

Wie kommt die Evaluation zu(m) Werten?

Transdisziplinäres Kolloquium „Zwischen Objektivität und Normativität: Wie kommt die Evaluation zu(m) Werten?“

Freitag, 27.02.2015, Justus-Liebig-Universität Gießen

Definitionen

Evaluation ist eine wissenschaftliche Dienstleistung, die insbesondere öffentlich verantwortete und/oder finanzierte Evaluationsgegenstände (Politiken, Programme, Projekte, Massnahmen oder Einrichtungen) systematisch, transparent und auf Daten gestützt beschreibt und ausgewogen bewertet, so dass Stakeholder (Auftraggebende etc.) die Ergebnisse für vorgesehene Zwecke wie Rechenschaftslegung, Entscheidungsfindung oder Verbesserung nutzen können.

Wert (eines Evaluationsgegenstandes). Summe der Eigenschaften eines Evaluationsgegenstands, die zu der Bewertung führt, dass er mehr oder weniger gut oder schlecht ist. Dabei kann die über Zeit und Raum relativ stabile Wertdimension der → Güte von der weniger stabilen der → Tauglichkeit unterschieden werden, die zwischen den Stakeholdern meist stark umstritten ist. Eine dritte Wertdimension ist die → Bedeutsamkeit. Die Bestimmung des Werts eines Evaluationsgegenstands, die Bewertung, ist die zentrale und unverzichtbare Aufgabe jeder Evaluation.

Werte – hier: **soziale** Werte – sind zunächst biographisch erworbene, kulturell basierte Dispositionen von Individuen, bestimmte Umstände/Lösungen/Handlungen anderen vorzuziehen. Werte beschreiben deskriptiv die tatsächlichen Dispositionen von Menschen; normativ benennen sie (soziale) Dispositionen, die – mit unterschiedlichen Begründungen – als verbindlich gelten oder eingefordert werden. Werte sind stärker auf längerfristiges Handeln bezogen und auch emotional verankert. Sie können mit kurzfristigeren, z. B. materiellen Interessen in Widerspruch geraten. Werte im sozialen Zusammenhang werden von gesellschaftlichen Milieus, Teilkulturen oder Gruppen getragen.

Wert, statistischer. Für die Auswertung mittels statistischer Verfahren vorbereitetes (quantitatives oder quantifiziertes) Datum: Werte (repräsentiert durch numerische und alphanumerische Zeichen) in Datenmatrizen (nach Konvention: Aussageeinheiten/Fälle in den Zeilen, Merkmale in den Spalten) ermöglichen die Verdichtung von Daten durch Auszählen der Häufigkeit ihres Vorkommens und – je nach statistischem Niveau – mächtigere mathematische Operationen gemäß den Konventionen der beschreibenden und der schließenden Statistik.

Kriterien sind in der Evaluation die Bezugsmaßstäbe, an denen bewertet wird. Etymologisch hergeleitet aus altgriechisch «krites» (der Richter) meint Kriterium (altgr. «kriterion»): Beurteilungsgesichtspunkt für den Richter. Kriterien basieren auf Werten. Nach Konkretisierungsgrad unterscheidbar sind allgemeinere Kriteriendimensionen (z. B. Lohngerechtigkeit) von operationalisierten Kriterien (z. B. summengleiche Bruttolohnhöhe bei Frauen und Männern mit gleicher Tätigkeit und gleicher Berufserfahrung). Für Bewertungsakte in der Evaluation sind in der Regel operationalisierte Kriterien erforderlich, für die zudem Kriterienpunkte etwa in Form von Prozentangaben im Vorhinein festgelegt werden.

Quelle: <http://www.eval-wiki.org>

Der Umgang von (Forschung und) Evaluation mit Werten

1 Forschung

Während das Gütekriterium der *Objektivität* – bestehend aus dem Dreiklang von Durchführungs-, Auswertungs- und Interpretationsobjektivität – in der psychologischen Forschung als weitgehend unproblematisch angesehen wird (Bortz/Döring 2006: 195), kann die quantitative Sozialforschung dieses Kriterium nur eingeschränkt einlösen: Objektivität bei der Interpretation sozialwissenschaftlicher Forschungsergebnisse (*Interpretationsobjektivität*) kann es nicht geben, da Interpretationen subjektiven Bewertungen (Werturteilen) unterliegen (können).

