtert aber das Verständnis des Buches, spätestens in den Kapiteln, die sich auf die Analyse dieser Fälle konzentrieren. 18 unterschiedliche Fälle auf insgesamt 90 Seiten hintereinander zu lesen, erscheint etwas mühselig. Einen Fall am Buchende zu lesen, wenn dieser zum ersten Mal erwähnt wird, unterbricht den Textzusammenhang. Es wäre doch

1 Centrum für Evaluation (CEval), Saarbrücken

trotz aller „Agilität“ leserfreundlicher, das Buch so aufzubauen, dass man es auch an einem Stück durchlesen könnte. Im Anschluss an die theoretischen Ausführungen der Kapitel zwei bis sieben (zum Teil mit kurzen Erläuterungen an konkreten Fällen), werden die 18 bereitgestellten Evaluationen, Beispiele aus Morells eigenen Evaluationen und von ihm erdachte Evaluationen in die Kapiteln acht bis zwölf auf Grundlage der zuvor entwickelten unterschiedlichen Theorien und Kategorien analysiert. Eine thematische Kombination von jeweils theoretischen Ausführungen und entsprechenden konkreten Falltypen anstatt der zerteilten Gliederung des Buches in theoretische Ausführungen und Fallanalysen würde den theoretischen Teil veranschaulichen. Es ist schade, dass durch den Aufbau des Buches das Lesen der interessanten Ausführungen erschwert wird.

Morell analysiert zunächst die Frage, wann Programme und Evaluationen anfällig dafür sind, nicht wie geplant zu verlaufen und entwickelt dann Konzepte für Evaluationen, die nicht durch unerwartete Ereignisse geschwächt werden. Programme haben grundsätzlich ein hohes Risiko, nicht wie geplant zu verlaufen, da sie nur Lösungen für einzelne Komponenten facettenreicher Probleme bereitstellen. Sie haben eine geringere Wahrscheinlichkeit, vom Programmplan abzuweichen, wenn ihre *Research and Development*-Komponente niedrig ist, sie über große „fidelity“, also dem Vermögen, das ursprünglich entwickelte Programmmodell beizubehalten, verfügen, sie eine große Widerstandsfähigkeit hinsichtlich ihrer Effektivität in einer Vielzahl von Settings besitzen, eine große Kenntnis der Programmumwelt vorhanden ist und sie sich nicht in ihrer Anfangsphase befinden. Die Frage, wann unerwartete Ereignisse zu Schwierigkeiten bei der Evaluation führen können, analysiert Morell mit dem „life cycle view“ und dem „social/organizational view“. Der „life cycle view“ bedeutet: Je kürzer die Dauer einer Evaluation in Bezug auf die Durchführungsphase eines Programms, desto geringer die Wahrscheinlichkeit, dass unerwartete Ereignisse des Programms die Evaluation beeinflussen. Der „social/organizational view“ unterteilt in interne und externe programmbezogene Veränderungen und evaluationsbezogene unerwartete Ereignisse, bei denen die Evaluation trotz intaktem Programm nicht stattfinden kann, z.B. wenn der Zugang zu Interviewpartner(inne)n versperrt wird. Der „social/organizational view“ gibt m.E. zwar Analysekatoren für den Ursprung von

unerwarteten Ereignissen, die Evaluationen behindern können, beantwortet aber nicht die Frage „when is surprise disruptive to evaluation?“ (S. 35), die als Überschrift über seiner Abhandlung steht.

Morell gibt drei Klassifizierungen für Maßnahmen, die angewendet werden können, um zu verhindern, dass unerwartete Veränderungen die Evaluation behindern, wobei er Veränderungen in ein idealtypisches Kontinuum mit den Enden ‚vorhersehbar‘ und ‚unvorhersehbar‘ unterteilt. Es erscheint wichtig, dass er aber auch die Frage „how much is too much?“ aufwirft und zu bedenken gibt, dass die Verwendung von zu viel Zeit und weiteren Ressourcen, um Unerwartetes zu vermeiden, eine Evaluation so komplex macht, dass das Risiko, dass etwas Unerwartetes die Evaluation vor Schwierigkeiten stellt, kontroverserweise wieder erhöht wird und die Ressourcen, die verwendet werden können, um den Informationsbedarf der Stakeholder zu bedienen, dadurch eingeschränkt werden.

