

---

# Nicht jede Evaluation ist eine gute Evaluation Warum gute Evaluationen fachliche Standards berücksichtigen sollten

Jan Hense

**Damit Evaluationen die in sie gesetzten Erwartungen erfüllen können, müssen fachliche Qualitätsansprüche berücksichtigt werden. Evaluationsstandards wie die der Gesellschaft für Evaluation formulieren solche Ansprüche. Sie geben damit allen, die in der Praxis für Evaluationen verantwortlich oder mit ihnen befasst sind, Orientierung für ihr Tun und helfen bei der Bewertung von Evaluationen und ihren Ergebnissen.**

## Auch Evaluationen lassen sich evaluieren

Auch wenn Evaluation und Forschung auf den ersten Blick viel gemeinsam haben und sich oft der gleichen Methoden wie z. B. Interviews, Fragebögen und Statistiken bedienen, gibt es wichtige Unterschiede. Einer der wichtigsten ist, dass Evaluationen immer auf einen oder mehrere konkrete, vorab bereits bekannte Zwecke hin durchgeführt werden. Die vier wichtigsten möglichen Zwecke sind die Verbesserung bzw. Qualitätssicherung des Gegenstands, etwa eines aufsuchenden Beratungsangebots, das Herstellen von Rechenschaft, z. B. über ein schulisches Aufklärungsprogramm, die Unterstützung von Entscheidungen, etwa über Fortführung oder Einstellung eines Fortbildungsangebots, und das Hinzuwachsen, z. B. bei der Erprobung eines innovativen digitalen Informationsportals. In all diesen Fällen ist Evaluation nie Selbstzweck, sondern immer Mittel zum Zweck. Vom Ende her gedacht könnte man also sagen, dass eine gute Evaluation eine ist, die unter den gegebenen Bedingungen eine oder mehrere dieser möglichen, ihr zugeordneten Funktionen in möglichst guter Weise erfüllen kann.

Mit dieser Zweckgebundenheit teilt Evaluation eine wichtige Gemeinsamkeit mit den Maßnahmen, Projekten oder Praktiken, die sie evaluiert: Sie will grundsätzlich etwas in Bewegung bringen und einen Unterschied in der Praxis machen, folgt also selbst einer Interventionslogik. Es sollte also nicht überraschen, dass Evaluationen ebenso evaluiert werden können wie die Gegenstände, die sie untersuchen, in der Fachsprache handelt es sich dann um Meta-Evaluationen (CASPARI, 2015). Doch auch jenseits einer systematischen Meta-Evaluation ist es für die praktische Arbeit mit Evaluationen wichtig, deren Qualität gestalten und bewerten zu können.

Doch woran erkennt man gute Evaluationen? Und mindestens ebenso wichtig: Was können jene, die für Evaluationen Verantwortung tragen, bei ihrer Vorbereitung, Planung und Durchführung tun, um gute Evaluationen zu gestalten, die die in sie gesetzten Erwartungen erfüllen können? Mit diesen Fragen hat sich die Evaluation als Disziplin schon früh aus-

einandergesetzt, was zur Formulierung und kontinuierlichen Weiterentwicklung eigener fachlicher Standards geführt hat. In diesem Beitrag werden diese vorgestellt und es wird beispielhaft erläutert, was bei ihrer Umsetzung zu beachten ist.

## Professionelle Standards guter Evaluationspraxis

Im deutschen Sprachraum wohl am meisten verbreitet sind die Standards für Evaluation der Gesellschaft für Evaluation (DeGEval e. V.). Die DeGEval wurde 1997 als Zusammenschluss von Personen und Institutionen gegründet, die im Bereich der Evaluation tätig sind, wobei die Entwicklung und Verbreitung fachlicher Standards eines der wichtigsten Ziele war und ist. Die »Standards für Evaluation« wurden 2001 erstmals verabschiedet und liegen aktuell in einer 2016 revidierten Fassung vor (DEGEVAL – GESELLSCHAFT FÜR EVALUATION, 2017). Sie stehen allen Interessierten als Kurz- oder Langfassung mit ausführlichen Hintergrundinformationen und Erläuterungen zum freien Download auf den Seiten der DeGEval ([www.degeval.org](http://www.degeval.org)) zur Verfügung oder sind dort als Druckfassung zu beziehen.

