

(veröffentlicht in: H. Holz, D. Schemme (Hg.): Wissenschaftliche Begleitung bei der Neugestaltung des Lernens. Innovation fördern, Transfer sichern, Bielefeld 2005: Bertelsmann, 64-71 (ISBN 3-7639-1047-6))

Beraten – Gestalten – Implementieren – Evaluieren – Forschen

Helmut Kromrey

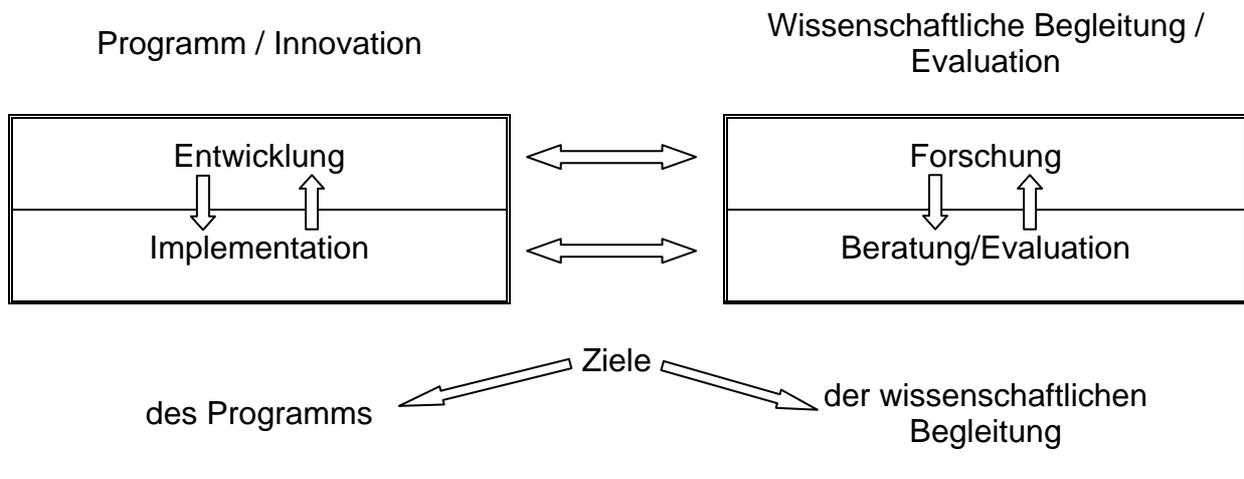
Das „Evaluieren“ (von Planungen, von Programmen, von Organisationen) ist zu einem gängigen Bestandteil von Politik und Praxis geworden. Allerdings kann dieses „Evaluieren“ unter Rückgriff auf verschiedenartige Konzepte und auf unterschiedlichste Weise mit unterschiedlichen Zielen von unterschiedlichen Instanzen vorgenommen werden. Ein Austausch von Meinungen und Erfahrungen zwischen allen Gruppen von Beteiligten und Betroffenen – wie auf dieser Tagung beabsichtigt – ist also ein nützliches Vorhaben.

Leider klingt dies einfacher als es ist. Denn zum einen ist der Begriff „Evaluation“ zu einem gern und häufig benutzten Modewort geworden, das in unterschiedlichen Kontexten in verschiedensten Variationen verwendet wird. Zum anderen sind die verschiedenen Konzepte des Evaluierens verbunden mit extrem unterschiedlichen Verfahrenslogiken und Methoden und eignen sich für jeweils unterschiedliche bis gegensätzliche Zwecke und Ziele. Es existiert also – anders formuliert – weder *die* Evaluation noch *die* Methodik des Evaluierens.

In dieser Situation ist es geradezu eine Bedingung für die Möglichkeit des Erfolgs jeden Erfahrungs- und Meinungsaustausches, eine Art „Propädeutikum Begriffsklärung“ vor- oder zumindest zwischenzuschalten. Die unverbundene Aneinanderreihung von Begriffen im Titel meines Beitrags steht für diese Absicht.

Nun ist zwar im Tagungsprogramm von *Evaluation* gar nicht die Rede, sondern von „Aufgaben wissenschaftlicher Begleitung“ im Rahmen der Ziele „Innovationen fördern, Transfer sichern“. Doch wie in anderen Kontexten, so spitzte sich auch hier am ersten Tagungsnachmittag in der Diskussion die „wissenschaftliche Begleitung“ sogleich auf „Begleitforschung“ zu, und diese wiederum wurde reduziert auf das Erkenntnisinteresse „Evaluation“. Vor einer solchen Verengung des Blickwinkels muss jedoch gewarnt werden. Der Gegenstand, mit dem wir uns bei der Aufgabe „Innovationen fördern, Transfer sichern“ konfrontiert sehen, ist viel zu komplex, als dass a priori eine solche Engführung zulässig wäre. Ich möchte das an folgender Skizze illustrieren:

Wissenschaftliche Begleitung bei der Programmentwicklung



In dieser Skizze sind sowohl die Kästchen wie die Pfeile in unterschiedlicher Weise empirisch interpretierbar.

