



EHB

EIDGENÖSSISCHE
HOCHSCHULE FÜR
BERUFSBILDUNG

Schweizer Exzellenz in Berufsbildung

VERSTÄNDNIS UND NUTZUNG VON BEWERTUNGSKRITERIEN

Véronique Eicher, Lars Balzer, Corinna Bumann, Céline Mavrot, Melike Ömerogullari, Oto Potluka, François-Xavier Viallon

27. DeGEval Tagung in Potsdam, September 2024

Hintergrund

- Bewertungskriterien sind zentral für jede Evaluation und Voraussetzung für die systematische und faire Bewertung des Evaluationsgegenstandes. (Balzer & Beywl, 2018)
- «Die Bestimmung von Bewertungskriterien ist von den Evaluationsexperten [...] als eine wichtige und gleichzeitig in der Praxis als unzureichend beachtete Erfolgsbedingung im Evaluationsprozess identifiziert worden». (Balzer, 2005, S. 201)
- «Die Evaluationsberichte, welche die Bundesverwaltung extern in Auftrag gibt, schenken dem Thema Bewertungskriterien und ihrer Verwendung in den untersuchten Fallbeispielen sehr unterschiedlich und der Tendenz nach eher wenig Aufmerksamkeit.» (Heuer, 2017, S. 343)
- In der Evaluationsliteratur gibt es unterschiedliche Ansichten, wer die Bewertungskriterien festlegen sollte:
 - Beteiligte und Betroffene (Stake, 1995; Stake & Schwandt, 2006)
 - Evaluierende (Scriven, 2007)
 - Auftraggebende (Balzer, 2005)
 - Partizipativer Prozess (Beywl, 2006)

Fragestellungen und Umsetzung

- Explorative Fragestellungen:
 - Was verstehen verschiedene Akteur:innen der Evaluationscommunity unter Bewertungskriterien?
 - Wie nutzen sie Bewertungskriterien?
 - Welche Bewertungskriterien werden genutzt?
 - Wer legt die Bewertungskriterien fest?
- Umfrage bei Akteur:innen der Evaluationscommunity
- Anlass: SEVAL-Tagung «Von der Bewertung zur Methode... und zurück» im September 2022
- Versand einer Online-Umfrage an Teilnehmende der Tagung
 - 3 offene Fragen (Auswertung in MAXQDA)
 - 1 geschlossene Frage (Auswertung in SPSS)
 - Fragen zur Person (Erfahrung, Evaluationsbereich, usw.)

Stichprobe: 66 Personen

Online-Umfrage 201mal angeklickt, davon 69mal abgesendet, davon 66 mindestens 1 offene Frage ausgefüllt

Hauptsächliche Evaluationserfahrung:

- 41 Personen (63.1%) führen selbst Evaluationen durch
- 11 Personen (16.9%) geben Evaluationen in Auftrag
- 5 Personen (7.7%) beraten bei Evaluationen
- 4 Personen (6.2%) nutzen Evaluationsergebnisse
- 3 Personen (4.6%) bilden im Bereich der Evaluation aus
- 1 Person (1.5%) forscht über Evaluation

Expertisegrad Evaluationstheorie

- 31 Personen (48.4%) mind. hohe Fachkenntnis

Hauptsächlicher Themenbereich:

- 15 Personen (23.1%) in der Bildung
- 12 Personen (18.5%) in der Gesundheit
- 7 Personen (10.8%) im Sozialen
- 7 Personen (10.8%) in der Umwelt
- 7 Personen (10.8%) in der Entwicklungszusammenarbeit
- und weitere

