DEGEVAL … INFO

Zeitschrift für Evaluation | 24. Jahrgang | 2025
Heft 2 | S. 351–359 | https://doi.org/10.31244/zfe.2025.02.13 |
© 2025 Waxmann

„Best credible und nützlich" – methodische Qualität, Nützlichkeit und Nutzung von Evaluationen

Bericht zur Frühjahrstagung 2025 des AK Methoden in der Evaluation

Franziska Heinze[1], Alexander Kocks[2], Klaus Heinle[3], Nils Verwohlt[3]

1. Anlass

Evaluationsergebnisse und Empfehlungen sollen von den beteiligten Stakeholder:innen und Adressat:innen nicht nur als nützlich wahrgenommen, sondern auch faktisch genutzt werden. Evaluierende stehen damit vor der Aufgabe, angemessene und qualitativ hochwertige Evaluationen umzusetzen und zugleich weitere Nützlichkeitsanforderungen zu berücksichtigen, um ‚best credible evidence' für eine bestmögliche Nutzung von Evaluation zu produzieren.

Diesem Thema widmete sich der AK Methoden in der Evaluation in der diesjährigen Frühjahrstagung am 10. und 11. Juli 2025. Mit der Stabsstelle Evaluierung der Hanns-Seidel-Stiftung e. V. in München stand dem Arbeitskreis dabei eine gastgebende Institution zur Seite, die nicht nur selbst Evaluationen durchführt und beauftragt, sondern zugleich an der Schnittstelle zu Auftraggeber:innen und Adressat:innen von Evaluation, insbesondere in der Politik, tätig ist.

Ziel der Tagung war es, auszuloten, in welchem Verhältnis die methodische Qualität einer Evaluation und deren wahrgenommene Nützlichkeit bzw. Nutzung stehen. Dies sollte entlang konkreter Beispiele, methodischer Designs sowie damit einhergehender Fragen der Responsivität und Einbindung von Stakeholder:innen geschehen.

1 Brandenburgische Technische Universität (BTU) Cottbus-Senftenberg; Sprecherin des AK Methoden
2 Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval); Sprecher des AK Methoden
3 Stabsstelle Evaluierung der Hanns-Seidel-Stiftung e. V.

2. Impulse

Den Auftakt machte *Klaus Heinle*, Leiter der Stabsstelle Evaluierung der Hanns-Seidel-Stiftung (HSS) in München. In seiner Begrüßung und Einführung stellte er kurz die Aufgaben der Stabsstelle dar, an der Schnittstelle zwischen operativer Arbeit und Geschäftsleitung. Ziel der Stabsstelle ist es, die Maßnahmen und Projekte der Stiftung zu verbessern und ihr volles Potential auszuschöpfen. Daher arbeitet die Stabsstelle an der Qualität von Evaluationen und ihrer Nutzung in Dialogpartnerschaft mit den operativen Abteilungen der HSS und ist den Standards von Evaluation und Sozialwissenschaften verpflichtet. Demnach braucht die Nutzung von Evaluation nicht nur gute Evaluationen, sondern auch einen Rahmen, in dem diese angemessen reflektiert werden können. Innerhalb der Stiftung wurde in den letzten Jahren das Thema Wirkungsorientierung in Gänze in Organisationsprozesse implementiert, beispielsweise im Zuge der Neukonzeptionierung der Berichtspflicht an die Geschäftsleitung in der HSS. Auch die Abläufe der wirkungsorientierten Planung, des Monitorings und der Evaluation wurden neu gedacht. Nützlichkeit von Evaluation beginnt für die HSS nicht erst mit der Durchführung einer Evaluierung – sondern sie beginnt mit der Planung der Projekte, beim Ansetzen des Monitorings, bei der Weitergabe der Daten an unterschiedliche Zielgruppen und auch bei der Auftragsklärung zu einer externen Evaluierung, bei der Frage: Was wollen wir wissen, wofür und für wen und wozu brauche ich den Bericht überhaupt? Die Stabsstelle arbeitet zudem offensiv an der Kommunikation über Ergebnisse sowie zusammen mit der Zentralstelle „Kommunikation und Öffentlichkeitsarbeit" an deren Verbreitung.

