

Editorial

Routinen und Gewohnheiten sind in der Evaluation wie im restlichen Leben etwas sehr Nützliches. Sie helfen uns in wiederkehrenden Situationen schnell die erforderlichen Handlungen auszuwählen und diese in vielfach eingeübter Weise sicher und effizient auszuführen. Wollten wir jede sich stellende Problemsituation erst einmal gründlich auf mögliche Handlungsalternativen hin analysieren, um die beste Option zu ermitteln, würde dies schnell zum weitgehenden Stillstand führen. Aufgrund dieser Effizienz und da sich mit jeder Wiederholung die Handlungssicherheit verfestigt, sind Routinen aus unserem Alltag nicht wegzudenken. In einem auch populärwissenschaftlich bekannt gewordenen Bestseller hat der Psychologe Daniel Kahneman basierend auf seinen Arbeiten mit Amos Tversky zwischen ‚schnellem‘ und ‚langsamem‘ Denken unterschieden. Während Problemsituationen beim langsamen Denken bewusst analysiert werden, entspricht das automatische, wenig bewusste schnelle Denken dem routinemäßigen Handeln. Kahnemann zeigt eindrücklich die verschiedenen Gefahren auf, die schnellem Denken innewohnen, etwa wie Heuristiken dazu führen, dass komplexe Fragen durch einfachere, leichter zu lösende ersetzt werden.

Mindestens als Analogie lassen sich diese Überlegungen gut auch auf die Evaluationspraxis übertragen. Auch hier sind Routinen unverzichtbar für eine effiziente Aufgabenbewältigung. Das Schreiben eines Evaluationskonzepts, der Entwurf eines methodisch sauberen Fragebogens, die genaue Auswertung qualitativer oder quantitativer Daten – all das wäre ohne Routinen kaum in ökonomischer Weise unter den Bedingungen der Praxis in der Evaluation umzusetzen. In Analogie zum schnellen und zum langsamen Denken sollte sich aber immer die Frage stellen, ob diese Routinen auch in der Evaluation dazu führen können, dass wir vorschnell auf Gewohntes und Vertrautes zurückgreifen, obwohl die Situation eigentlich eine spezifischere, alternative oder ergänzende Problemlösung verlangen würde.

Ganz in diesem Sinne hält *Kalle Hauss* im ersten Beitrag dieses Hefts der Evaluation den Spiegel vor und hinterfragt eine oft wenig bewusst ausgeübte Routine, die Anwendung inferenzstatistischer Methoden auf nicht probabilistische Stichproben und Vollerhebungen. Er weist völlig berechtigt darauf hin, dass gerade ein empiriegestütztes Verfahren wie die Evaluation immer eine Selbstreflexion ihrer Methoden braucht. Im Kontext der Professionalisierung der Evaluation verweist dies auf die Notwendigkeit der empirischen Forschung über Evaluation und ihre Praxis, für die der Beitrag ein lesenswertes Beispiel darstellt. Denn der Autor argumentiert nicht nur aus einer methodologischen Perspektive, sondern gründet seine Kritik auf eine empirische Analyse der derzeitigen Praxis in publizierten Evaluationsstudien. Seine Inhaltsanalyse von 355 Journalbeiträgen und Praxisberichten belegt, dass inferenzstatistische Methoden häufig auch in Situationen zum Einsatz kommen, in

denen ihre Voraussetzungen nicht gegeben oder zumindest fraglich sind, so dass die Interpretation ihrer Ergebnisse ungenau oder inhaltlich uneindeutig sein kann. Konkret geht es um zwei Situationen: Einerseits quasi- oder nicht experimentelle Studiendesigns, in denen die Annahme einer zufallsbasierten Stichprobe nicht vorausgesetzt werden kann, andererseits Designs mit Vollerhebungen, in denen der Zufall gar keine Rolle bei Bildung der Stichprobe spielt und zusätzlich unklar ist, was die Grundgesamtheit ist, auf die per inferenzstatistischem Test geschlossen werden soll. Lesenswert macht den Artikel vor allem, dass diese Praxis nicht wie in manch ähnlich gelagerten methodologisch argumentierenden Beiträgen aus einer ‚puristischen‘ Perspektive nur als grundsätzlich verwerflich kritisiert wird. Vielmehr zeigt der Autor unter Bezug auf die Spezifika der Evaluationspraxis und ihrer Erfordernisse gut begründete aber anwendungsnahe Möglichkeiten zum Umgang mit inferenzstatistischen Methoden bei nicht zufallsbasierten Stichproben auf und macht uns damit ein Angebot zum ‚langsamen‘ Denken im Sinne Kahnemanns. Eine weitere in diesem Kontext zu hinterfragende Gewohnheit wäre übrigens die routinemäßige Verwendung jener Kenngröße, die vielleicht am häufigsten Gegenstand inferenzstatistischer Tests in Evaluationen ist, und zwar des Mittelwerts. Ob er als Maß der zentralen Tendenz, das alle Heterogenität der Befunde ausmittelt, automatisch das geeignetste Maß zur Bewertung von Maßnahmenwirkungen ist, wäre ebenfalls ein lohnender Gegenstand der Selbstreflexion in Evaluationen.

