

Evidenz und Evaluation

Positionspapier der DeGEval – Gesellschaft für Evaluation¹

Die *DeGEval- Gesellschaft für Evaluation* hat auf ihrer Jahrestagung 2012 mit „Evidenz und Evaluation“ ein Thema diskutiert, das in Politik und Praxis aktuell sehr bedeutsam ist. Entscheidungen zur Einführung oder Umsetzung von politischen Programmen sollen wissenschaftsbasiert sein. Der Zusammenhang von Evidenz und Evaluation ist naheliegend, soll letztere doch wesentliche Informationen für die rationale Gestaltung und Steuerung von Programmen und Organisationen erzeugen. Entscheider und Durchführende von Evaluationen erwarten wissenschaftlich basierte Erkenntnisse, die das eigene Handeln absichern helfen. Anders als im Alltag, in dem ein Verweis auf „Evidentes“ häufig synonym für „selbstverständlich“ oder „offensichtlich“ steht, meint Evidenz im Kontext von Evaluation „Beleg“ oder „Beweis“, damit gut informierte – und möglichst robuste – Entscheidungen getroffen werden können.

Noch vor wenigen Jahren fanden Gespräche im Vorfeld von Evaluationen auf relativ einfachem Niveau statt: Wie erkläre ich den Evaluierten, was eine Evaluation ist? Wie erkläre ich den Auftraggebenden, was eine Evaluation leisten kann? Heute hingegen stellen Evaluatorinnen und Evaluatoren immer häufiger ein entwickeltes Verständnis und vertieftes Wissen über Evaluation fest. Eine Reife im Umgang mit Evaluationen ist inzwischen weit verbreitet.

In der Vergangenheit waren Evaluationen thematisch breit aufgestellt: Sie prüften grundlegend den Bedarf für Projekte oder Programme, bewerteten ihre Sinnhaftigkeit und Struktur, es gab ein starkes Interesse daran, Maßnahmen durch Evaluation zu begleiten und kontinuierlich in Zusammenarbeit mit Evaluierten zu entwickeln. Verbunden damit war ein großes Spektrum eingesetzter Methoden.

¹ Positionspapiere enthalten in knapper und allgemeinverständlicher Form Stellungnahmen zu aktuellen Themen der Evaluation und werden vom Vorstand der Gesellschaft für Evaluation ausgearbeitet.

Das breite thematische und methodische Spektrum von Evaluation wird angesichts eines zunehmenden (Kosten-)Drucks auf Durchführende und Initiatoren von politischen Programmen in dem Sinne reduziert, dass nunmehr von Evaluationen immer häufiger erwartet wird, Evidenzen – im Sinne von robusten Belegen – für die Wirksamkeit von Interventionen zu liefern. Das Eintreten von angestrebten Effekten soll möglichst sicher belegt sein.

Innerhalb der sozialwissenschaftlichen Methodendebatten dominiert aktuell die Vorstellung, dass Evidenzen für Wirkungen von Programmen lediglich durch den methodischen „Goldstandard“ garantiert werden können: Wirkungen können – etwas vereinfacht ausgedrückt – demnach nur bewiesen werden, wenn ein systematischer Vergleich von wenigstens zwei statistisch identischen Gruppen, von denen die erste eine „Intervention“ erhält und die zweite nicht, eine signifikante Differenz ergibt. Dazu muss die Intervention kontrolliert sein, sie muss also in einer exakt vorgeschriebenen Art und Weise durchgeführt werden und demnach wiederholbar sein.

Ein solches randomisiertes und kontrolliertes Experiment ist allerdings aufgrund einer Vielzahl von Gründen in sozialen Kontexten nur schwer umsetzbar: In sozialen oder politischen Programmen lassen sich die den Erfolg beeinflussenden Variablen kaum isolieren, Programme sind in der Regel kaum in mechanischem Sinn durchführbar, vielen Programmen fehlt ein logisches Wirkmodell, Kontrollgruppen sind aus ethischen und praktischen Gründen kaum festzulegen, quantitative Studien benötigen große Stichproben.

Darüber hinaus ist auch zu fragen, was quantitative Effektstärken von Programmen inhaltlich aussagen. Der Blick auf Effekte verstellt – insbesondere bei unterfinanzierten Evaluationen – zudem allzu leicht den Blick auf nicht beabsichtigte oder sogar negative Effekte von Programmen.

Die Erwartung von Wirkungsbelegen ist gut nachvollziehbar. Wirkungsevaluationen waren auch in der Vergangenheit immer wesentliches Element der Arbeit von Evaluatoren. Aber die Dominanz der Forderung nach Evidenzen für Wirkungen übersieht, dass die Qualität eines Programms nicht zureichend durch Effektmessungen zu bewerten und vor allem nicht weiter zu entwickeln ist. Mit dieser Verschiebung bzw. Verengung ändert sich zunehmend die Funktion von Evaluation: sie dient weniger der Optimierung, sondern mehr der Legitimation eines Programms bzw. einer Maßnahme.

Ein Blick aus anderer Perspektive: Für eine Evaluation, die Evidenzen belegt hat, ist es enttäuschend, wenn Entscheider oder Praktiker dann nicht den

Belegen entsprechend handeln. Evaluierende müssen verstehen lernen, dass Entscheidungen durch Evidenzen zwar fundiert werden können, sie aber immer auch von anderen Variablen mitbestimmt werden. Evidenzen determinieren nicht bestimmte Entscheidungen. Handeln ist immer auch an Kontexten und Normen orientiert, die jenseits von Beweisen liegen. Entscheidungen können zwar durch Evidenzen informiert sein, sich aber dennoch an anderen Maßstäben als gemessene Belege orientieren. Handeln in politischen, sozialen oder pädagogischen Kontexten ist nicht auf „Vermessbares“ im Sinne einer normativ neutralen Währung reduzierbar.

Im Austausch zwischen Evaluierenden und Auftraggebenden sollte die – plausible – Forderung nach Erzeugung von Evidenzen die Schwierigkeiten im Blick behalten, die mit einem solchen Anspruch verbunden sind. Evidenzen sollen auch dazu dienen, Programme zu verbessern, ihren Nutzen zu steigern und die Menschen zu stärken –, nicht nur messbare Effekte zu belegen. Hierzu sind Rückkopplungs-Schleifen und deren Auswertung im Sinne von strukturierten Lernprozessen nötig. Auch erfordert ein solcher Ansatz multimethodische Zugriffe. Das geht nicht ohne einen angemessenen Aufwand und damit verbundene Kosten.

Vorstellungen von schlichten und unproblematischen Wirkungsevaluationen sind unrealistisch und reduzieren das Potenzial von Evaluationen. Der Hauptredner der DeGEval-Konferenz 2012, Professor Geert Biesta (nunmehr Universität Luxemburg), machte deutlich, dass die Mächtigkeit der Idee der Evidenz die Tendenz erzeugt, dass Entscheidern nur noch als bedeutend erscheint, was in seinen Wirkungen messbar ist. Stattdessen sollte umgekehrt gefragt werden, was uns – politisch, gesellschaftlich, pädagogisch, ökologisch – wichtig ist. Ob das Wichtige messbar ist, ist dann erst die zweite Frage.