Erkenntnisse aus Theorie und Praxis

In der Keynote *„Nützlichkeit von Evaluation"* vermittelte *Susanne von Jan*, smep-consult, Salzburg, und Mitglied des Vorstands der DeGEval, grundlegende *„Erkenntnisse aus Theorie und Praxis"* zum Tagungsthema. Nach begrifflichen Klärungen stellte sie dar, dass die Nutzung von Evaluation erstens davon beeinflusst wird, wie jene durchgeführt wird. Dabei sind die Qualität der Kommunikation in einer Evaluation und die Kompetenz der Evaluierenden, die Qualität der Evaluation selbst sowie ihre Glaubwürdigkeit, ihre Ergebnisse, die Rechtzeitigkeit und die Relevanz der Ergebnisse wichtige Einflussfaktoren. Zweitens beeinflussen die Entscheidungskontexte, in die eine Evaluation eingebettet ist, deren Nutzung. Hierzu zählen persönliche Eigenschaften und Rollen der Nutzer:innen, das politische Klima und die Einstellung der Nutzer:innen zu Evaluation, ihre Informationsbedarfe sowie die jeweiligen Merkmale der zu treffenden Entscheidung. Auch konkurrierende Informationen oder ein Umsetzungsmonitoring können sich darauf auswirken, wie Evaluationen genutzt werden. Drittens hat die Einbindung von Stakeholder:innen einen direkten positiven Effekt auf die Nutzung von Evaluation und wirkt zugleich moderierend auf andere Faktoren, die die Nutzung von Evaluation beeinflussen.

Von Jan verwies darauf, dass diese Aspekte auch in den Nützlichkeitsstandards der DeGEval (DeGEval, 2017) sowie in weiteren Qualitätsstandards von Evaluationen verankert sind. Anschließend thematisierte von Jan verschiedene Forschungserkenntnisse

dazu, wie die Qualität einer Evaluation mit deren Nutzung zusammenhängt. Beispielsweise ist nach Johnson et al. (2009) die Qualität weniger wichtig für die Nutzung als Empfehlungen und Kommunikation. In einer Befragung von Evaluator:innen kommen Fleischer und Christie (2009) zu dem Befund, dass jene die methodische Qualität einer Evaluation als weniger bedeutsam für die Nutzung einschätzen als andere Aspekte. Eine Meta-Evaluation von Bundi et al. (2021) konstatiert, dass die in der Meta-Evaluierung festgestellte Qualität einer Evaluierung nicht im Zusammenhang mit der wahrgenommenen Qualität einer Evaluation steht. Die Anwendung eines Methodenmixes korrelierte sogar negativ mit ihrem wahrgenommenen Nutzen (ebd.). Silvestrini (2023) arbeitet ebenfalls in einer Meta-Evaluation heraus, dass Evaluationen teilweise eine hohe Nützlichkeit bescheinigt wird, trotz teils deutlicher Schwächen, v. a. hinsichtlich methodischer Qualität und Ergebnisdarstellung. Was bedeutet dies? Drei Thesen bildeten den Abschluss der Keynote: 1. Eine hohe Qualität der Evaluierung ist die Grundvoraussetzung für ihre positive Wirkung. Dies ist ein professioneller Standard und zentral für die Glaubwürdigkeit einer Evaluierung. Von Jan verwies unter Bezugnahme auf die Meta-Evaluierung von Silvestrini (2023) darauf, dass gute Leistungsbeschreibungen und angemessene Budgets einen positiven Beitrag zur Qualität von Evaluierungen leisten können. 2. Eine hohe Qualität der Evaluierung wird nach Ansicht von Susanne von Jan oft nicht erkannt und häufig nicht bezahlt. Dies betreffe insbesondere die Aufbereitung und Auswertung von (qualitativen) Daten, aber auch Kommunikationsprozesse, insbesondere wenn Auftraggebende nur eingeschränkte Evaluierungskapazitäten hätten. Daher braucht es ein angemessenes Verständnis für diese Kosten, die die Basis für qualitativ hochwertige Arbeit bilden. 3. Eine Umsetzung von Evaluierungen nach wissenschaftlichen Standards geht – ganz im Sinne allgemeiner Trends in den Wissenschaften – möglicherweise auch mit einer wissenschaftlich komplizierten Sprache einher. Dies kann laut von Jan die Nützlichkeit und Nutzung von Evaluation einschränken. Eine verständliche Sprache und zielgruppenangemessene Kommunikationsformate sind für von Jan von besonderer Bedeutung. Diese sollten dennoch keine inhaltliche Substanz einbüßen und wissenschaftlich glaubwürdig sein.