Gemäß den DeGEval-Standards zeichnen sich gute Evaluationen durch vier zentrale Attribute aus: Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit. Die Standards differenzieren und präzisieren diese vier Attribute in Form von insgesamt 25 Einzelstandards, die in der Langfassung jeweils begründet und in Bezug auf ihre praktische Bedeutung erläutert werden. Im Folgenden werden sie beispielhaft anhand der aus den vier Attributen gebildeten Standardgruppen vorgestellt.

## Gute Evaluationen sind nützlich

Die Nützlichkeitsstandards fordern, dass Evaluationen grundsätzlich danach streben sollten, einen Mehrwert für die Praxis zu erbringen, z. B., indem konkrete Verbesserungs-

möglichkeiten aufgezeigt werden oder die Entscheidungsfindung über einen Evaluationsgegenstand erleichtert wird. Die Gruppe umfasst acht Einzelstandards.

So fordert etwa der Standard »N2 – Klärung der Evaluationszwecke«, dass frühzeitig im Prozess der Evaluation ausgehandelt, festgelegt und priorisiert wird, welche Zwecke sie konkret verfolgen soll. Dies ist wichtig, um die Evaluation passgenau auf diesen Zweck und die intendierte Verwendung ihrer Ergebnisse hin gestalten zu können, denn wenn etwa ein Beratungsangebot evaluiert werden soll, macht es bereits bei der Planung der Evaluation einen großen Unterschied, ob es um die Weiterentwicklung der Beratungspraxis gehen soll oder darum, einer Rechenschaftspflicht gegenüber den Mittelgebenden zu genügen.

Ein weiteres Beispiel dieser Gruppe von Standards ist »N1 – Identifizierung der Beteiligten und Betroffenen«. Er verlangt, dass wichtige Interessensgruppen auf Seiten der Evaluation und des Evaluationsgegenstands berücksichtigt werden, um deren Anliegen und Informationsbedürfnisse zu klären und nach Möglichkeit in der Evaluation zu berücksichtigen. Der Standard N1 trägt der aus langjähriger Forschung über die Nutzung von Evaluationen gewonnenen Erkenntnis Rechnung, dass Evaluationen umso eher genutzt werden, je mehr Beteiligte und Betroffene in den Evaluationsprozess einbezogen werden (JOHNSON et al., 2009).

Die weiteren Standards der Nützlichkeitsgruppe thematisieren u. a. die Kompetenzen und die Glaubwürdigkeit von Evaluierenden als wichtige Voraussetzung von Nützlichkeit, eine angemessene Auswahl von Informationen, die Transparenz von Werthaltungen, die der systematischen Bewertung zugrunde liegen, die Vollständigkeit und Klarheit der Berichterstattung und die Rechtzeitigkeit von Ausschreibung bzw. Beauftragung und Durchführung der Evaluation.

### Gute Evaluationen sind durchführbar

Evaluationen sind in der Praxis häufig mit erheblichen Herausforderungen bezüglich ihrer Umsetzbarkeit konfrontiert. Zu ihnen zählen u. a. zeitliche oder budgetäre Restriktionen, Schwierigkeiten beim Feldzugang und bei Datenerhebungen oder fehlende Akzeptanz oder gar Widerstände von Beteiligten oder Betroffenen. Die Gruppe der drei Durchführbarkeitsstandards trägt diesem Sachverhalt Rechnung. Sie sollen sicherstellen, dass angemessene Verfahren zum Einsatz kommen, dass die Evaluation bewusst und sensibel im Handlungsfeld agiert und dass dabei Aufwand und Nutzen in einem angemessenen Verhältnis stehen.