Expertisegrad Evaluationspraxis

- 54 Personen (83.1%) mind. hohe Fachkenntnis

Offene Fragen: Kodierung der ersten beiden Fragen

- Standards der grösseren Evaluationsgesellschaften (SEVAL, DeGEval, Program Evaluation Standards) wurden bzgl. Bewertungskriterien analysiert, um Codesystem zu erstellen
«Die Bewertungskriterien definieren die Aspekte, anhand derer der Evaluationsgegenstand bewertet wird. Dabei kann es sich z.B. um die Wirksamkeit, die Effizienz oder die inhaltliche Kohärenz handeln. Die Bewertungskriterien müssen ausreichend präzise definiert und explizit dargelegt werden. Allen an einer Evaluation Beteiligten muss klar sein, nach welchen Kriterien der Evaluationsgegenstand bewertet wird. Dies setzt voraus, dass sie die Bewertungskriterien diskutieren und sie gegebenenfalls gemeinsam entwickeln. Die Kriterien sind zu operationalisieren, d.h. es gilt festzulegen, wie sie konkret gemessen werden.»
(SEVAL Standards 2016, Seite 13, als Beispiel)
- Codes basierend auf diesen Quellen
Bewertung Evaluationsgegenstand / Merkmal des Evaluationsgegenstands (das evaluiert wird) / Stakeholder : gemeinsam diskutiert, entwickelt / explizit, transparent, Begründung / Operationalisierung, messbar, präzise, konkret / Beispiele (Wirksamkeit, Effizienz, inhaltliche Kohärenz) / Schwellenwert (aus DeGEval Standards) / Vor Datenerhebung festgelegt (aus American Evaluation Standards)
- Zusätzliche Codes basierend auf den Antworten
Herleitung aus Evaluationsgegenstand / Zusammenhang mit Fragestellungen / Deduktive Herleitung / Unterscheidung quanti-quali

deduktiv

Code	Beschreibung	Beispiel
Bewertung des Evaluationsgegenstands	Es wird ein Werturteil gefällt, es geht darum die Güte / den Nutzen zu bestimmen	«Bewertung eines Evaluationsgegenstandes»
Merkmal des Evaluationsgegenstands	Es werden Merkmale / Aspekte des Gegenstands genannt, der bewertet wird	«Dimensionen, an denen der Evaluationsgegenstand am Ende gemessen wird»
Schwellenwert	Es wird von einem Schwellenwert gesprochen, ab dem die Bewertung kippt	«Massstäbe, nach denen in Evaluation gemessen wird»
Stakeholder	Es werden Stakeholder genannt, die einbezogen werden (auch implizit, wenn gemeinsam entwickelt / diskutiert wird)	«mit dem Auftraggeber zusammen zu definierende Kriterien»
Begründung	Es wird explizit oder transparent gemacht, wie die Kriterien hergeleitet wurden, auch objektiv und systematisch fällt hierunter	«Kriterien sind im Sinne der Transparenz offenzulegen»
Operationalisierung	Es wird darauf hingewiesen, dass die Kriterien messbar sein müssen, konkret(isiert) sein müssen, oder auch präzise	«messbar»
Vor Datenerhebung festgelegt	Es wird gesagt, dass das Kriterium vor der Datenerhebung festgelegt sein muss (auch implizit, z.B. bei der Konzeptualisierung, usw.)	«im voraus festgelegt»
Beispiele für Kriterien	Es werden Beispiele für Kriterien genannt	«six criteria from the OECD DAC»
Herleitung aus Evaluationsgegenstand	Es werden Ziele, Merkmale vom Evaluationsgegenstand, Dokumente genannt (z.B. Konzepte), aus denen die Kriterien hergeleitet sind	«Ziele des Projekts in Bewertungskriterien operationalisiert» (auch bei Operationalisierung)
Zusammenhang mit Fragestellungen	Es wird ein direkter Zusammenhang zwischen Kriterien und Evaluationszweck oder Evaluationsfragestellungen gemacht	«zur Beantwortung einer Evaluationsfrage dienen»
Deduktive Herleitung	Es werden «Standard»-Kriterien genannt, die in der Praxis üblich sind (u.a. auch gesetzlicher Rahmen)	«rechtliche Grundlagen»
Unterscheidung quanti-quali	Es wird beschrieben, dass die Kriterien quantitativ oder qualitativ sein können	«Diese können ganz unterschiedlicher Art und qualitativer und/oder quantitativer Natur sein»

induktiv

Ergebnisse

Nach Ihrer Erfahrung sind Bewertungskriterien in Evaluationsprojekten, mit denen Sie zu tun haben, oftmals (alles ankreuzen, was zutrifft):

	Alle (n = 66)	Evaluierende (n = 41)	Auftraggebende (n = 11)
Irrelevant	6%	7%	9%
Vorher vorgegeben	47%	39%	64%
Vor Datenauswertung ausgearbeitet	74%	73%	82%
Während Datenauswertung ausgearbeitet	26%	17%	27%
Im Nachhinein bestimmt	9%	12%	0.0%
In Verantwortung der Auftraggebenden	30%	29%	46%
In Verantwortung der relevanten Akteure	67%	61%	73%
In Verantwortung der Evaluierenden	44%	54%	27%

Lesebeispiel: 74% aller Personen geben an, dass nach ihrer Erfahrung, Bewertungskriterien «vor Datenauswertung ausgearbeitet» werden.