Umkämpfte Glaubwürdigkeit und ungewollte Nutzungsweisen

Frank König vom Deutschen Jugendinstitut, Außenstelle Halle (Saale), griff die grundlegende Bedeutung einer hohen methodischen Qualität von Evaluationen in seinem Vortrag *„Cui bono? Weshalb die methodische Qualität von Evaluation bei der Nutzung ihrer Ergebnisse nicht alles, aber ohne sie alles nichts ist"* auf und diskutierte dies an dem Fallbeispiel der unerwünschten Nutzung von Evaluationsergebnissen aus dem Feld der Demokratieförderung und Extremismusprävention. Ausgangspunkt des Vortrags war die an die Evaluation des Bundesprogramms „Demokratie leben!" in der Förderphase 2020 bis 2024 herangetragene Aufgabe, die Gesamtanlage sowie die Erträge des Programms zu analysieren und zu bewerten. Die Erhebungen und deren Ergebnisse wurden in unterschiedlichen Publikationsformaten aufbereitet, in denen die Methodendarstellung und -reflexion einmal ausführlich und einmal eher knapp ausfiel. Das für das Programm und die Evaluation verantwortliche Ministerium nutzte ausgewählte Erkenntnisse der Evaluation zu Legitimationszwecken in Haushalts-

verhandlungen. Dies war u. a. eingebettet in eine politische Situation, in der eine intensive Auseinandersetzung um Ziele des Programms stattfand. Die in einer Pressemitteilung des Ministeriums verkürzt und zusammengefasst dargestellten Evaluationsbefunde wurden anschließend in verschiedenen Medien aufgegriffen. König stellte am Beispiel eines Onlinemagazins dar, wie die Medienberichterstattung selektiv, ohne Rückgriff auf die öffentlich zugänglichen und ausführlichen Evaluationsberichte erfolgte und entlang (politischer) Ziele das methodische Vorgehen sowie die Evaluation selbst zu delegitimieren versuchte. Das Beispiel verdeutlicht, dass verschiedene Akteur:innen die Evaluationsergebnisse aus ihren eigenen Logiken heraus nutzen. Gerade in politisch umkämpften Feldern, wie dem hier vorgestellten, ist dabei auch die jeweilige (politische) Positionierung der Akteur:innen für Nutzungsweisen entscheidend. Während Programminitiator:innen Erfolgsmeldungen verbreiten wollen, um eine Weiterführung des Programms zu legitimieren, nutzen Medien diese beliebig zu unterschiedlichen Zwecken. Wenn dabei Daten in der Öffentlichkeitsarbeit oder in der Beantwortung von Presseanfragen unzureichend kontextualisiert und diese Informationen auf Medienseite nicht sorgfältig geprüft werden, kann dies zu mitunter folgenreichen Nutzungsweisen führen. Als Lernerfahrungen leitete König ab, dass eine qualitativ hochwertige Methodenanwendung zwar die Evaluation absichert, aber keine Garantie für eine adäquate Nutzung darstellt. Evaluation ist daher gefordert, sich gerade auch in politisch umkämpften Feldern mit ungewollten Nutzungsweisen auseinanderzusetzen und diese im Vorfeld möglichst zu antizipieren.