So können einerseits beide Seiten der Grafik für *verschiedene Funktionen* stehen, die in einem Modellvorhaben mit wissenschaftlicher Begleitung zu erfüllen sind, *ohne* dass damit eine *personelle Trennung* impliziert sein muss:

- Es könnte sich also um ein Projekt mit (arbeitsteilig) selbst-organisierter Forschung handeln (Konzept F&E);
- die gleiche Darstellung könnte aber auch für ein Projekt stehen, das nach dem Design der Handlungsforschung verfährt und in dem alle Mitwirkenden an beiden Funktionen verantwortlich mitbeteiligt sind.

Anders sieht das empirische Feld aus, wenn die beiden Seiten der Skizze für *verschiedene Personengruppen* stehen, die jede mit unterschiedlichen Verantwortlichkeiten zum Gelingen des Gesamtprojekts beitragen wollen. Auch hier sind wieder verschiedene Konstellationen denkbar:

- z.B. als ein Entwicklungsprojekt mit partizipativer Beteiligung der Forschung oder
- als ein Entwicklungsprojekt mit wissenschaftlicher Beratung oder auch
- als ein Entwicklungsprojekt mit unabhängig davon durchgeführter Begleitforschung, die das Modellvorhaben im Auftrag einer externen Institution dokumentierend begleitet.

Wissenschaftliche Begleitung würde in jedem dieser Fälle *auch* empirische Informationserhebung (= Forschung) bedeuten; sie *kann* (muss aber nicht) zugleich Evaluation beinhalten. Selbst wenn sie Evaluation beinhaltet, muss damit nicht impliziert sein, dass „das Evaluieren“ durch die Begleitforschung vorgenommen wird. Dieses kann auch in die Zuständigkeit einer anderen – projektinternen oder –externen – Instanz fallen (etwa die Lenkungsgruppe des Entwicklungsprojekts oder ein externer Projektbeirat). Die Begleitforschung würde in diesem Fall durch Sammlung und Aufbereitung bewertungsrelevanter Informationen der Evaluation „lediglich“ zuarbeiten.

Man sieht, es handelt sich um ein Puzzle, das in vielfältiger Weise sinnvoll zusammengesetzt werden kann: die einzelnen Teile passen problemlos an unterschiedliche Stellen und konstituieren je nach Zusammenordnung unterschiedliche Bilder.

Um den vorgegebenen Rahmen des Beitrags nicht zu sprengen, beschränkte ich mich bei dem Versuch einer semantischen Klärung im Folgenden auf das Konzept „Evaluation“ und stelle die simpel erscheinende Frage: Was *ist* das eigentlich? Die Antwort fällt allerdings gar nicht so leicht.

Schon beim ersten Anlauf zu einer Antwort fällt auf, dass wir es hier mit einem geradezu – zumindest für nicht alltagssprachliche Zwecke – „schlampigen“ Begriff zu tun haben, der eigentlich in eine wissenschaftliche Terminologie keinen Zugang hätte finden dürfen, denn er steht gleichzeitig für verschiedene Arten von Referenzobjekten:

- Er bezieht sich zum einen auf eine „symbolische“ Ebene: als aus dem Amerikanischen übernommene *Bezeichnung* für Bewertung / Bewerten sowie für Denkmodelle bzw. Konzepte systematischer Bewertungsverfahren. Wenn wir uns über Methoden oder Standards der Evaluation unterhalten, beziehen wir uns auf diese Ebene.
- Der Begriff kann aber auch auf ein spezifisches *empirisches Handeln*, auf einen Prozess zielorientierten Informationsmanagements gemünzt sein: auf eine (methodisch kontrollierte) verwertungs- und bewertungsorientierte Form des Sammelns und Auswertens von Informationen. Wenn von Begleitforschung die Rede ist, dann steht dieser Aspekt im Vordergrund.
- Und schließlich kann damit auch das *Resultat* dieses Prozesses, ein Werturteil über das Zielobjekt des Evaluierens (das „Evaluationsprodukt“) angesprochen werden. Dies ist der Fall, wenn wir eine Aussage hören, das Programm X oder das Institut Y sei „positiv“ oder „negativ“ evaluiert worden.