Nach Ihrer Erfahrung sind Bewertungskriterien in Evaluationsprojekten, mit denen Sie zu tun haben, oftmals (alles ankreuzen, was zutrifft):

	Alle (n = 66)	Bildung (n = 15)	Gesundheit (n = 12)
Irrelevant	6%	-	17%
Vorher vorgegeben	47%	47%	33%
Vor Datenauswertung ausgearbeitet	74%	80%	75%
Während Datenauswertung ausgearbeitet	26%	33%	25%
Im Nachhinein bestimmt	9%	13%	8%
In Verantwortung der Auftraggebenden	30%	27%	25%
In Verantwortung der relevanten Akteure	67%	73%	67%
In Verantwortung der Evaluierenden	44%	33%	42%

Lesebeispiel: 74% aller Personen geben an, dass nach ihrer Erfahrung, Bewertungskriterien «vor Datenauswertung ausgearbeitet» werden.

Wer legt die Bewertungskriterien fest?

- Alleinige Verantwortung der relevanten Akteure: 36%
 - Alleinige Verantwortung der Evaluierenden: 15%
 - Gemeinsame Verantwortung der relevanten Akteure + Evaluierende: 14%
 - für 65% der TN sind Auftraggebende nicht in der Verantwortung!
- Evaluierende werden als **wichtiger als Auftraggebende** gesehen
- **Relevante Akteure** → Beteiligte und Betroffene bzw. partizipativer Prozess?
- Zusätzlich spannend:
 - 6%: irrelevant
 - 14%: weder «vorher vorgegeben» noch «vor Datenauswertung»

Bitte beschreiben Sie, was Sie unter Bewertungskriterien verstehen.

Häufigste Codes ($\geq 15\%$) – insgesamt 12 Codes	Alle (n = 66)	Evaluierende (n = 41)	Auftraggebende (n = 11)
Bewertung des Gegenstands	79%	78% (1)	82% (1)
Schwellenwert	32%	27% (4)	55% (2)
Operationalisierung	30%	29% (3)	36% (3)
Merkmal des Gegenstands	27%	27% (4)	18% (5)
Beispiele	26%	34% (2)	18% (5)
Begründung	21%	17% (6)	18% (5)
Zusammenhang mit Fragestellungen	12%	17% (6)	9% (8)
Unterscheidung quanti-quali	12%	15% (8)	9% (8)
Vor Datenerhebung festgelegt	11%	10% (9)	27% (4)
Anzahl Antworten	96% (63)	95% (39)	100% (11)
Anzahl Code-Nennungen pro Person	2.7	2.7	2.8

Bitte beschreiben Sie, was Sie unter Bewertungskriterien verstehen.

Häufigste Codes ($\geq 15\%$) – insgesamt 12 Codes	Alle (n = 66)	Bildung (n = 15)	Gesundheit (n = 12)
Bewertung des Gegenstands	79%	73% (1)	92% (1)
Schwellenwert	32%	27% (4)	17% (5)
Operationalisierung	30%	47% (2)	25% (4)
Merkmal des Gegenstands	27%	40% (3)	42% (3)
Beispiele	26%	7% (7)	50% (2)
Begründung	21%	20% (5)	17% (5)
Zusammenhang mit Fragestellungen	12%	7% (7)	8% (7)
Unterscheidung quanti-quali	12%	7% (7)	-
Vor Datenerhebung festgelegt	11%	13% (6)	8% (7)
Anzahl Antworten	96% (63)	100% (15)	100% (12)
Anzahl Code-Nennungen pro Person	2.7	2.5	2.6

Was verstehen Akteur:innen der Evaluationscommunity unter Bewertungskriterien?

- **Bewertung des Gegenstands** wird von den meisten Teilnehmenden genannt
 - **Operationalisierung** wird ebenfalls von ca. einem Drittel aller Teilnehmenden genannt
 - Von Auftraggebenden wird zusätzlich der **Schwellenwert** genannt und dass die Kriterien **vor der Datenerhebung** festgelegt sind
 - Evaluierende und Personen aus dem Gesundheitsbereich nennen zudem häufig **Beispiele**
- **Grosser Konsens** beim Verständnis der Bewertungskriterien
- mit **unterschiedlichen Schwerpunktsetzungen**

Wie werden Bewertungskriterien in Evaluationsprojekten, mit denen Sie zu tun haben, verwendet? Beschreiben Sie bitte kurz, was für Sie wichtig ist.