Nützlichkeit stärken durch Methoden der Zukunftsforschung

Jan-Tobias Polak vom Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit, Bonn, thematisierte im Vortrag *„In die Zukunft schauen“*, inwiefern eine *„Stärkung der Nutzung von Evaluierungen durch Methoden der Zukunftsforschung“* möglich sei. Eingangs skizziert er das Spannungsfeld, dass Empfehlungen ein wichtiges Instrument für die Nutzung von Evaluierungsergebnissen und prospektiv ausgerichtet sind, jedoch vor allem auf retrospektiven Daten und Erkenntnissen basieren. Anschließend definierte Polak ‚Foresight‘ als „a structured and systematic approach of exploring plausible futures to anticipate and better prepare for change“ (OECD, o. J.). Damit sind verschiedene Methoden der Zukunftsforschung gemeint, die eine Erhebung und Analyse multipler und plausibler Zukünfte ermöglichen. Diese beziehen sich beispielsweise darauf, wie zukünftige Rahmenbedingungen erfasst und aufkommende Themen durch Prognosen auf Basis von bestehenden Daten identifiziert (z. B. Trend Analysis, Forecasting), wie alternative Entwicklungen systematisch erhoben (u. a. Scenario Analysis, Delphi), wie zukünftige Entwicklungen frühzeitig identifiziert (u. a. mittels Horizon Scanning, Weak Signals) oder kreative Spekulationen über Zukünfte möglich werden können (z. B. Design Fiction, Wild Cards). Polak plädierte dafür, durch eine *futures-informed evaluation* die strategische und systematische Auseinandersetzung mit (wünschenswerten) zukünftigen Entwicklungen in Evaluationen einzubeziehen. Die Integration von Foresight-Methoden kann unterschiedliche Funktionen und Aufgaben übernehmen, abhängig davon, ob sich der Evaluationsgegenstand gerade in der Design- oder Implementierungsphase befindet oder ob Foresight-Methoden am

Ende eines Programms zum Einsatz kommen und in welcher Phase einer Evaluierung sie einbezogen werden. Die grundlegenden Überlegungen konkretisierte Polak anschließend am Beispiel der DEval-Evaluierung des BMZ-Aktionsfelds „Bevölkerungsdynamik; sexuelle und reproduktive Gesundheit und Rechte". Dabei kombinierte die Evaluation retrospektive Methoden zur Bewertung des Gegenstands mit einer Szenarioanalyse für die systematische Erfassung zukünftiger Rahmenbedingungen des Gegenstands. Auf dieser Basis formulierte sie Empfehlungen, wobei die Ergebnisse auf Basis der retrospektiven Methoden mit den Ergebnissen der prospektiven Methode abgeglichen wurden (future proofing). In der Reflexion des Vorgehens merkte Polak an, im Design der prospektiven Methoden könne eine methodenintegrative Vorgehensweise gestärkt werden. Zugleich erwies sich in der Datenerhebung das für Zukunftsszenarien bedeutsame ‚Out-of-the-box'-Denken als Herausforderung, sowohl aufseiten der Befragten als auch aufseiten der Evaluierenden in der konkreten Interviewgestaltung. Mit Blick auf den methodischen und methodologischen Status der genutzten Erhebungsmethoden ergab sich zudem ein Spannungsverhältnis zwischen der Legitimität und der Qualität der (retro- und prospektiv) entwickelten Szenarien. Nicht zuletzt spiegelten die finalen Szenarien nur sehr eingeschränkt wider, wie aufwendig die methodische Umsetzung eines solchen Vorgehens ist.