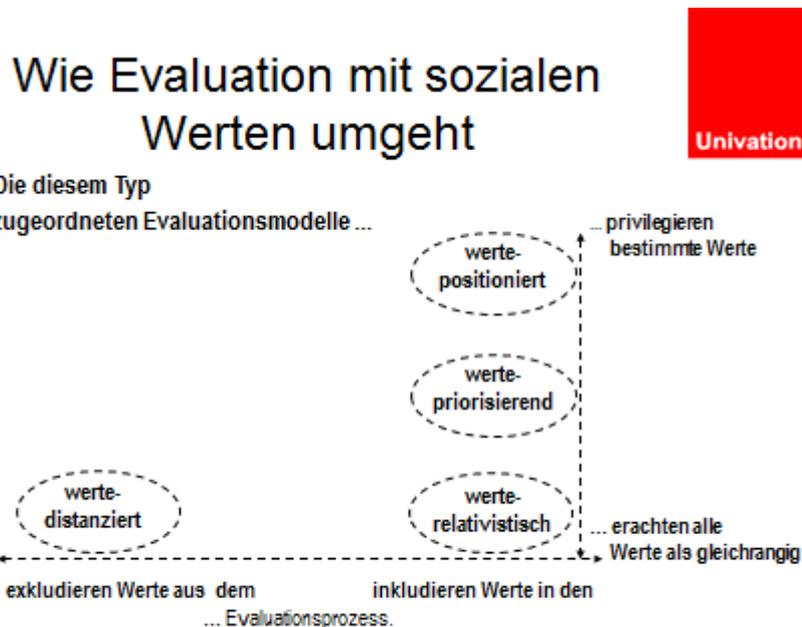
Durchführungsobjektivität kann weitgehend durch die *Standardisierung* von Fragebögen (alle Befragten erhalten identisch formulierte Fragen, also identische Stimuli) realisiert werden (Reinecke und Porst, Kapitel 44 und 50 in diesem Band). Allerdings des Fragebogens, beispielsweise Effekte der Reihenfolge von Fragen (Klößner/ Friedrichs, Kapitel 49 in diesem Band) und/oder Antwortvorgaben (Franzen, Kapitel 51 in diesem Band), oder Merkmale der Befragungssituation und/oder der Interviewer (...) die Durchführungsobjektivität einschränken. Diese Effekte können jedoch durch geeignete *a priori-Maßnahmen* vor der Befragung, z.B. durch Variation der Fragenreihenfolge sowie durch *ex post-Analyse* der Antworten in Abhängigkeit von Interviewermerkmalen, wenn nicht ausgeschaltet, so doch kontrolliert werden (Sudman et al. 1996; Tourangeau et al. 2000).

Die *Auswertungsobjektivität* (das ist die intersubjektive Nachvollziehbarkeit der – auf der Basis des vorliegenden Datenmaterials – durchgeführten Analysen) wird durch eine sorgfältige *Dokumentation der Datenaufbereitung* gewährleistet, bei der die Daten auf fehlerfreie Eingabe und Vollständigkeit überprüft werden. Außerdem ist zu dokumentieren, wie die Daten für *statistische Analysen* (...) *modifiziert* (...) wurden. So können Werte einer Variablen zusammengefasst oder in andere Werte umkodiert werden. Sorgfältig zu dokumentieren ist auch der *Umgang mit fehlenden Werten* (...).

Die Dokumentation der angewandten statistischen Verfahren sowie der Methoden von Skalierung (...) und Indexbildung (...) tragen ebenso zur Sicherstellung der Auswertungsobjektivität bei. So werden zur Erfassung einer latenten Variable (einer Einstellung) oft nicht einzelne sondern mehrere Indikatoren verwendet, die (additiv oder multiplikativ) miteinander verknüpft, d.h. zu einem Index zusammengefasst werden.

Quelle: Krebs, Dagmar/Menold, Natalja (2014): "Gütekriterien quantitativer Sozialforschung". In: Baur, Nina/Blasius, Jörg (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer Fachmedien Wiesbaden, S. 425-438. [http://dx.doi.org/10.1007/978-3-531-18939-0_30]

2 Evaluation



Beywl, Wolfgang (2006): "Evaluationsmodelle und der Stellenwert qualitativer Methoden". In: Flick, Uwe (Hg.): *Qualitative Evaluationsforschung. Konzepte - Methoden - Umsetzung*. Reinbek: Rowohlt, S. 92-116