Eine erste Klassifizierung sind Maßnahmen, die für das Aufdecken vorhersehbarer Veränderungen angewendet werden können. Vorhersehbare Veränderungen zur Programm- oder Evaluationsplanung können schon in frühen Phasen der Evaluationsplanung/-durchführung identifiziert werden und sind dann keine unerwarteten Ereignisse mehr. Methoden, diese Ereignisse zu antizipieren, sind die Verwendung von Theorien (z.B. life cycle behavior, Charakteristika der freien Marktwirtschaft, Systemtheorie, Programmtheorie) und Erfahrungen aus ähnlichen Programmen. Weiterhin kann eine Evaluation bereits so geplant werden, dass der Zeitrahmen zwischen Programmdurchführung und Messung der erreichten Wirkungen bezüglich zeitlicher als auch kausaler Dimensionen der Wirkungskette möglichst gering gehalten wird, wobei Morell Fälle, bei denen bewusst langfristige Wirkungen gemessen werden sollen, ausnimmt.

Die zweite Klassifizierung sind Maßnahmen, die Veränderungen von der Programm- und Evaluationsplanung während der Evaluationsdurchführung so früh wie möglich aufdecken, um Zeit für die Reaktion auf bevorstehende Probleme zu gewinnen. Morell schlägt vor, Prognosetechniken aus Planungsmethoden zu adaptieren, regelmäßiges Monitoring von Logic Models im Programmverlauf durchzuführen und diese entsprechend anzupassen sowie Logic Models, um die Systeme, in die ein Programm eingebettet ist, zu erweitern. Diese Ratschläge

bezüglich Logic Models sind m.E. sinnvolle Maßnahmen, um mit relativ geringen Ressourcen ungeplante Programmverläufe regelmäßig zu verfolgen.

Die dritte Klassifizierung ist schließlich die „agile Evaluation“, die auch bei Veränderungen, die durch die oben beschriebenen Maßnahmen nicht vorhersehbar sind, durchführbar bleibt. Die „agile Evaluation“ soll so konstruiert sein, dass sie sich unvorhergesehenen Umständen anpassen kann und Methoden verwendet, deren Risiko, dass sie nicht erfolgreich implementiert werden können, gering ist. „Agile Evaluationen“ betreffen die Wahl der Datenquellen, das Design und die Programmtheorie. Bei der Wahl der Datenquellen sollte darauf geachtet werden, dass (1) Entwicklungskosten gering sind, wenn die Datenquellen verändert werden müssen, z.B. die Datenerhebungsinstrumente nicht neu konstruiert werden müssen, (2) die Evaluation eine geringe Belastung für die Stakeholder bedeutet, da eine große Belastung der Stakeholder zu Problemen für die Evaluation und das Verhältnis zwischen Evaluator(inn)en und Stakeholdern führen könnte und (3) die Zeit, die Erhebungsmethoden bei Problemen benötigen, um verändert zu werden, mit der Zeit übereinstimmt, die dafür zur Verfügung steht. Dies ist z.B. schwierig, wenn eine Evaluation so geplant ist, dass eine Befragung nur in einem bestimmten, kurzen Zeitraum stattfinden kann. Agile Designs bedeuten, dass es (1) wenig Abhängigkeiten zwischen unterschiedlichen Teilen des Erhebungsdesigns gibt, (2) die Anzahl organisatorischer Grenzen, die überschritten werden müssen, z.B. Kontrollgruppen, die nicht Teil des Programms sind, gering sind, und (3) das Ausmaß, bis zu welchem ein Evaluationsdesign unterteilt werden kann, groß ist, also z.B. Interviews zu unterschiedlichen Zeitpunkten geplant werden. Die Programmtheorie ist umso agiler, je größer die Möglichkeit ist, dass sie unter neuen Umständen verändert werden kann, was z.B. durch ein oben beschriebenes Monitoring des Logic Models durch Stakeholder gefördert werden kann. Je größer die Möglichkeit ist, dass die Programmtheorie verändert werden kann, desto größer ist auch die Möglichkeit, dass das Evaluationsdesign verändert werden kann. Weiterhin wird bei „agilen Evaluationen“ während des gesamten Evaluationszeitraums mit den Stakeholdern kooperiert und es werden Beratergruppen für die Evaluation gebildet, die mit Methoden der Gruppendynamik moderiert werden. Mitglieder dieser Beratergruppen kommen sowohl aus unterschiedlichen Programmberei-

chen als auch aus Bereichen, die nicht Teil des Programms sind. Über diese Art der Kooperation hinaus sollten Stakeholder aber nicht weiter belastet werden.