Auch die weiteren Beiträge dieses Hefts laden in verschiedener Weise zur Reflexion etablierter Gewohnheiten und Routinen der Evaluationspraxis ein. *Benjamin Froncek*, *Agostino Mazziotta*, *Pamela McCann*, *Ilka Würpel* und *Anette Rohmann* tun dies mit einer weiteren Arbeit, die der empirischen Forschung über Evaluation zugeordnet werden kann. Sie beleuchten einen bisher eher selten untersuchten Gegenstand und zwar verschiedene Ausprägungen institutionalisierter Evaluationspraxis. Denn Evaluation wird ja nicht nur im Rahmen ‚klassischer‘ Auftrags- oder Dienstleistungskonstellationen durch eine extern beauftragte Fremdevaluation durchgeführt, sondern zunehmend auch organisationsintern in Form von Selbstevaluationen oder internen Fremdevaluationen durch entsprechend spezialisierte Unterheiten. Die Autorinnen und Autoren des Beitrags untersuchen konkret die Praxis im Handlungsfeld der psychosozialen Beratungsarbeit. Hier berichten sie eine – wie in vielen anderen Bereichen – in den letzten Jahren gewachsene Bedeutung der Evaluation, die insbesondere mit der Notwendigkeit zur Qualitätsentwicklung und dem Nachweis fachlicher Standards begründet wird. Zu welchen Zwecken, mit welchen thematischen Schwerpunkten und mit welchen Rollenverteilungen und Methoden dabei zu welchen Zeitpunkten evaluiert wird, untersucht der Beitrag anhand einer Online-Befragung von 339 in der psychosozialen Beratungsarbeit tätigen Personen, im Rahmen deren Arbeit Evaluationsaktivitäten stattfinden. Interessant und zur Reflexion einladend ist insbesondere der Vergleich zwischen verschiedenen organisationalen Kontexten der evaluierten Beratungsarbeit. Dabei zeigt sich zunächst wenig überraschend, dass Evaluation in der öffentlich geförderten institutionellen Beratung stärker der Dokumentation gegenüber Mittelgebenden und in der gewerblich ausgeübten Beratungsarbeit eigener Praxis stärker dem eigenen Lernen dienen soll. Damit gehen anscheinend auch unterschiedliche Rollenkonstellationen in der Evaluation

einher. Selbstevaluation durch die Beratenden selbst erweist sich im untersuchten Handlungsfeld zwar insgesamt als dominante Konstellation, sie ist allerdings in der eigenen Praxis häufiger, während im institutionellen Kontext unabhängige Dritte eine größere Rolle spielen. Auch Schwerpunktsetzungen der Evaluationsarbeit unterscheiden sich. So werden als Bewertungskriterien bei der institutionellen Beratung eher strukturelle Merkmale herangezogen, während in der eigenen Praxis stärker auf die eingesetzten Methoden und Veränderungen Wert gelegt wird. Methodisch spielen schriftliche Erhebungen im institutionellen Kontext eine stärkere Rolle, während qualitative Methoden wie Beobachtungen, Interviews und Dokumentationen in der eigenen Beratung häufiger sind. Auch an diese Befunde lässt sich die im Rahmen der vorliegenden Studie nicht adressierte Frage anschließen, inwiefern die untersuchte Praxis eher Routinen und Gewohnheiten folgt als einer dem ‚langsamem‘ Denken vergleichbaren Vorgehensweise. Denn zumindest für Evaluationszwecke wie das Weiterentwickeln der Praxis oder das Dazulernen über die eigene Beratungsarbeit sind sicherlich angepasste Evaluationsverfahren jenseits einer routinemäßigen prozessbegleitenden Datenerhebung und -auswertung erforderlich. Die Frage, ob institutionalisierte Evaluationssysteme so ausgestaltet sind, dass die intendierten Evaluationszwecke tatsächlich erreicht werden können, stellt sich sicherlich auch in vielen anderen Handlungsfeldern.