Konkret formuliert der Standard »D1 – Angemessene Verfahren« den Anspruch, dass Evaluationsverfahren, insbesondere bei der Datenerhebung, nicht nur professionellen Gütekriterien entsprechen (vgl. unten die Gruppe der Genauigkeitsstandards), sondern auch passgenau für die jeweiligen Evaluationszwecke und -fragen ausgewählt werden und die Praxis nicht über Gebühr belasten. Das aus dem Datenschutz bekannte Prinzip der Datensparsamkeit gilt also auch hier. Standard »D2 – Diplomatisches Vorgehen« geht darauf zurück, dass Evaluationen oft in Konstellationen erfolgen, die von Interessenskonflikten geprägt sind. Er verlangt, mit diesen bewusst und sensibel umzugehen, um eine möglichst hohe Akzeptanz der Evaluation sicherzustellen. Standard »D3 – Effizienz von Evaluation« thematisiert die eingangs angesprochenen Gedanken, dass Evaluation selbst einen Mehrwert für die Praxis haben muss, und geht sogar darüber

hinaus, indem er fordert, dass dieser Mehrwert den eingesetzten Aufwand sowohl auf Seiten der Evaluation als auch der evaluierten Praxis aufwiegen sollte.

### Gute Evaluationen sind fair

Evaluationen setzen angemessene Verfahren ein, um in systematischer und nachvollziehbarer Weise Bewertungen vorzunehmen. Die fünf Fairnessstandards sollen gewährleisten, dass Werturteile wie »wirksam«, »gut geeignet«, »wenig effektiv« oder »verbesserungswürdig« transparent zustande kommen und dabei die Interessen von Beteiligten und Betroffenen angemessen gewahrt werden.

Der Standard »F3 – Umfassende und faire Prüfung« wendet sich etwa gegen einseitige Betrachtungen, bei denen z. B. nur Schwachpunkte einer Maßnahme betont werden, mögliche Stärken oder positive Nebeneffekte aber außer Betracht bleiben. Im Standard »F2 – Schutz individueller Rechte« wird u. a. der seit Einführung der Datenschutz-Grundverordnung (DSGVO) auch rechtlich sehr relevante Datenschutz in Evaluationen thematisiert. Weitere Einzelstandards der Fairnessgruppe behandeln die Notwendigkeit formaler Vereinbarungen, die im Zweifel Schutz vor unfairen Vorgehensweisen bieten, fordern eine unparteiische Durchführung und Berichterstattung und die Offenlegung von Ergebnissen und Berichten für die Allgemeinheit.

### Gute Evaluationen sind genau

Schließlich thematisieren die Standards der letzten Gruppe die Notwendigkeit der Genauigkeit von Evaluationsverfahren und -ergebnissen. Hier finden sich mehrere Standards, wie sie auch sonst für Methoden empirischer Forschung gelten, wie etwa zur vollständigen Angabe von Informationsquellen, zur Nutzung genauer und gültiger Informationen, zur systematischen Fehlerprüfung oder zu angemessenen Auswertungsverfahren. Dabei verhalten sich die Standards bewusst neutral zur Methodenwahl, sind also sowohl für quantitative wie für qualitative Verfahren anwendbar.