Neun generische Kriteriendimensionen für Gegenstände der Bildungsevaluation

Kirkpatrick revisited

Themenbereiche	Beispiele für generische Fragestellungen
1. Bedarfsgerechtigkeit	In welchem Masse passen die (Lern-)Ziele und Inhalte der angebotenen Bildungsmassnahme auf die Bedarfe (Diskrepanzen zwischen den aktuellen und den zu schaffenden Kompetenzen) der Zielgruppen? Inwieweit werden Erwartungen der Zielgruppenmitglieder (z. B. zum Angebotsformat – Vollzeit vs. Teilzeit) und Ansprüche der Organisation getroffen?
2. Zufriedenheit*	Wie zufrieden sind die Teilnehmenden (und andere Kundengruppen) mit der Servicequalität des Angebots und den Produkten? In welchem Masse entsprechen diese den Dienstleistungsstandards der anbietenden Organisation?
3. Lernförderung	In welchem Masse unterstützen die didaktische Konzeption und ihre Umsetzung in Lehr-/Lernhandeln den fachlichen, methodischen und sozialen Kompetenzerwerb der Teilnehmenden, insbesondere das an differenzierten Lernzielen ausgerichtete selbstgesteuerte Lernen?
4. Lernerfolg*	In welchem Masse erreichen die Teilnehmenden die für das Lernfeld gesetzten Ziele, zeigen im (geschützten) Rahmen des Lernarrangements, dass sie über das geforderte Wissen verfügen und dass sie entsprechend handeln können?
5. Lernzuwachs	In welchem Masse haben Teilnehmende durch die Bildungsmassnahme dazu gelernt, insbesondere vorhandene Kompetenzen ausgebaut? Wie viel haben diejenigen, die bereits auf hohem Kompetenzniveau begonnen haben, dazu gelernt?
6. Transfererfolg /-nutzen*	Wie gelingen Transfer des Wissens und Anwendung des Könnens in der aktuellen und künftigen Bildungslaufbahn und Berufspraxis der Teilnehmenden? Welcher konkrete Nutzen entsteht z. B. am Ausbildungs- oder Arbeitsplatz? In welchem Masse werden Transferhemmnisse im Lernsetting antizipiert und bearbeitet? Welche bestehenden Transferchancen werden genutzt, welche bleiben ungenutzt?
7. Kosten-Nutzen- Angemessenheit*	Stehen die erzielten Erfolge auf den Ebenen Zufriedenheit, Lernförderung, Lernerfolg/-zuwachs, Transfererfolg/-nutzen und Systementwicklung in einer angemessenen Relation zu den eingesetzten finanziellen u. a. Ressourcen? Welche Maßnahmenalternativen sind bei gleichen Erträgen kostengünstiger?
8. Individuelle Entwicklung**	In welchem Maße entstehen für Teilnehmende neue Bildungs- und berufliche Chancen (weiterführende Abschlüsse, Berufseintritt, abschlussadäquater Stellen-, Arbeitgeber- oder Berufswechsel)? Inwieweit können sie Kontakte und Einflussmöglichkeiten in sozialen Beziehungsnetzen sichern und erweitern? Inwiefern gelingt Aneignung von Bildungsangeboten und Integration in eigene Sinnstrukturen?
9. Systementwicklung***	In welchem Masse trägt die Bildungsmassnahme dazu bei, dass ein lernförderliches Klima und das kollaborative Wissensmanagement in der Organisation gestärkt werden? Welche weiteren relevanten organisationalen Verfahrensweisen werden dadurch stabilisiert oder verbessert?

* Die Dimensionen 2, 4, 6 und 7 entsprechen den klassischen Kirkpatrick-Levels (Kirkpatrick & Kirkpatrick, 2006) 1, 2, 3 und 4 (vgl. dieses Buch, Kap. 6.1.2.2); 3 und 5 können Kirkpatricks 4. Level zugeordnet werden.

** Individuelle Entwicklung betont den Aspekt der eigensinnigen Bildung durch Aneignung gegenüber dem intentionalen Lernen, der Qualifizierung oder dem Kompetenzerwerb; den Hinweis darauf verdanken wir Katrin Kraus (2014).