Häufigste Codes ($\geq 15\%$) – insgesamt 14 Codes	Alle (n = 66)	Evaluierende (n = 41)	Auftraggebende (n = 11)
Operationalisierung	38%	32% (3)	46% (2)
Vor Datenerhebung festgelegt	35%	34% (1)	46% (2)
Begründung	33%	29% (5)	55% (1)
Stakeholder	32%	32% (3)	18% (7)
Zusammenhang mit Fragestellungen	32%	34% (1)	36% (4)
Beispiele	29%	29% (5)	18% (7)
Herleitung aus Gegenstand	23%	17% (7)	9% (10)
Bewertung des Gegenstands	21%	17% (7)	36% (4)
Grosse Bedeutung	20%	12% (9)	18% (7)
Schwellenwert	15%	7% (12)	27% (6)
Anzahl Antworten	91% (60)	90% (37)	100% (11)
Anzahl Code-Nennungen pro Person	3.4	3.0	3.2

Wie werden Bewertungskriterien in Evaluationsprojekten, mit denen Sie zu tun haben, verwendet? Beschreiben Sie bitte kurz, was für Sie wichtig ist.

Häufigste Codes ($\geq 15\%$) – insgesamt 14 Codes	Alle (n = 66)	Bildung (n = 15)	Gesundheit (n = 12)
Operationalisierung	38%	13% (10)	33% (3)
Vor Datenerhebung festgelegt	35%	27% (3)	50% (1)
Begründung	33%	27% (3)	25% (6)
Stakeholder	32%	47% (1)	33% (3)
Zusammenhang mit Fragestellungen	32%	7% (11)	33% (3)
Beispiele	29%	27% (3)	42% (2)
Herleitung aus Gegenstand	23%	47% (1)	17% (8)
Bewertung des Gegenstands	21%	27% (3)	25% (6)
Grosse Bedeutung	20%	20% (8)	17% (8)
Schwellenwert	15%	20% (8)	8% (10)
Anzahl Antworten	91% (60)	80% (12)	83% (10)
Anzahl Code-Nennungen pro Person	3.4	3.7	3.7

Wie nutzen Akteur:innen der Evaluationscommunity Bewertungskriterien?

- **Operationalisierung, Festlegung vor Datenerhebung und Zusammenhang mit Fragestellungen** werden am häufigsten genannt, nur von Personen im Bildungsbereich werden alle relativ wenig genannt
 - **Gemeinsame Entwicklung mit Stakeholdern** für Evaluierende wichtiger als für Auftraggebende
 - **Begründung und Bewertung des Gegenstands** für Auftraggebende wichtiger als für Evaluierende
- Antworten sind **deutlich heterogener** als bei Verständnis der Bewertungskriterien
- Antworten gehen **über Standards hinaus** (Zusammenhang mit Fragestellungen, Herleitung aus Gegenstand)
- Antworten unterscheiden sich von erster Frage (nur Operationalisierung wird ähnlich häufig genannt)

Falls Bewertungskriterien in Evaluationsprojekten, mit denen Sie zu tun haben, verwendet werden:
Gibt es solche, die immer wieder verwendet werden? Welches sind die wichtigsten?

Häufigste Codes ($\geq 10\%$) - insgesamt 31 Codes	Alle (n = 66)	Evaluierende (n = 41)	Auftraggebende (n = 11)
Wirksamkeit/Effektivität	46%	46% (1)	55% (1)
Kosten-Nutzen/Effizienz	38%	29% (2)	55% (1)
Relevanz	20%	27% (3)	18% (6)
Jedes Mal andere Kriterien	20%	22% (4)	36% (3)
Kohärenz	17%	20% (5)	18% (6)
Nachhaltigkeit	15%	12% (6)	36% (3)
Zweckmässigkeit	14%	12% (6)	27% (5)
Zufriedenheit	11%	7% (9)	9% (8)
Zielerreichung	9%	7% (9)	-
Rechtmässigkeit	8%	7% (9)	9% (8)
Akzeptanz	3%	2% (15)	9% (8)
Anzahl Antworten mit Beispielen	74% (49)	73% (30)	55% (6)
Anzahl Beispiele pro Person	3.0	3.0	5.0

Falls Bewertungskriterien in Evaluationsprojekten, mit denen Sie zu tun haben, verwendet werden:
Gibt es solche, die immer wieder verwendet werden? Welches sind die wichtigsten?