Besonders werden hier aber Aspekte angesprochen, die evaluationsspezifischer sind. Darunter fällt etwa »G1 – Beschreibung des Evaluationsgegenstandes«, wo u. a. eine Beschreibung des Wirkungsmodells (Theory of Change) der Maßnahme empfohlen wird, das darstellt, unter welchen Bedingungen und mittels welcher Ressourcen und Mechanismen eine Maßnahme zu ihren Wirkungen beiträgt (HENSE & TAUT, 2021). Auch die Forderung nach methodisch nachvollziehbaren Bewertungen wird hier konkret im Standard »G8 – Begründete Bewertungen und Schlussfolgerungen« angesprochen. Denn Bewertungen beruhen immer auf bestimmten Kriterien wie z. B. Wissenszuwachs, Einstellungs- oder Verhaltensänderung von Beratenen, Beteiligungsquote, Akzeptanz der Zielgruppen etc. und auf Vergleichswerten, auf deren Basis Aussagen wie »eine sehr gute Beteiligung« oder »eine wenig wirksame Beratungsleistung« fußen. Beides, die Auswahl von Kriterien wie die der verwendeten Vergleichswerte, muss explizit und transparent erfolgen, da Bewertungen und Schlussfolgerungen sonst willkürlich und nicht nachvollziehbar sind.

## Für die Qualität von Evaluation sind viele mitverantwortlich

Ein häufiges Missverständnis in Bezug auf Evaluationsstandards ist, dass diese sich im Sinne einer Handlungsanleitung ausschließlich an jene richten, die die Evaluation verantwortlich durchführen. In vielen Fällen können die Evaluierenden aber wesentliche Faktoren der Evaluationsqualität nicht oder kaum beeinflussen.

Dies zeigt z. B. der Standard »N7 – Rechtzeitigkeit der Evaluation«. Zwar ist es Sache der Evaluierenden, gemäß einem vereinbarten Zeitplan fristgemäß vorzugehen. Wenn aber ein vorab vereinbarter Datenzugang, z. B. Interviews mit wichtigen Maßnahmenbeteiligten, durch die Projektleitung erst verspätet ermöglicht wird, kann die Evaluation dagegen wenig tun. Noch häufiger ist das Problem, dass Evaluationen zu spät, z. B. erst gegen Projektende, ausgeschrieben oder beauftragt werden, so dass die Evaluationsergebnisse für wichtige Entscheidungen oder Verbesserungen gar nicht mehr pünktlich vorliegen und berücksichtigt werden können.

Ähnliches gilt für »F5 – Offenlegung von Ergebnissen und Berichten«. Im Sinne der Transparenz und des fachlichen Erkenntnisfortschritts sollten Evaluationsergebnisse, insbesondere wenn sie sich auf öffentlich getragene Maßnahmen beziehen, immer allgemein zugänglich gemacht werden. Vor allem, wenn es sich um (mikro-)politisch umstrittene Themen handelt, scheuen Auftraggebende von Evaluationen aber oft genug diese Transparenz und setzen durch, dass Evaluationsergebnisse nicht oder nur selektiv veröffentlicht werden.

Die Qualität von Evaluationen hat also viele Eltern. Neben den Evaluierenden gehören dazu Auftraggebende, Entscheidungstragende, Projektbeteiligte, Zielgruppenangehörige und je nach Konstellation weitere Akteure. Vor allem die erstgenannten Gruppen sollten sich dessen bewusst sein und mindestens mit den für sie relevanten Evaluationsstandards vertraut sein, damit sie ihren Beitrag leisten können, um nützliche, durchführbare, faire und genaue Evaluationen zu ermöglichen.

## Standards schaffen Vertrauen und Akzeptanz

Seit ihrer Einführung 2001 haben die Standards für Evaluation große Akzeptanz und Verbreitung gefunden. Dies lässt sich etwa daran ablesen, dass Ausschreibungen immer häufiger ihre Berücksichtigung einfordern. Obwohl die Evaluationsstandards von ihrer Formulierung her nicht auf bestimmte Handlungsfelder festgelegt sind, ist ihre Verbreitung nicht einheitlich. So spielen in manchen Bereichen wie etwa dem der Entwicklungszusammenarbeit auch alternative Standards internationaler Organisationen wie der UN (UNITED NATIONS EVALUATION GROUP, 2016) eine übergeordnete Rolle. Dies ist aber insofern problematisch, als dass zwischen den Regelwerken inhaltlich große Überschneidungen bestehen. Eine Übersicht über die Verbreitung, Nutzung und Herausforderungen in verschiedenen Politik- und Praxisfeldern, darunter auch dem Gesundheitsbereich, gibt der Sammelband von HENSE et al. (2019).

