

## Evaluation

»Ihre Meinung ist gefragt!«, »Helfen Sie uns, unseren Service zu verbessern!«, »Bitte bewerten Sie die Lehrveranstaltung!«, »Wir wollen, dass Sie zufrieden sind!«, »Kritik erwünscht!« – Die Überschriften variieren, die Aufforderung ist immer die Gleiche: Wer heute in einem Hotel absteigt oder in einem Restaurant speist, wer ein Uni-Seminar belegt oder eine Fortbildung absolviert, wer eine Beratungsstelle aufsucht oder einen Pflegedienst in Anspruch nimmt, kann so gut wie sicher sein, am Ende einen Fragebogen in die Hand gedrückt zu bekommen. Es folgt die freundliche Bitte anzukreuzen (auf einer Skala von 1 »sehr zufrieden« bis 6 »total unzufrieden«), was missfallen und was gefallen hat, was im Argen liegt und was besser laufen könnte. Nur selten fehlt die Zusicherung »Ihre Angaben bleiben selbstverständlich anonym!«; manchmal soll ein Preisausschreiben die Bereitschaft zum Ausfüllen der Bögen wecken. Was mit den erhobenen Daten passiert, bleibt meist im Dunkeln; bisweilen werden die Zufriedenheitsbarometer am Schwarzen Brett ausgehängt oder können bei der Abteilung für Qualitätsentwicklung eingesehen werden.

Kunden- und Teilnehmerbefragungen sind nur das augenfälligste Symptom jener Epidemie, der böse Zungen den Namen Evaluationitis gegeben haben. Erfasst hat sie die bundesdeutsche Gesellschaft nach längerer Inkubationszeit irgendwann in den 1990er Jahren, ihr endgültiger Durchbruch fiel zusammen mit dem »PISA-Schock«. Inzwischen sehen manche Beobachter erste Immunisierungstendenzen. Dass das Zauberwort Evaluation etwas von seiner Strahlkraft eingebüßt hat und der Sport über die Auswertungsmanie wächst, besagt jedoch keineswegs, dass auch die tatsächliche Ausbreitung von Evaluationsverfahren zurückginge. Im Gegenteil: Vieles deutet darauf hin, dass die Praktiken der Aus- und Bewertung sich veralltäglicht und Evaluierer wie Evaluerte gelernt haben, das Evaluationsspiel zu spielen, ohne sich davon im gewohnten Gang der Geschäfte sonderlich aufhalten zu lassen.

Der Begriff Evaluation entstammt dem Französischen und wanderte während der Aufklärungszeit mit der Bedeutung »Schätzung« oder »Wertbestimmung« in die deutsche Sprache

ein. Da man ein Wort lateinischen Ursprungs vor sich zu haben glaubte, korrigierte man dabei die als »unlateinisch« empfundene Buchstabenfolge und machte aus der Evaluation eine Evalvation, die noch in Wörterbüchern der 1960er Jahre auftaucht. Auf dem Umweg über das Englische fand die Evalvation (oder gleichbedeutend: Evaluierung) im heutigen Sinne dann ab den 1970ern Eingang in die deutschsprachige sozialwissenschaftliche Literatur und diffundierte von dort aus in die Alltagssprache.

Evaluationsverfahren knüpfen an so verschiedene Praktiken an wie die kameralistische Rechenschaftslegung, Bonitätsprüfungen im Finanzgewerbe, das betriebliche Controlling oder Konsumentbefragungen im Rahmen der Marktforschung. Ihr Aufstieg als politisches Steuerungsinstrument fällt in die Zeit der Kennedy- und Johnson-Administration in den USA. Dort hatte die Welle sozialpolitischer Reformen auch den Bedarf an Begleitforschung explosionsartig anwachsen lassen. Ein Heer von Sozialwissenschaftlern und -wissenschaftlerinnen hatte sich aufgemacht, um die Effekte jener Programme zu überprüfen, mit denen die Behörden die Armut und Kriminalität zu bekämpfen, das Bildungssystem und Gesundheitssystem zu modernisieren und die Städte zu sanieren versuchten. Die Evaluationspezialisten dienten sich dabei an als »Wegweiser zu rationaler Sozialplanung«: »In Anbetracht der unbedingten Notwendigkeit, effektivere Programme zu entwickeln«, hieß es in einem 1974 ins Deutsche übersetzten Methodenhandbuch, »um die sozialen Bedürfnisse zu befriedigen, mit denen unsere Nation konfrontiert ist, gewinnt Evaluierungsforschung zunehmend an Bedeutung als Quelle gesicherten Wissens und begründeter Handlungsanweisungen« (Weiss 1974, 8f.). Mittels Evaluation sollten ganz im Sinne der damals wirkmächtigen Sozialkybernetik Rückkopplungsschleifen installiert werden, um die Interventionen des politischen Systems mit den gesellschaftlichen Erfordernissen abzustimmen und seine Steuerungskapazität zu erweitern.