Morell schreibt, dass Evaluationen an sich agil sind, wenn sie nicht von historischen Daten abhängen und einen Fokus auf Langzeitwirkungen ebenso wie Kontrollgruppen und rigide Datenerhebungsinstrumente vermeiden. Wenn man obige Ausführungen beachtet, fällt auf, dass demnach auch partizipative Evaluationsansätze, bei denen Stakeholder z.B. Aufgaben bei der Datensammlung übernehmen, nicht agil sind, da sie eine große „Belastung“ für die Stakeholder bedeuten. Designs, die Programmwirkungen nachweisen können, also auf Baseline-Studies und Kontrollgruppen beruhen, sind nicht agil. Morell gesteht ein, dass eine solche agile Evaluation den Informationsbedürfnissen der Stakeholder nicht genügen könnte. Eine Lösung sieht er darin, zu prüfen, ob man auch ohne Methoden auskommt, deren erfolgreiche Durchführung bei Veränderungen schwieriger ist (also Kontrollgruppen, rigide Datenerhebungsinstrumente etc.) und dass man Agilität bei der Wahl der Methoden berücksichtigen sollte. Über konkrete ihm bereitgestellte Evaluationsfälle schreibt er: „These evaluations could produce credible findings without control groups, without preinnovation historical data, and without long-term follow-up past the formal ending of the program“ (S. 181). Die m.E. grundlegende Frage, wie sehr für eine agile Evaluation, die eine größere Chance hat, erfolgreich durchgeführt zu werden, auf ein anspruchsvolles Evaluationsdesign, deren erfolgreiche Durchführung zu einem weitaus größeren Informationsgehalt führt, verzichtet wird, erscheint damit aber noch nicht ausreichend diskutiert. Anstatt auf grundlegende Bestandteile wertvoller Evaluationen zu verzichten, sollte die Diskussion eher dahin gehen, wie man etwa innerhalb eines Kontrollgruppendesigns oder eines partizipativen Ansatzes die Agilität der Evaluation steigern kann.

Schließlich analysiert Morell die 18 ihm bereitgestellten Evaluationsfälle, von ihm selbst durchgeführte sowie erdachte Evaluationen anhand der in den vorherigen Kapiteln entwickelten Theorien und Methoden. Er kategorisiert die in den Fällen enthaltenen unerwarteten Ereignisse, beschreibt, wie mit ihnen umgegangen wurde und zeigt auf, welche Alternativen möglich gewesen wären. Dabei folgert er von in den Fällen enthaltenen Schwierigkeiten auf Tendenzen für die Häufigkeiten der Probleme. Er

schreibt z.B., dass das Verhältnis der Fälle, bei denen es zu Schwierigkeiten bei der Durchführung eines Evaluationsdesigns kam, zu denen, bei denen es Probleme bei der Datenverfügbarkeit gab, 13:7 ist oder dass man bei fast 70% der Fälle die Veränderungen hätte voraussehen können. Aus den 18 Evaluationsfällen Typisierungen herzuleiten, ist sicherlich sinnvoll. Aus 18 ohne systematische Datensammlung von Kolleg(innen) erhaltenen Fällen kann aber nicht bereits auf Tendenzen für Häufigkeiten von verschiedenen Typen von Veränderungen und unerwarteten Ereignissen bei Evaluationen verallgemeinert werden. Morell erkennt dies und verwendet etwa die Prämisse „if the pattern were to hold for a larger sample“ (S. 157) für seine Generalisierungen. Es bleibt die Frage, warum diese Verallgemeinerungen, die noch keine Aussagekraft besitzen, dann überhaupt vorgenommen werden.

Die Evaluationsforschung zum Umgang mit Unsicherheiten, unerwarteten Ereignissen und Veränderungen, die eine Evaluation vor Schwierigkeiten stellen, steht noch am Anfang. Es ist Jonathan A. Morells Verdienst, dieses wichtige Thema angestoßen zu haben.



WAXMANN
Münster • New York • München • Berlin



www.waxmann.com
info@waxmann.com

Unter dem Label „Neue Steuerung“ vollzieht sich gegenwärtig in allen Bereichen des Bildungssystems ein grundlegender Wandel der Handlungskoordination in und zwischen Politik, Verwaltung und pädagogischer Praxis. Der sich darin manifestierende Veränderungsdruck begründet sich nicht zuletzt in der wieder in den Fokus der Aufmerksamkeit gerückten Reproduktion sozialer Ungleichheiten. Daraus ergibt sich die diesen Band charakterisierende Perspektivierung des Gegenstandes „Neue Steuerung“: Gefragt wird nach dem Verhältnis von Neuer Steuerung und tradierten sozialen Ungleichheiten im Bildungssystem.



Fabian Dietrich,
Martin Heinrich,
Nina Thieme (Hrsg.)

Neue Steuerung – alte Ungleichheiten?

Steuerung und Entwicklung
im Bildungssystem

2011, 360 Seiten, br., 34,90 €
ISBN 978-3-8309-2569-9