Die Entwicklung und Aushandlung von fachlichen Standards ist ein wichtiger Schritt in der Entwicklung aller neuen Disziplinen. Verglichen mit anderen Tätigkeitsfeldern ist Evaluation ein relativ junges Handlungsfeld, das mit den Standards für Evaluation bereits früh diesen Schritt vollzogen hat. Bisher fehlen aber – vielleicht auch aus gutem Grund

– Mechanismen, die die Berücksichtigung von Evaluationsstandards durchsetzbar oder einklagbar machen. Auch ist der Begriff der Evaluation bisher nicht geschützt und die Ausübung von Evaluationstätigkeiten nicht auf den Nachweis einschlägiger fachlicher Kompetenzen angewiesen (vgl. DeGEVAL – GESELLSCHAFT FÜR EVALUATION, 2008). Daher erscheint es umso wichtiger, dass alle, die egal in welcher Rolle mit Evaluation befasst sind, wissen, was gute Evaluationen ausmacht und welchen Beitrag sie dafür leisten können, kurzum: dass sie mit fachlichen Standards für Evaluation vertraut sind.



Prof. Dr. Jan Ulrich Hense, freier Berater für Evaluation und Wirkungsorientierung, Gießen.

**Kontakt:**  
mail@jan-hense.de

### Zitierweise

Hense, J. U. (2021). Nicht jede Evaluation ist eine gute Evaluation. Warum gute Evaluationen fachliche Standards berücksichtigen sollten, FORUM Sexualaufklärung und Familienplanung: Informationsdienst der Bundeszentrale für gesundheitliche Aufklärung (BZgA), 1, 3-5.  
[https://doi.org/10.17623/BZgA\\_SRH:forum\\_2021-1\\_beitrag\\_eval\\_standards](https://doi.org/10.17623/BZgA_SRH:forum_2021-1_beitrag_eval_standards)

### Literatur

- CASPARI, A. (2015). Well done? Who knows ... Ein Plädoyer für Meta-Evaluationen. In HENNEFELD, V., MEYER, W., & S. SILVESTRINI (Hrsg.), *Nachhaltige Evaluation? Auftragsforschung zwischen Praxis und Wissenschaft* (S. 143–166). Münster: Waxmann.
- DeGEVAL – GESELLSCHAFT FÜR EVALUATION (2008). *Empfehlungen zur Aus- und Weiterbildung in Evaluation: Anforderungsprofile an Evaluatorinnen und Evaluatoren* (2. Aufl.). DeGEVAL. <https://www.degeval.org/publikationen/aus-und-weiterbildung/>
- DeGEVAL – GESELLSCHAFT FÜR EVALUATION (2017). *Standards für Evaluation. Erste Revision 2016*. Gesellschaft für Evaluation e. V. <https://www.degeval.org/de/degeval-standards/>
- HENSE, J., BÖTTCHER, W., KALMAN, M., & MEYER, W. (Hrsg.). (2019). *Evaluation: Standards in unterschiedlichen Handlungsfeldern: Einheitliche Qualitätsansprüche trotz heterogener Praxis?* Münster: Waxmann.
- HENSE, J., & TAUT, S. (2021). Wie Wirkungsmodelle zur Wirkung kommen. Nutzen und Kosten in der Evaluationspraxis: Einführungsbeitrag zum Themenheft »Wirkungsmodelle und Programmtheorien in der Evaluationspraxis«. *Zeitschrift für Evaluation*, 20(2).
- 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>
- UNITED NATIONS EVALUATION GROUP (2016). *Norms and Standards for Evaluation*. UNEG. <http://www.uneval.org/document/detail/1914>