Um Effizienzsteigerung und Feedback geht es noch immer, allerdings hat inzwischen der Optimierungsfuror der Qualitätsmanager die technokratische Hybris der Sozialplaner abgelöst. Konzentrierte Evaluation sich früher darauf, die erzielten Wirkungen an den gesteckten Zielen zu messen, Defizite zu identifizieren und Schritte zu ihrer Behebung aufzuzeigen, so dienen die fortwährenden Aus- und Bewertungsrituale jetzt vor allem dazu,



Leistungen zu vergleichen, den internen wie externen Wettbewerb zu stimulieren und so einen kontinuierlichen Verbesserungsprozess, kurz: KVP, in Gang zu setzen: Zuerst wird evaluiert; dann wird ein Ranking erstellt; anschließend werden die Werte der Spitzenreiter als Benchmarks festgelegt und das Feld verpflichtet, den »Besten der Besten« nachzueifern (Parole: »Einholen und überholen!«); das Ergebnis der entsprechenden Anstrengungen wird wiederum evaluiert; und so geht es immer weiter. Das Ganze gleicht einem sportlichen Ausscheidungskampf, bei dem in jeder Saison über Auf- und Abstieg neu entschieden wird. Bundesliga ist überall.

Im Unterschied zum Sport ist jedoch bei der Evaluation eines Hochschulinstituts oder einer Außendienstmitarbeiterin keinesfalls immer schon klar, was mit welchen Verfahren gemessen und welche Leistung wie gewichtet werden soll: eingeworbene Drittmittel, Studienabsolventen oder Anzahl der Veröffentlichungen? Umsatzsteigerung, neue Kundenkontakte oder Rückgang der Reklamationen? Leistung ist eine Definitions- und damit eine Machtfrage. Wer die Indikatoren festlegt, entscheidet, wenn auch indirekt, über die Ergebnisse. Die Evaluationsmacht legitimiert sich über ihre Objektivität: Sie weist nicht willkürlich Ränge zu, sondern gibt einheitliche Maßstäbe vor, nach denen alle beurteilt werden. Das zeitigt paradoxe Effekte: Weil die Position im Ranking weit reichende Folgen hat – Beförderung oder Entlassung, Aufstockung oder Kürzung des Budgets –, richten die Evaluatoren ihr Verhalten prospektiv auf die zu Grunde gelegten Kriterien hin aus. Man tut, was gemessen, und unterlässt, was vom Bewertungsraster nicht erfasst wird. Evaluation schafft so erst die Wirklichkeit, die sie zu bewerten vorgibt, und erzeugt statt der allseits beschworenen Innovationsfähigkeit »einen Aggregatzustand betrieblicher Konformität« (Koschorke 2004, 151).

Beispiel *citation index*: Diese nach Disziplinen geordnete Rangliste dokumentiert, wie häufig die Veröffentlichungen eines Wissenschaftlers in den einschlägigen, nach Einfluss gewichteten Fachorganen erwähnt werden, und gilt inzwischen als internationale Leitwährung wissenschaftlicher Reputation. Dass die Zitationsrate über die Besetzung von Lehrstühlen oder die Vergabe von Forschungsmitteln entscheidet, hat den akademischen Betrieb selbst nachhaltig verändert: »Wissenschaftler schreiben keine Monografien mehr, weil diese weniger »wert« sind. Forschung,

die Ergebnisse produziert, die besser in Bücher passen als in Aufsätze, unterbleibt. Koautorenschaften nehmen zu. Interessantes Material, das gut in einem Aufsatz Platz gefunden hätte, wird auf zwei oder drei Aufsätze ausgewalzt. [...] Nicht mehr die Aufhellung ungeklärter Phänomene und erst recht nicht die Lösung praktischer Probleme ist das zentrale Anliegen, im Vordergrund steht vielmehr die Herstellung von Aufsätzen, deren Chancen, in Zeitschriften mit hohem *impact factor* veröffentlicht zu werden, sehr gut sind« (Kieser 2003). Weil Evaluation ihre Maßstäbe festlegen muss, bevor es ans Messen geht, bleibt sie blind für das Neue. Das »Exzellenz«-Siegel erhalten jene, die dem *Mainstream* folgen. Wer gegen den Strom schwimmt, landet auf den hinteren Rängen. Der Leistungsvergleich stärkt gerade nicht die innovativen Kräfte, nach denen man angeblich doch händeringend sucht. Kreativität ist nicht evaluierbar.

Evaluationen brauchen Evaluationsexperten, die – freilich nach Maßgabe der jeweiligen Auftraggeber – Fragebögen entwerfen, Kenngrößen definieren und schließlich die erhobenen Daten durch ein Statistikprogramm jagen, um daraus eindrucksvolle Schaubilder und voluminöse Dokumentationen zu destillieren. Im gleichen Maße, wie die Praktiken des Messens und Vergleichens sich verfeinern, professionalisieren sich die Evaluierer. So hat der *citation index*, der im Übrigen von einem Privatunternehmen, dem *Institute of Scientific Information* (ISI), erstellt wird, den Beruf des Bibliometrikers hervorgebracht. Die Evaluation hat sich von einer Unterabteilung der empirischen Sozialforschung zu einer eigenständigen Disziplin mit zertifizierten Ausbildungen, Forschungseinrichtungen und Publikationsorganen gemauert – und zu einer Industrie, die der Beratungsbranche, was Umsatz und Mitarbeiter angeht, kaum nachstehen dürfte. Ohnehin sind beide eng miteinander verflochten: Zunächst kommen die Berater und diagnostizieren Evaluationsbedarf; wurde dann evaluiert, darf man sicher sein, dass die Ergebnisse wiederum Beratungsbedarf signalisieren. Oder umgekehrt.