Der Beitrag von *Steffen Wild*, *Ernst Deuer* und *Philipp Pohlenz* fokussiert mit dem Hochschulbereich ein Handlungsfeld, in dem solche Evaluationssysteme ebenfalls eine wichtige Rolle spielen. Sie untersuchen mit dem Studienerfolg ein in letzter Zeit zunehmend thematisiertes Erfolgskriterium bei der Evaluation von Studium und Lehre. Ihre Arbeit kann zur Reflexion dienen, dass zunächst scheinbar einfach zu fassende Konstrukte durch unterschiedliche Stakeholder sehr unterschiedlich interpretiert und gewichtet werden können und ihr eindeutiges und homogenes Verständnis daher nicht immer vorausgesetzt werden sollte. Mit der Methode der latenten Profilanalyse zeigen sie auf Basis einer Online-Befragung von 359 hauptamtlich Lehrenden der Dualen Hochschule Baden-Württemberg, dass ein alltagssprachlich scheinbar einfacher Begriff wie ‚Studienerfolg‘ inhaltlich durchaus unterschiedlich verstanden wird. Während etwa die Hälfte der Befragten mit Studienerfolg besonders stark gute Noten und niedrige Abbruchquoten der Studierenden (‚Studierendenfokussierung‘) verbanden, betonte ca. ein Drittel stärker die im Studium erworbenen fachlichen Fähigkeiten (‚Fachfokussierung‘) und die verbleibende Restgruppe die Vorbereitung auf den Arbeitsmarkt (‚umfassende Qualifikationsfokussierung‘). Bei Differenzierung nach Professor(inn)en für Lehraufgaben und Studiengangsleitungen zeigten sich teils deutliche Unterschiede in diesen Wahrnehmungen: Unter den Professor(inn)en war die Fachfokussierung weit häufiger anzutreffen, bei den Studiengangsleitungen dagegen die Studierendenfokussierung. Stellt man sich nun vor, diese Gruppen würden im Rahmen einer Evaluation eines Projektes zur Verbesserung des Studienerfolgs befragt, zeigt sich, dass dies ohne weitere Elaboration des Begriffs risikobehaftet ist: Dass unter dem alltagssprachlich scheinbar eingängigen Begriff tatsächlich dasselbe verstanden wird, muss vor dem Hintergrund der Ergebnisse des Beitrags sicherlich hinterfragt werden.

Lena Zimmer, Till Seipp und Uwe Schmidt fokussieren in ihrem Beitrag einen weiteren Fall, der zumindest metaphorisch dem (zu) ‚schnellen‘ Denken in der Evaluation zugeordnet werden kann. Es geht um die verkürzte Annahme, dass Evaluation automatisch zur Rezeption ihrer Ergebnisse und diese Rezeption automatisch zu optimierten Handlungskonsequenzen führt. Obwohl die Forschung über Evaluation diese einfache Wirkungskette spätestens seit den 1970er Jahren als verkürzt und unterkomplex entlarvt hat, kommt sie noch oft genug zur Anwendung. Auch verschiedenen Evaluationsverfahren im Schulbereich, seien es regelmäßige Leistungstests und deren Rückmeldung oder die hier untersuchten externen Schulevaluationen, liegt grundsätzlich diese Logik zugrunde. Der Beitrag untersucht auf Basis von umfassenden Befragungsdaten von Schulleitungen und Lehrkräften in mehrerebenenanalytischer Betrachtung Einflussfaktoren auf die Nutzung von externen Evaluationsergebnissen aus inspektionsähnlichen Verfahren. Die Befunde bestätigen erneut, dass nur eine eher oberflächliche Auseinandersetzung mit verschiedenen Evaluationsergebnissen berichtet wird und eine Nutzung zur Weiterentwicklung der eigenen Arbeit in noch geringerem Maße ausgeprägt ist. Für die hier fokussierte externe Schulevaluation fällt diese Diskrepanz sogar besonders deutlich aus. Dass diese Befunde aber nicht zur grundsätzlichen Kritik an Evaluationsverfahren dienen können, sondern eher auf die Notwendigkeit ihrer bewussten Gestaltung, Einbettung und Kommunikation verweisen, zeigen die weiteren Analysen des Beitrags. Denn das Ausmaß der Rezeption und Nutzung kann u.a. von der Schulleitung, ihrem Führungsstil und ihrer Haltung gegenüber der Evaluation beeinflusst werden.

Der letzte Originalbeitrag des Hefts stammt von *Lutz Thieme* und berichtet über eine haushaltstheoretische Forschungsstudie zu Effekten des Ehrenamtsmanagements auf ehrenamtliches Engagement. Ehrenamtsmanagement, das hier eher als indirekt erschlossenes Konstrukt denn als konkrete Intervention betrachtet wird, soll prinzipiell die Leistung von ehrenamtlich Engagierten erhöhen und zu einem besseren Matching von Personen, die sich engagieren wollen, und ehrenamtlichen Organisationen beitragen. Unter Rückgriff auf Daten des sozioökonomischen Panels wird die Frage adressiert, inwiefern veränderte Budgetrestriktionen die Entscheidung über Beginn, Ausweitung oder Reduktion eines ehrenamtlichen Engagements beeinflussen. Zu diskutieren wäre, inwiefern die vorgeschlagene Methodologie, die im Beitrag exemplarisch angewendet wurde, über das Untersuchungsfeld des ehrenamtlichen Engagements hinaus und in konkreten Evaluationskontexten zur Anwendung kommen kann.