*** Systementwicklung wurde als 5. Level den vier Kirkpatrick-Levels durch Guskey (2000, s. 80) hinzugefügt.

Quelle: Balzer, Lars/Beywl, Wolfgang (2015): *Bildungsevaluation*. Bern: h.e.p. (im Erscheinen)

Hauptquellen für wertebasierte Kriterien in Gleichstellungsevaluationen

In vielen Evaluationen sind Personen durch die Beschreibungen, und insbesondere durch die Bewertungen in ihren ideellen oder materiellen Interessen, in ihrer Bildungs- oder Berufs- oder Lebensplanung und in anderen Aspekten tangiert. «Um Widerstände gegen die Evaluation möglichst zu vermeiden, ist es wichtig, die kulturellen und sozialen Werte der involvierten Personen zu kennen und zu verstehen» (Widmer, 2005, S. 19). Dies gilt für Evaluationen im Gleichstellungsbereich mit besonderer Deutlichkeit, da die Beteiligten und Betroffenen sich in aller Regel eindeutig einem der beiden Geschlechter zuordnen, was eng mit ihrer persönlichen Identität und Integrität verbunden ist. Sie verfügen über – oft implizite – auf individuellen Biographien, Rollenerfahrungen und Lebensentwürfen basierende Vorstellungen von Gleichstellung oder Geschlechtergerechtigkeit. Diese können sich auf die Akzeptanz und Nutzung von Evaluationen auswirken. Aufgabe einer wertorientierten Evaluation ist es nicht, die individuellen Werthaltungen in Frage zu stellen, sondern diese zu klären und darauf aufbauend einen kriterienbasierten Bewertungsrahmen für die Evaluation zu formulieren.

Drei Ausgangspunkte

- a) die zentralen Werte der Gleichstellungspolitik, die jeweils konkretisiert werden. D. h., in Gleichstellungsevaluationen müssen **spezifisch auf Gleichstellung bezogene Wertvorstellungen** oder «Werteschemata» (Fraser, 2003, S. 36) beigezogen werden, um die Güte und Tauglichkeit eines Evaluationsgegenstandes zu beurteilen;
- b) **generische Werte**, die abgestützt auf evaluationstheoretische Überlegungen eingebracht werden, falls diese in den beiden anderen Quellen nicht vorhanden sind. So wird Evaluation immer die drei Kriteriendimensionen Güte, Tauglichkeit und Bedeutsamkeit (Scriven, 2007) in Erwägung ziehen;
- c) **lokal bzw. situativ fundierte Werte** der konkret am Evaluationsgegenstand Beteiligten und von ihm Betroffenen. Sie sind oft in den Konzeptionen oder Zielen der jeweiligen Gleichstellungsmassnahmen explizit oder implizit angesprochen.

zu a)

Die Gleichstellungs-Debatte kennt als zentrale Bezugspunkte die Werte «Gerechtigkeit» und «Anerkennung»:

- (Geschlechter-)Gerechtigkeit meint vielfach «Verteilungsgerechtigkeit» («Umverteilung», «redistribution»), welche die sozioökonomische Dimension betont. Nach Leitner (2007) kann sie «formal» oder «substantiell» definiert sein, sich auf «Prozess» oder «Ergebnis» beziehen. Formal sind «Gleichbehandlung» (Prozess, z. B. gleicher Lohn für gleiche Arbeit) und «Gleichberechtigung» (z. B. gleicher Lohn für gleichwertige Arbeit); substantiell sind «Chancengleichheit» (Prozess: z. B. «gleiche Zugangschancen für leitende Positionen») und «Gleichstellung» (Ergebnis: z. B. «gleiche Teilhabe an leitenden Positionen»).
- Anerkennung kann sowohl als Aspekt von Gerechtigkeit begriffen werden oder als eigenständige Wertedimension. Sie bezeichnet «[...] eine ideale reziproke Beziehung zwischen Subjekten, in der jeder den anderen als seinesgleichen und zugleich von sich getrennt sieht ...» und betrifft vordergründig wesentlich «Ziele der Selbstverwirklichung und des guten Lebens» (Fraser, 2003, S. 19). Das Statusmodell der Anerkennung ist darüber hinausgehend darauf gerichtet, «... diejenigen kulturellen Wertschemata ihrer kulturellen Geltung zu berauben, die gleiche Beteiligungschancen verhindern, um sie durch Muster zu ersetzen, welche solche Chancen befördern» (S. 46).

Quelle: Beywl Wolfgang/Ryter Elisabeth (2010): Wertorientierung – Schlüsselfaktor für Nützlichkeit von Evaluationen in der Gleichstellungspolitik. Full Proposal zum Nationalen Forschungsprogramm NFP 60 „Gleichstellung der Geschlechter“, Bern 2010 (nicht angenommen)

Vier generische Kriteriendimensionen für Evaluationen

1. **Nützlichkeit** (utility): „Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet.“
2. **Durchführbarkeit** (feasability): „Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.“
3. **Fairness** (propriety): „Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird.“
4. **Genauigkeit** (accuracy): „Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand