

Rezension zu:

Hense, Jan Ulrich/Böttcher, Wolfgang/Kalman, Michael/Meyer, Wolfgang (Hg.): Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis? Münster: Waxmann, 2019. 368 Seiten, 39,90 EUR, ISBN 978-3-8309-4074-6

*Niklas Zierke*¹

Überblick

Mit dem im Jahr 2019 unter dem Titel „Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis?“ im Waxmann Verlag erschienenen Sammelband legen die vier Herausgeber Jan Ulrich Hense, Wolfgang Böttcher, Michael Kalman und Wolfgang Meyer einen umfassenden und bislang in der Evaluationsforschung vermissten Beitrag zur Bedeutung von Evaluationsstandards in der gegenwärtigen Evaluationspraxis vor. Das knapp 300-seitige Werk behandelt die zentrale Frage, ob trotz heterogener Handlungsfelder in denen Evaluation praktiziert wird, etwaig von einheitlichen Qualitätsstandards gesprochen werden kann. Insbesondere möchten die Herausgeber – so ist dem Vorwort zu entnehmen – die Rolle der 2001 ins Leben gerufenen und 2016 revidierten „Standards für Evaluation“ der deutschen Gesellschaft für Evaluation (DeGEval) im heutigen nationalen und internationalen Evaluationsgeschehen beleuchten und vergleichend auf den Prüfstand stellen. Mit dem Werk richten sie sich an ein gemischtes, zwischen Evaluationspraxis und -forschung zu verortendes Zielpublikum.

Aufbau

Der Sammelband besteht aus insgesamt 19 sich an das Vorwort der Herausgeber anschließenden Artikeln angesehener Evaluationsexpertinnen und -experten. Nach dem gelungenen Vorwort, in welchem sektorenübergreifender Umsetzungsstand und Zukunftspotenzial von Evaluationsstandards als zentrale Problemstellung des Bandes vorgestellt werden und einem nachfolgenden

Einleitungskapitel zur historischen Genese und zum künftigen Potenzial der Standards der Evaluation, gliedern sich alle weiteren Beiträge des Bandes in vier Hauptkapitel auf. Nachdem sich im ersten Hauptkapitel vier Beiträge mit den DeGEval-Standards aus einem querschnittsbezogenen Blickwinkel beschäftigen, konzentrieren sich im daran anschließenden Kapitel insgesamt neun Beiträge auf die zentrale Frage des Bandes: Die Institutionalisierung und Bedeutung von Standards der Evaluation im Lichte der Praktiken unterschiedlicher Handlungsfelder. Die hier behandelten Praxisfelder reichen vom beruflichen, schulischen und hochschulbezogenen Bildungssystem, über die Gesundheitsförderung, die Entwicklungszusammenarbeit und das Wirtschaftssystem bis hin zum Wissenschaftssystem, der öffentlichen Verwaltung und der Struktur- sowie Regionalentwicklungspolitik. In den drei nachfolgenden Artikeln wird eine internationale Perspektivierung angeführt, die die Standards im Verhältnis zu globalen Entwicklungslinien zu interpretieren und einzuordnen sucht. Im synthetisierenden Schlusskapitel wird mit zwei Beiträgen eine übergreifende Bilanz zu dem aktuellen Stand der Verbreitung der Standards gezogen. Der Band enthält darüber hinaus eine detaillierte Auflistung und Beschreibung der aktualisierten DeGEval-Standards entlang ihrer vier zentralen Kategorien: Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit.

Inhalt

In dem ersten Teil des Buches werden die DeGEval-Standards in *umfassenden Perspektiven* – das heißt jeweils ausgehend vom Standpunkt