Gegenstandsimmanente Anforderungen an Glaubwürdigkeit und Nützlichkeit

Den zweiten Tagungstag eröffneten *Anna Gieschen* und *Thomas Eckert* von der Ludwig-Maximillians-Universität München mit einem Input zu „*Evaluation im Spannungsfeld zwischen methodischen Ansprüchen, legitimatorischen Notwendigkeiten und pädagogischem Selbstverständnis*". Die Vortragenden erläuterten zunächst, dass Evaluationen in pädagogischen Kontexten an Bedeutung gewinnen und sich zugleich das Spektrum relevanter Stakeholder:innen ausdifferenziere. In der Konsequenz müssten Evaluationsergebnisse zahlreichen unterschiedlichen Adressat:innen vermittelt werden. Wenn jene nicht über das methodische Wissen verfügen, um Evaluationsergebnisse angemessen einschätzen zu können, führe dies Gieschen und Eckert zufolge zu drei Entscheidungsproblemen: Erstens stelle sich die Frage, wie vor diesem Hintergrund (scheinbar) widersprüchliche Befunde vermittelt werden können. Zweitens bedürfe es der Klärung, welche normative Referenz als Bezugspunkt für die Interpretation und Bewertung der Ergebnisse herangezogen wird. Drittens müsse Evaluation klären, wie sie Akzeptanz aufseiten der pädagogisch Tätigen herstellen könne. Diese Entscheidungsprobleme erläuterten die Vortragenden am Beispiel der Studie „Erhebung zum Lese- und Mathematikverständnis in den Jahrgangsstufen 4 bis 6 in Berlin" (ELEMENT). Die Studie kam zu dem Ergebnis, dass Schüler:innen nach der vierten Klassenstufe bis zum Ende der sechsten Klassenstufe in Bezug auf das Leseverständnis sowohl in sechsjährigen Grundschulen als auch in Frühstarterklassen an Gymnasien eine vergleichbare Leistungsentwicklung durchlaufen. Zugleich zeigte die Studie, dass Schüler:innen mit gleicher Lernausgangslage in Gymnasien eine höhere Leistungssteigerung aufwiesen. Die Ergebnisse der Berliner Studie wurden zu einem Zeitpunkt publiziert, als in Hamburg gerade über eine sechsjährige Primarstufe diskutiert wurde. Während für Eltern möglicherweise das zweite Ergebnis im Hinblick auf die indivi-

duelle Förderung, den Bildungserfolg und die Leistungsentwicklung von Kindern bedeutsamer ist, ist für die Bildungsadministration der erste Befund zur vergleichbaren Leistungsentwicklung in Frühstarterklassen und sechsjährigen Grundschulen von Bedeutung. Dies verweist bereits auf die normative Referenz als relevante Bewertungsgrundlage in der Einordnung der Befunde in pädagogischen Kontexten. Die Herstellung von übergreifender Akzeptanz und Glaubwürdigkeit betrifft dann vier Ebenen (Individuum, Institution, System und Interaktion). Auf der individuellen Ebene, so Gieschen und Eckert, beeinflussen das professionelle pädagogische Selbstverständnis, die Haltung gegenüber Evaluation sowie die Evaluationskompetenz und -erfahrung, inwieweit Evaluierungen Akzeptanz erfahren und als glaubwürdig anerkannt werden. Anhand einer eigenen Studie zeigten sie unterschiedliche Wahrnehmungen von Pädagog:innen zu den einzelnen Aspekten auf. Sie leiteten ab, dass beispielsweise die Anerkennung pädagogischer Expertise und die Verknüpfung der Evaluation mit pädagogischen Zielstellungen eine wesentliche Bedingung für Akzeptanz und Glaubwürdigkeit sei. Entsprechend ist eine dialogische Klärung des Evaluationszwecks und dessen pädagogische Relevanz bedeutsam, ebenso wie Transparenz und Beteiligung der Stakeholder:innen bei der Planung, Durchführung und Interpretation von Evaluation.