Wie immer besteht das Heft nicht nur aus Originalbeiträgen. In der Rubrik Praxisberichte freue ich mich zunächst besonders über einen Beitrag eines geschätzten Gießener Kollegen. *Christian Diller* stellt in seinem Beitrag vor, wie im Rahmen des Masterstudienganges Wirtschaftsgeographie und Raumentwicklungspolitik an der Justus-Liebig-Universität Gießen Studierende gegenstandsbezogene Evaluationskompetenzen erwerben. Wie nötig dies ist, zeigt er anhand von Ergebnissen einer bundesweiten Befragung, die zeigen, dass Evaluation zwar zum späteren Anwendungsportfolio in diesem Handlungsfeld gehört, sie aber oft nicht im Studium vermittelt wird. Um dieses Problem zu adressieren, umfasst der vorgestellte Masterstudiengang mehrere Pflichtmodule mit Evaluationsschwerpunkt und ermöglicht in

weiteren Modulen eine Fokussierung auf Evaluationsinhalte. Insgesamt kann er als exemplarisches Beispiel dienen, wie es in einem fachspezifischen Studiengang gelingt, umfassendere Evaluationskompetenzen, die über rein methodische Kenntnisse zu spezifischen Untersuchungsdesigns hinausgehen, so zu vermitteln, dass sie im konkreten Handlungsfeld zum Einsatz kommen können.

Ebenso lesenswert ist ein weiterer Praxisbericht von *Reinhard Stockmann*, der ebenfalls den Aufbau von Evaluationskompetenzen zum Gegenstand hat, hier aber umfassender unter der Perspektive des Capacity Development. Konkret steht der Evaluation-Capacity-Development (ECD)-Ansatz des Centrums für Evaluation (CE-val) im Mittelpunkt. Der Beitrag geht von der Beobachtung aus, dass zwar weltweit eine steigende Nutzung der Evaluationsfunktion sowie eine Professionalisierung der Evaluation zu beobachten ist, für einen umfassenden Nutzen für Politik und Gesellschaft aber eine Systemperspektive notwendig ist. Das vorgestellte ECD-Angebot umfasst daher Maßnahmen auf verschiedenen Systemebenen mit jeweils zielgruppenspezifischen Programmen, die von der Sensibilisierung des Top-Managements über Angebote für die operative Ebene bis hin zu train-the-trainer-Elementen reichen. Informativ für alle, die an der Ausbildung von Evaluationskompetenzen interessiert sind, sind hier insbesondere die jeweils berichteten Erfahrungen.

Abgerundet wird das Heft durch drei Rezensionen und die Rubrik DeGEval ... Info. Wolfgang Beywl bespricht das neueste Werk von *Michael Q. Patton* ‚*Principles-Focused Evaluation: The Guide*‘, das erneut hohen Anregungsgehalt bezüglich einer Reflexion etablierter Rollenverständnisse in der Evaluation bietet. Ulrich Degen stellt ‚*Chinas Berufsbildung im Wandel. 30 Jahre Entwicklungszusammenarbeit mit der Hanns-Seidel-Stiftung*‘ von *Reinhard Stockmann* und *Wolfgang Meyer* vor, das eine Evaluation der Berufsbildungszusammenarbeit der Hanns-Seidel-Stiftung mit China untersucht und dabei Vergleiche zu einer vor 17 Jahren durchgeführten, analogen Studie ziehen kann. Reinhard Stockmann selbst schließlich rezensiert den von *Fritz Sager*, *Thomas Widmer* und *Andreas Balthasar* herausgegebenen Band ‚*Evaluation im politischen System der Schweiz*‘, das eine intensive Länderstudie bietet und damit zu internationalen Vergleichen anregt. In der Rubrik DeGEval ...Info schließlich lesen Sie einen Bericht von *Maximilian Schmid*, *Amy O'Brien*, *Thorsten Bär* und *Susanne von Jan* über die Frühjahrstagung des DeGEval-Arbeitskreises Entwicklungspolitik und Humanitäre Hilfe, die das OECD/DAC-Kriterium der Relevanz in den Mittelpunkt stellte.

Viel Vergnügen beim Lesen und hoffentlich viele Anregungen zum ‚langsamen‘ Denken wünscht Ihnen

Jan Hense

Für die Herausgeberinnen und Herausgeber