Auch methodisch ist das Angebot kaum mehr zu überblicken. Da gibt es formative und summative, externe und interne, Fremd- und Selbstevaluationen, man unterscheidet Produkt-, Personal-, Prozess- und Projektevaluationen, und allein die staatlichen Instanzen können wählen zwischen Reform-, Maßnahmen-, Gesetzes- oder generellen Politikevaluationen. Nicht zu vergessen die

Meta-Evaluationen, mit denen die Auswertungsspezialisten die Qualität ihrer eigenen Arbeit überprüfen und sie mit anderen Evaluationsstudien vergleichen (vgl. Widmer 2000). Auch dafür existieren bereits Kriterienkataloge. So präsentieren die Qualitätsstandards der *Deutschen Gesellschaft für Evaluation* eine nach Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit gegliederte und aus 25 Einzelstandards zusammengesetzte Liste, die zugleich das fachliche Ethos der Disziplin formuliert. Die Anforderungen erstrecken sich nicht nur auf das methodisch angemessene Vorgehen bei der Datenerhebung, -auswertung und -dokumentation, sondern auch auf die Auswahl der beteiligten Personen: »Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.« Als genereller Orientierungspunkt gelten »die Gütekriterien quantitativer und qualitativer Sozialforschung« (DeGEval 2001).

Methodische Ansprüche wie Validität und Reliabilität, aber auch der Einsatz standardisierter Messinstrumente verbindet Evaluationsverfahren mit Tests. Wie diese etablieren sie ein Kontrollregime und folgen dabei dem demokratischen Prinzip des *one man, one questionnaire*. Gleich sind sich alle zumindest darin, dass sie mit dem gleichen Maß gemessen werden. Während jedoch Eignungs- oder Intelligenztests die Leistungsfähigkeit beziehungsweise -bereitschaft erheben und die Probanden dazu unter Laborbedingungen standardisierte Aufgaben ausführen lassen, bewerten Evaluationen die im »Normalbetrieb« tatsächlich erbrachten Leistungen. Wichtiger noch ist eine andere Differenz: Tests liefern keine Beurteilungen Dritter und sind deshalb anders als Evaluationen außer Stande, das zu messen, was den gemeinsamen Bezugspunkt aller Qualitätsdiskussionen markiert – die Kundenzufriedenheit. Wer einen Evaluations-Fragebogen ausfüllt, wird stets als Kunde angesprochen, gleich ob er eine Vorlesung besucht, Wohngeld beantragt oder sich einer Operation unterzogen hat. Er ist der König, dessen Meinung, wenn auch nur statistisch gemittelt, zählt. Diejenigen, die beurteilt werden, finden sich dagegen in der Rolle von Lieferanten wieder. Sie sollen mit ihren Produkten oder Dienstleistungen die Wünsche und Bedürfnisse derjenigen befriedigen, die ihnen genau darüber ein gutes oder schlechtes Zeugnis ausstellen. Negativbewertungen fun-

gieren dabei als Optimierungsmotor. Sie signalisieren Anpassungsbedarf und sollen ein flexibles Aussteuern ermöglichen, auf dass Angebot und Nachfrage immer besser aufeinander austariert werden. Die Rollen wechseln, und weil einer nicht nur des anderen Kunde und Lieferant, sondern auch sein Konkurrent ist, zeigen die Zufriedenheitsbarometer immer auch Marktpositionen an. Nicht »Sein oder Nichtsein«, sondern Bestseller oder Ladenhüter, »das ist hier die Frage«. Die Antwort liegt meist irgendwo dazwischen. Wo genau, das entnimmt der Hamlet von heute dem neuesten Ranking.

#### Literatur

- Deutsche Gesellschaft für Evaluation (DeGEval) 2001, *Standards für Evaluation*, [www.degeval.de/standards/standards.htm](http://www.degeval.de/standards/standards.htm) (28.1.2004).  
Kieser, Alfred 2003, »Forschung vom Fließband«, in: *Die Zeit*, Nr. 30, 21. Juli.  
Koschorke, Albrecht 2004, »Wissenschaftsbetrieb als Wissenschaftsvernichtung. Einführung in die Paradoxologie des deutschen Hochschulwesens«, in: Kimmich, Dorothee/Thumfart, Alexander (Hg.), *Universität ohne Zukunft?*, Frankfurt/M., 142–157.  
Power, Michael 1997, *The Audit Society. Rituals of Verification*, Oxford.  
Weiss, Carol H. 1974, *Evaluierungsforschung. Methoden zur Einschätzung von sozialen Reformprogrammen*, Opladen.  
Widmer, Thomas 2000, »Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird«, in: Stockmann, Reinhard (Hg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*, Opladen, 77–102.

Ulrich Bröckling