1 Centrum für Evaluation (CEval), Saarbrücken

einzelner Querschnittsbereiche – thematisiert. Die vier Artikel dieses Kapitels beschäftigen sich jeweils mit einem ausgewählten Querschnittsthema, das für die Evaluationspraxis besondere Relevanz besitzt. Wolfgang Meyer und Theodor Leiber gehen beispielsweise in ihrem Beitrag auf die Rolle der Ethik in der Evaluation ein und untersuchen dabei den ethischen Gehalt der vier zentralen Merkmale guter Evaluation. Sie können einerseits plausibel zeigen, dass die DeGEval-Standards zentralen ethischen Wertbezügen verpflichtet sind und mit diesen grundlegend vereinbar sind, auch wenn diese nicht explizit in Form eines axiomatisch zu verstehenden ethischen Entscheidungskriteriums gefasst sind; andererseits blieben die Standards aufgrund ihrer Abstraktheit immer Gegenstand von offenen Interpretationsprozessen. Der ethische Aspekt von Evaluation wird durch einen Beitrag der Autorin Angela Wroblewski inhaltlich weiter fundiert und ertragreich um die Genderperspektive ergänzt. Die zentrale Aussage ihres Artikels ist, dass die Betrachtung der DeGEval-Standards aus einer geschlechtsbezogenen Blickrichtung zur Sensibilisierung beitragen könne und Gendermainstreaming ein zentrales Anliegen der Weiterentwicklung von Evaluationsstandards sein muss. Wünschenswert wäre, in Zukunft zusätzlich auch einen Beitrag aufzunehmen, der sich explizit auf intersektionale Muster sozialer Disparität im Kontext von Evaluation und ihren Standards bezieht. So ist zum Beispiel – gerade in Hinblick auf Evaluation, die sich im Entwicklungskontext ansiedelt – die ethische Frage der Interkulturalität anhand der bestehenden Standards in gleicher Weise wie die Genderfragen zu reflektieren. Eine feststellbare Lücke besteht im Übrigen im Querschnittsthema der *digitalisierten Bewertung*, die ebenso einen lohnenden Blick auf Evaluationsstandards zuließe (vgl. Kropf/Laser 2019). So könnten beispielsweise neue digitale Evaluationstechnologien hinsichtlich ihrer Kompatibilität zu den Standards untersucht werden. Dies ist keine grundlegende Kritik, sondern ein inhaltliches Potenzial, das man inspiriert durch die Beiträge weiter ausschöpfen kann.

Ein weiterer Beitrag schildert aus der Perspektive der Methoden der empirischen Sozialforschung die methodologischen Kernpunkte der Evaluationsstandards. Caspari und Kollegen kommen dabei zu dem Schluss, dass die DeGEval-Standards im Kern durch eine besondere Verpflichtung zu einem wissenschaftlichen Ethos gekennzeichnet seien: Entsprechend könne nur wer angemessen methodisch professionalisiert ist, glaubwürdig Evaluationen durch-

führen und einschätzen, wie groß die zeitlichen, politischen oder finanziellen Restriktionen seitens des Auftraggebers sind, um sodann eine Praxis durchzuführen, die methodischen Gütekriterien genügt. Eine Schlussfolgerung, die überzeugend aufzeigt, dass qualitativ hochwertige Evaluation durch akademische Professionalisierung und Sozialisierung erreicht wird, die ihren zentralen Ausgangspunkt in der adäquaten Beherrschung qualitativer und quantitativer Methoden der empirischen Sozialforschung nimmt. Dem Schlagwort der Professionalisierung widmet sich auch der Beitrag von Schott, Niestroj und Friedrich, die auf die Wichtigkeit von Standards für die Professionalität der Evaluationspraxis verweisen. Sie unterstreichen zudem stichhaltig, dass hierfür auch Forschung über Evaluation wichtig sei, da man auf diese Weise den berufspraktischen Kern und professionsprägende Qualitätsverständnisse herausstellen könne, was beides wiederum die Konsolidierung zu einer distinkten professionellen Fachgemeinschaft befördere. In jedem Fall können die Beiträge, die sich mit den Standards aus der Perspektive einzelner Querschnittsbereiche befassen, glaubhaft darlegen, dass die Standards grundsätzlich kompatibel zu den ausgewählten Querschnittsthemen sind.

Der Blick auf die Bedeutung der Evaluationsstandards in den verschiedenen Handlungsfeldern liefert ein heterogeneres Bild, denn hier wird offenkundig, dass die Standards je nach Feld zu unterschiedlichem Grade Anwendung finden und teils von handlungsfeldspezifischen und lokalen Standards überlagert werden, wie beispielsweise Philipp Pohlenz und Uwe Schmidt überzeugend am Beispiel des Hochschulsystems darlegen. Während die DeGEval-Standards im Feld der Wirtschaft aktuell nahezu keine praktische Bedeutung – wohl aber Zukunftspotenzial – besitzen (siehe Beitrag von Kraus und Rost), sind sie in anderen Feldern wie der deutschen Entwicklungszusammenarbeit von anhaltend großer Relevanz und gewähren auch eine Anpassungsfähigkeit gegenüber neueren Entwicklungen dieses Praxisfelds (siehe Beitrag von Meyer et al.). So wird im letztgenannten Artikel mit starken Argumenten dargelegt, dass trotz gegenläufiger globaler Entwicklungen, eine Übernahme der rein politisch mandatierten Evaluation in das deutsche Evaluationssystem unwahrscheinlich sei. Gleichzeitig gestalten sich die Standards in diesem Handlungsfeld flexibel genug, um auch innovative Impulse aus dem globalen Diskurs und im Zuge der Definition ‚großer gesellschaftlicher Herausforderungen‘ aufzunehmen. So könnten beispielsweise

wichtiger werdende Nachhaltigkeitsziele durchaus adäquat in den Standards berücksichtigt werden.

Diese prinzipielle *Anpassungsfähigkeit* der Standards wird dabei in den meisten Praxisfeldern, in denen sie bereits angewandt werden, gesehen. In Feldern, in denen die Beitragenden noch deutliche Lücken in der Anwendung der DeGEval-Standards identifizieren, wird gleichzeitig stets auf Entwicklungspotenziale, die beispielsweise in einem verstärkten Dialog mit der DeGEval oder in systematischerer Implementierung der Standards liegen kann, verwiesen (siehe z.B. Müller und Renz zum Schulsystem oder Beutner zum Berufsbildungssystem). Auch im Bereich der Regionalentwicklungspolitik werden solche Entwicklungspotenziale gesehen, die primär davon abhängen, wie innerhalb des Faches der Diskurs um evidenzbasierte Planung mittelfristig ausfiele (siehe Beitrag von Weith und Kollegen). Diese Diagnose gilt sicherlich auch für die anderen untersuchten Handlungsfelder: Die zukünftige Anwendung und Verbreitung der DeGEval-Standards sind inhärent damit verbunden, auf welche fachkulturelle Anerkennung diese jeweils stoßen. Bis dahin werden sie stets auch mit lokalen oder verbandspezifischen Logiken konkurrieren – dies können die Beiträge zusammengekommen überzeugend darlegen.

Highlights

Als ein inhaltlicher Höhepunkt des Werks mag wohl eindeutig die aus dem Vergleich der verschiedenen Praxisfelder hervorgehende synthetisierende Synopse der vier Herausgeber erscheinen. Hier wird ersichtlich, dass Standards zur Evaluationsqualität nicht in jedem Praxisfeld gleichermaßen Relevanz besitzen. Einen solchen gelungenen systematischen Vergleich hat man in der Evaluationsforschung bisher noch zu selten gesehen. Der Sammelband schafft es darüber hinaus sogar eine internationale Forschungsperspektivierung einzuführen, was besonders positiv hervorsteht. In Zukunft ließe sich genau hier mit weiterer, dem ‚curiositas‘-Motiv verpflichteter Forschung anknüpfen.

Dem interessierten Leser respektive der interessierten Leserin fallen einzelne Beiträge wohl in ihrer inhaltlichen Güte oder in ihrer Novität besonders ins Auge. Ausdrücklich hervorzuheben ist hier, dass die querschnittsbezogene Sichtweise eine sehr gelungene und in der bisherigen Evaluationsforschung selten anzutreffende ist. Insbesondere der Aspekt der Ethik ist vielversprechend und könnte künftig eventuell noch kontrastiver im Verhältnis zu anderen berufsethischem Standards betrachtet werden.

Kritische Würdigung

Trotz der relativ großen inhaltlichen Güte der Beiträge im Einzelnen sowie ihrer bündigen Synthese im Allgemeinen sind zwei Kritikpunkte anzuführen, die gewiss aber als ‚Kritik auf hohem Niveau‘ zu verstehen sind. So könnten zunächst einzelne Beiträge insgesamt noch stärker von einem kritischen Reflexionswissen profitieren, das dazu beitragen kann, geltende Praktiken und Standards nicht nur zu beschreiben, sondern auch zu überdenken und dadurch zu optimieren. Eine solche Reflexivität der eigenen Praxis gehört dabei vielleicht sogar gerade in jenen berufsethischen Kanon der Evaluation (vgl. Zembylas 2017: 29). Außerdem könnten die Beiträge noch stärker von einer theoretischen Perspektivierung profitieren, welche versuchen sollte, darzulegen, warum es in manchen Fällen zu einer *sektoralen Homogenisierung* von Standards kommt und in anderen eine solche ausbleibt. Dies hängt eventuell mit ihrer Nähe zu bestimmten teilsystemspezifischen Handlungslogiken zusammen. So ist beispielsweise in der Wissenschaft eine vollkommene Diffusion relativ zu anderen Praxisfeldern gesehen unwahrscheinlicher, da wir es mit einem System zu tun haben, welches in langer Tradition etablierter Formen der Selbststeuerung steht; wie es beispielsweise anhand des Peer-Review-Verfahrens veranschaulicht werden kann (vgl. Reinhart 2012). Diese feldspezifischen theoretischen Besonderheiten wünschte man sich manchmal in den Beiträgen stärker berücksichtigt. Darüber hinaus ist auf stilistischer Ebene einzelner Beiträge zu bemängeln, dass der Lesefluss manchmal durch eine zu stichpunktartige Darstellungsweise gestört wird; eine Form, die zudem inhaltlich häufig zu sehr daran orientiert ist, die Kriterien sukzessive abzuhandeln, anstatt auch im gleichen Maße systematisch typologisierende Quintessenzen der Analyse darzulegen. Hier würden dem Zielpublikum sicherlich auch grafische und tabellarische Darstellungen, die zentrale Ergebnisse synthetisierend und kategorisierend beinhalten, entgegenkommen.

Relevanz

Das Werk hält eine weitreichende wissenschaftliche Anschlussfähigkeit an aktuelle sozialwissenschaftliche Forschungslinien zu dem Gegenstand der Evaluation bereit. So lässt sich sowohl an eine hypothesenprüfende, international vergleichende Sozialforschung über Evaluation denken, der das Werk als Ausgangspunkt dienen kann; als auch an eine breitere, theoretische Einbettung der Fragestellung, die versucht, Antworten zu finden auf Fragen nach den theoretischen Ursachen der in den Beiträgen immer wieder unterstrichenen

feldspezifischen Muster. Die von Hense et al. (S. 318) gut begründet abgeleiteten Bedingungsfaktoren für die Diffusion der Standards können als Grundlage hierzu dienen.

Auch für Praktiker(innen) und Anwender(innen) besitzt der Sammelband vielversprechende Anknüpfungspunkte. Das wird beispielsweise dann evident, wenn die Beitragenden explizit den Praxiswert von Standards unterstreichen. Beispielsweise verweisen Grohs und Piesker schlüssig auf Synergiepotenziale der DeGEval-Standards für die öffentliche Verwaltung. Oder es wird auch dann offenkundig, wenn man den Blick auf die Rolle von Standards im Feld der Hochschul- und Wissenschaftsbewertung wirft. Hier könnte zukünftig von einem stärkeren Dialog profitiert werden, der immer auch Fragen der sinnvollen feldspezifischen Übersetzbarkeit der Standards adressieren sollte.

Während für die letztgenannten zwei Lesergruppen eine grundsätzliche Leseempfehlung ausgesprochen werden kann, eignet sich der Sammelband hingegen weniger für Einsteiger(innen) in das Themengebiet der Evaluation. Das liegt beispielsweise daran, dass viele Beiträge ohne umfangreiche Literatur-

verzeichnisse auskommen, was eine grundlegende und weiterführende Lektüre erheblich einschränkt. Eine stilistisch vielleicht nicht immer anregende, aber inhaltlich wertvolle und im Gesamtbild konsistente Lektüre bietet der Band davon abgesehen in jedem Fall für Zielgruppen der professionellen Evaluationspraxis und auch für jene Akteure, die in den behandelten Handlungsfeldern aktiv sind.

Literatur

- Kropf, Jonathan/Laser, Stefan (2019): Eine Bewertungssoziologie des Digitalen. In: Kropf, Jonathan/Laser, Stefan (Hg.): Digitale Bewertungspraktiken. Soziologie des Wertens und Bewertens. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-21165-3_1
- Reinhart, Martin (2012): Soziologie und Epistemologie des Peer Review. Baden-Baden: Nomos. <https://doi.org/10.5771/9783845239415>
- Zembylas, Tasos (2017): Evaluation im kulturpolitischen Wirkungsbereich. In: Zeitschrift für Kulturmanagement, 3 (1), S. 13-36. <https://doi.org/10.14361/zkmm-2017-0102>