Häufigste Codes ($\geq 10\%$) - insgesamt 31 Codes	Alle (n = 66)	Bildung (n = 15)	Gesundheit (n = 12)
Wirksamkeit/Effektivität	46%	7% (5)	75% (1)
Kosten-Nutzen/Effizienz	38%	13% (3)	67% (2)
Relevanz	20%	7% (5)	25% (6)
Jedes Mal andere Kriterien	20%	13% (3)	-
Kohärenz	17%	-	33% (3)
Nachhaltigkeit	15%	7% (5)	33% (3)
Zweckmässigkeit	14%	-	33% (3)
Zufriedenheit	11%	33% (1)	-
Zielerreichung	9%	7% (5)	17% (7)
Lehr-Kriterien	6%	27% (2)	-
Akzeptanz	3%	-	17% (7)
Anzahl Antworten mit Beispielen	74% (49)	67% (10)	92% (11)
Anzahl Beispiele pro Person	3.0	2.3	3.9

Welche Bewertungskriterien werden genutzt?

- **Effektivität** und **Effizienz** klar am häufigsten genutzt, nur von Personen im Bildungsbereich werden beide wenig genannt
 - **Relevanz** ist für Evaluierende wichtig, Kohärenz für Personen aus dem Gesundheitsbereich
 - **Nachhaltigkeit** und **Zweckmässigkeit** sind für Auftraggebende und Personen aus dem Gesundheitsbereich wichtig
 - **Jedes Mal andere Kriterien** sagen vor allem Auftraggebende
 - **Evaluierende** und **Personen aus dem Gesundheitsbereich** nennen **häufiger Beispiele** als andere
-
- Homogene Antworten: **Wenige Beispiele häufig genannt**
 - Ergebnisse zu Effektivität und Effizienz decken sich mit Heuer (2017)
 - Relativ hoher **Konsens zwischen den Akteur:innen**
 - **OECD DAC Kriterien** werden am häufigsten genannt (mit Ausnahme: Impact)

Fazit

- Akteur:innen der Evaluationscommunity haben grundlegendes Verständnis davon, was Bewertungskriterien sind
 - Diesbezüglich viel Konsens (z.B. Bewertung des Gegenstands, Operationalisierung), aber unterschiedliche Schwerpunktsetzung
 - Kriterien relativ homogen, aber teils starke Unterschiede zwischen Evaluierenden und Auftraggebenden
- Gemeinsames Verständnis von Bewertungskriterien schaffen und Bewertungskriterien bei jeder Auftragsklärung explizit festlegen
- Je nach Themenbereich/Disziplin ist dies mehr oder weniger wichtig

Grenzen der explorativen Studie

Selektion der Stichprobe:

1. Selektion: Teilnehmende der SEVAL Tagung zum Thema «Von der Bewertung zur Methode... und zurück»
2. Selektion: Personen ohne Wissen zu Bewertungskriterien haben vermutlich nicht teilgenommen

Grösse der Stichprobe

- Kleine Fallzahl
- Schwierig, vertiefte Analysen für verschiedene Akteursgruppen / Themenbereiche durchzuführen

Kontakt

Véronique Eicher: veronique.eicher@ehb.swiss

Lars Balzer: lars.balzer@ehb.swiss

Quellenangaben

- Balzer, L. (2005). *Wie werden Evaluationsprojekte erfolgreich? – Ein integrierender theoretischer Ansatz und eine empirische Studie zum Evaluationsprozess*. Landau: Verlag Empirische Pädagogik.
- Balzer, L. & Beywl, W. (2018). *Evaluiert*
- Beywl, W. (2006). Demokratie braucht wirkungsorientierte Evaluation. Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In Projekt eXe (Hg.), *Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis* (S. 25-46). München: Deutsches Jugendinstitut.
- Heuer, C. (2017). Verwendung von Bewertungskriterien in den externen Evaluationen der Bundesverwaltung: Ergebnisse einer Forschungsstudie. *LeGes - Gesetzgebung & Evaluation*, 28(2), 327–345.
- Scriven, M. (2007). *Key Evaluation Checklist* (von Daniela Schröter und Wolfgang Beywl ins Deutsche übersetzt und kommentiert). Michigan: Western Michigan University und Köln: Univation.
https://www.univation.org/download/Checklist_Key_Evaluation_ger.pdf (abgerufen am 13.09.2024)
- Stake, R.E. (1995). *The Art of Case Study Research*. Thousand Oaks: Sage Publications.
- Stake, R.E. & Schwandt, T. (2006). On Discerning Quality in Evaluation. In I.F. Shaw, J.C. Greene & M.M. Marks (Eds.), *The SAGE Handbook of Evaluation. Policies, Programs and Practices*. (pp. 404-418). London: Sage Publications.