Nützlichkeit durch systematische User-Responsiveness

Schließlich gaben *Gerald Leppert* und *Folake Kareem* vom Global Center on Adaptation, Groningen, einen Einblick in ihre Herangehensweise an einen partizipativen Entscheidungsprozess zu Evaluierungsfragen und -methoden in Evaluierungen. Unter dem Titel „*From Rigour to Usefulness: User-Responsive Evaluation Design in Rapidly Changing Environments – Insights from the Global Center on Adaptation*“ stellten sie dar, wie die konzeptionelle Einbindung der Stakeholder:innen einer Evaluation dazu beitragen soll, deren spezielle Erkenntnisinteressen zu berücksichtigen und konkreten Nutzen zu generieren. Im Vordergrund des Beitrags stand hier der am DEval entwickelte systematische Ansatz zur Methodenintegration (Leppert et al., i. V.), und die dabei erfolgende Einbindung der an der Evaluation Beteiligten. Leppert und Kareem berichteten, wie dabei in einem mehrstufigen Verfahren die von der Evaluationseinheit festgelegten Evaluationsfragen mit den Stakeholder:innen besprochen und von diesen priorisiert werden. Dieser Prozess unterstütze Evaluator:innen, etwas über die Präferenzen der Stakeholder:innen zu erfahren, und helfe den Beteiligten, sich eigener Präferenzen bewusst zu werden. Die so zustande kommende Gewichtung der Teilkomponenten ermögliche es den Evaluator:innen, ihre Kapazitäten verstärkt den Interessen der Stakeholder:innen zukommen zu lassen. Dadurch könne sie konkrete Antworten geben auf die Fragen, die sich die Beteiligten wirklich stellen. Außerdem entwickele sich so eine intensivere Art der geteilten Ownership zwischen Evaluator:in und Evaluierten. Die führt nach Ansicht von Leppert und Kareem dazu, dass die Evaluation als Entscheidungshilfe für die Verbesserung der Praxis und nicht als losgelöste akademische Fingerübung wahrgenommen wird.

3. Diskussion und Fazit

Die Q&A-Runden zu jedem Vortrag betonten, dass Evaluationen grundlegend von hoher methodischer Güte sein sollten. Angesichts unzureichender oder unerwünschter Nutzung von Evaluation wurde kontrovers diskutiert, inwieweit Evaluator:innen zur Stärkung von Evaluationskompetenzen aufseiten von Auftraggeber:innen und weiteren Stakeholder:innen beitragen können bzw. sollen. Zum einen wurden ‚Lesehilfen‘ für das Verständnis komplexer Evaluationsmethoden in Berichten vorgeschlagen, zum anderen wurden kommunikative Kompetenzen aufseiten der Evaluierenden gefordert, um das Lernen in und durch Evaluation sowie das Verstehen des methodischen Vorgehens zu fördern. Neben dem geteilten Ansinnen, adressat:innengerechte Kommunikations- und Ergebnisformate bereitzustellen, mahnten Teilnehmende an, dass Evaluierende sich auf ihr Kerngeschäft, die Analyse und Bewertung, konzentrieren sollten. Gerade in Evaluationsaufträgen und -budgets sind Vermittlungsleistungen, die darauf abzielten, die Evaluationskapazitäten der Auftraggebenden und weiterer Stakeholder:innen zu stärken, grundsätzlich zwar wünschenswert, aber selten finanziell abgedeckt. Ein weiterer Diskussionsstrang fokussierte, wie Stakeholder:innen in Evaluationen eingebunden werden können, um die Nützlichkeit und Nutzung von Evaluationen zu erhöhen. Vorgehensweisen, die Ziele, Priorisierungen oder angestrebte Nutzungsweisen der zentralen Stakeholder:innen in den Fokus rückten und die (methodische) Ausgestaltung von Evaluation daran orientierten, erfuhren viel Zustimmung. Die Frage der ‚Nützlichkeit für wen?‘ müsse bereits in Leistungsbeschreibungen adressiert werden. Auch unterschiedliche Wertigkeiten von Ergebnisdarstellungen (z. B. Zahlen vs. Geschichten) kamen dabei zur Sprache. Die Einbindung von zentralen Stakeholder:innen könne unterschiedlichen Zukunftsszenarien Legitimität verschaffen, werfe aber auch Fragen zum Verhältnis von Aufwand und Nutzen von Foresight-Methoden auf.

Die Impulse aus den Vorträgen und Q&A-Runden diskutierten die Teilnehmenden am zweiten Tag in Arbeitsgruppen entlang der Fragen: 1. Welche Erkenntnisse nehmen Sie zum Zusammenhang von methodischer Qualität, Nützlichkeit und Nutzung mit? 2. Welche offenen Fragen und weiteren Forschungsbedarfe zu diesem Thema bleiben bestehen? Die Arbeitsgruppen präsentierten die nachfolgend zusammengefassten Ergebnisse hierzu im Plenum.

Eine zentrale Erkenntnis der Tagung ist, dass die methodische Qualität von Evaluierungen zwar eine notwendige, aber keine hinreichende Bedingung für die Nützlichkeitsperzeption und Nutzung durch die Stakeholder:innen der Evaluierungen darstellt. Hiervon ausgehend wurde diskutiert, wie sich die Kommunikation mit den relevanten Stakeholder:innen und die Vermittlung von Methodik und evaluativer Evidenz verbessern lassen. Zielgruppenspezifische Formate des Dialogs bzw. der Vermittlung sind hier bedeutsam, ebenso die Verständlichkeit von Evaluierungen und ihren Ergebnissen. Letztere könne zum Beispiel durch die Nutzung künstlicher Intelligenz (KI) oder durch Lesbarkeitsformeln (*readability formulas*, zur Messung der Verständlichkeit von Texten) sichergestellt werden. Die Nutzung von KI in Evaluationen wurde von den Tagungsteilnehmenden zudem als Feld weiterer Auseinandersetzung (u. a. im AK Methoden) identifiziert – hinsichtlich ihrer generellen Bedeutung sowie konkreter Nutzungsmöglichkeiten in Evaluationen.

Mit Blick auf Stakeholder:innen-Interaktion betonte die Arbeitsgruppe, intendierte Evaluierungsansätze und -designs noch stärker unter Nützlichkeitsaspekten in den *terms of references* und den *evaluation proposals* zu verankern. Zudem sollte mit den Auftraggebenden einer Evaluierung ausreichend Zeit für die Vermittlung der Methodik eingeplant werden, denn je verständlicher und akzeptierter die gewählte Methodik, desto wahrscheinlicher wird diese als nützlich wahrgenommen. Die Möglichkeit, Stakeholder:innen in die Auswahl der Evaluierungsfragen und der für ihre Beantwortung geeigneten Methoden einzubeziehen, wurde erörtert. Ein geteiltes Verständnis davon, welche Evaluierungsfragen mit welchen Methoden und welchen Ressourcen prioritär beantwortet werden sollen, könnte sich auch unter Effizienzgesichtspunkten positiv auf die wahrgenommene Nützlichkeit und Nutzung einer Evaluierung auswirken. Genauer wäre dabei aus Sicht der Tagungsteilnehmenden zu klären, wie in einem solchen partizipativen Prozess mit den unterschiedlichen Interessen von Stakeholder:innen umgegangen werden kann.

Der Zusammenhang der methodischen Qualität von Evaluierungen und deren Nutzung wurde auch aus einem politischen Blickwinkel beleuchtet und dabei die Gefahr der politisch motivierten Instrumentalisierung von Evaluierungen kritisch reflektiert. Die Teilnehmenden betonten, wie wichtig eine klare, möglichst nicht-mehrdeutige Sprache im Evaluierungsbericht und in der Kommunikation ist. Zugleich bestand Konsens darüber, dass ‚ungewollte Nutzungsweisen' zwar nicht gänzlich verhindert werden können, eine hohe methodische Qualität aber sehr wohl als ‚Lebensversicherung' fungiere, also weniger Angriffsfläche böte als ein objektiv kritikwürdiges Methodendesign. Damit zusammenhängend diskutierten die Teilnehmer:innen, wie widersprüchliche Evaluierungsbefunde (und Zielkonflikte) in einem politisch sensiblen Umfeld kommuniziert werden können. Als eine Möglichkeit wurde hier angeführt, Empfehlungen partizipativ zu erarbeiten – nach Sachstandsbericht im Falle einer Expertisen-Lücke oder zur differenzierten Bewertung widerstreitender Befunde.

Schließlich diskutierten die Teilnehmer:innen auch die Möglichkeit der Nützlichkeits- und Nutzungssteigerung von Evaluierungen auf Ebene konkreter Methoden. Der Einsatz prospektiver Methoden, die beispielsweise künftige Entwicklungen antizipieren und diese u. a. bei der Formulierung von Empfehlungen berücksichtigen, fand dabei großen Zuspruch. Die weitere Beschäftigung mit konkreten Methoden/Tools der Zukunftsforschung wurde als wichtiges Lernfeld identifiziert (z. B. unter anderem mit Blick auf ihre potentielle Nutzung für Personalbedarfsprognosen). Wie innovative (z. B. KI-basierte oder Foresight-)Methoden angesichts begrenzter Ressourcen genutzt werden können, wurde auch kritisch erörtert: Worauf sollen sich Evaluator:innen bei zunehmend komplexen Evaluierungsgegenständen, Skill-Ansprüchen und methodischen Angeboten bei gleichzeitig oftmals geringen Ressourcenausstattungen konzentrieren? Wie kann methodische Qualität auch bei geringen Ressourcen sichergestellt werden? Die hiermit verbundene Frage nach methodischen Mindeststandards im Spannungsverhältnis zwischen (gegenstandangemessenen) Methodenansprüchen und Möglichkeiten der Umsetzung wird auch weiterhin Gegenstand der Arbeit des AK Methoden in der Evaluation sein – sowohl in allgemeiner Betrachtung als auch mit Blick auf neue, konkrete Methoden(-anwendungen).

Die Dokumentation zur Tagung ist auf der Website des AK Methoden verfügbar.[4] Wir danken allen Referent:innen und Teilnehmer:innen für die anregenden Impulse und Diskussionen sowie seitens der AK-Sprecher:innen den Gastgeber:innen Katharina Chandon, Julia Dietl, Nils Verwohlt und Klaus Heinle von der Stabsstelle Evaluierung der Hanns-Seidel-Stiftung für die sehr gute Zusammenarbeit, für die Möglichkeit, die Räumlichkeiten des Konferenzzentrums zu nutzen und für ein tolles Ambiente rund um die Tagung.

Literatur

Bundi, P., Frey, K., & Widmer, T. (2021). Does Evaluation Quality Enhance Evaluation Use? *Evidence & Policy*, *17*(4), 661–687. https://doi.org/10.1332/174426421x16141794148067

DeGEval – Gesellschaft für Evaluation (2017). *Standards für Evaluation. Erste Revision 2016.* https://www.degeval.org/fileadmin/content/Z03_Publikationen/DeGEval-Standards/DeGEval-Standards_fuer_Evaluation.pdf

Fleischer, D. N. & Christie, A. C. (2009). Evaluation Use: Results From a Survey of U. S. American Evaluation Association Members. *American Journal of Evaluation*, *30*(2), 158–175. https://doi.org/10.1177/1098214008331009

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation, 30*(3), 377–410. https://doi.org/10.1177/1098214009341660

Leppert, G, Kocks, A., Noltze, M, Erler, J, & Harten, S. (in Vorbereitung). *A Systematic Approach to Method Integration (SAMI) for Compex Inquiries in Intervention Research*.

OECD (o. J.). *Strategic Foresight*. https://www.oecd.org/en/about/programmes/strategic-foresight.html

Silvestrini, S. (2023). Meta im Quadrat – was wir aus Meta-Evaluationen lernen können. *Zeitschrift für Evaluation, 22*(1), 120–146. https://doi.org/10.31244/zfe.2023.01.07

Dr. Franziska Heinze | Brandenburgische Technische Universität Cottbus-Senftenberg | Lipezker-Straße 47, D-03048 Cottbus | E-Mail: ak-methoden@degeval.org

Dr. Alexander Kocks | Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) | Fritz-Schäffer-Str. 26, D-53113 Bonn | E-Mail: ak-methoden@degeval.org

Klaus Heinle | Hanns-Seidel-Stiftung, Stabsstelle Evaluierung | Lazarettstr. 33, D-80636 München | E-Mail: heinle@hss.de

Nils Verwohlt | Hanns-Seidel-Stiftung, Stabsstelle Evaluierung | Lazarettstr. 33, D-80636 München | E-Mail: verwohlt-n@hss.de

4 https://www.degeval.org/arbeitskreise/methoden-in-der-evaluation/aktuelles