Dieser Mangel an begrifflicher Präzision mag vielleicht harmlos klingen, ist aber verantwortlich für die Ergebnislosigkeit mancher Diskussionen, in denen die Beteiligten nicht immer bemerken, dass sie wechselseitig über völlig Unterschiedliches und damit aneinander „vorbei“ reden. Man stelle sich einmal vor, in der Praxis des Ingenieur- oder des Bauwesens würde mit ähnlich unpräzisen Begriffen gearbeitet und es wäre beispielsweise eine sprachliche Abgrenzung zwischen dem Bauplan des Architekten (Modell/Konzept), der Durchführung der Baumaßnahme durch Unternehmer und Bauarbeiter (Handeln/Prozess) sowie dem fertig gestellten Ergebnis, dem Gebäude (Produkt) nicht vorgesehen. In einer Diskussion über Baumängel wäre dann ohne Nachfragen nicht erkennbar, ob eine Aussage sich auf die statischen Berechnungen (Modell), den Einsatz minderwertiger Baumaterialien bzw. auf fehlerhafte Arbeit der Maurer (Prozess) oder auf die Haltbarkeit der Kellerisolierung bei steigendem Grundwasserstand (Resultat) bezieht. Bei der „Evaluation“ leisten wir uns diesen Luxus mangelnder sprachlicher Präzision.

Die mangelnde Präzision reicht aber noch einen Schritt weiter: Nicht nur bezieht sich der Begriff „Evaluation“ auf unterschiedliche Arten von Referenzobjekten, sondern seine Bedeutung variiert zudem in verschiedenen Verwendungskontexten erheblich:

- (1) *Alltagssprachlich* handelt es sich lediglich um ein „wohlklingendes“ und deshalb gern benutztes Fremdwort für jegliche Art von Bewertung: Irgendjemand bewertet irgendetwas irgendwie unter irgendwelchen Gesichtspunkten.
- (2) Im öffentlichen, im weitesten Sinne „*politischen*“ *Sprachgebrauch* ist damit die Überprüfung eines Programms oder einer Einrichtung durch ein zu diesem Zweck berufenes Expertengremium gemeint, das informationsgestützt und auf der Grundlage seiner Expertise ein Evaluationsgutachten erstellt, ggf. ergänzt um Empfehlungen für die Einstellung oder Fortführung, die Weiterentwicklung oder Optimierung des bewerteten Vorhabens.

- (3) Im *wissenschaftlichen Kontext*, insbesondere unter empirisch-methodischer Perspektive, verbirgt sich hinter „Evaluation“ das Design für einen spezifischen Typ von Sozialforschung, im Idealfall realisiert als experimentelle oder quasi-experimentelle Wirkungsforschung. Subsumiert werden darunter aber auch die Ansätze der Implementations- und der Begleitforschung, der empirischen Erfolgskontrolle sowie der formativen empirisch-wissenschaftlichen Beratung, die sämtlich nach je unterschiedlicher Logik und Methodik verfahren.
- (4) Und schließlich wird ärgerlicherweise auch noch „ganz gewöhnliche“ *Umfrageforschung* mit Evaluation gleichgesetzt, wenn sie – analog zur Meinungsforschung – bewertende (also evaluierende) Aussagen der Befragten erhebt und aufbereitet. Zielgruppen solcher Evaluationsbefragungen sind Personen, die in besonderer Weise als geeignet angesehen werden können, einen zu bewertenden Sachverhalt oder Gegenstand zu beurteilen: etwa Teilnehmer/innen von Lehrveranstaltungen als Evaluatoren der Qualität der Lehre, Klienten einer sozialen Dienstleistung als Beurteiler der Dienstleistungsqualität, Wahlberechtigte als Begutachter von Politikern (s. „Politbarometer“). Genau besehen handelt es sich hierbei allerdings nicht um methodisch kontrollierte Bewertungen mit dem Anspruch intersubjektiver Geltung (wie bei den vorgenannten Typen 2 und 3), sondern entweder um die Erhebung und Aufbereitung subjektiver „Alltagsevaluationen“ (deren zugrunde liegende Kriterien und Standards nicht rekonstruierbar sind) oder um Zufriedenheitsmessung (falls anstelle expliziter Bewertungen zu präzise benannten Sachverhalten um sog. „Alles-in-allem“-Urteile gebeten oder ausdrücklich nach der „Zufriedenheit mit ...“ gefragt wird).

Methodisch kontrollierte Evaluation unterscheidet sich von Alltagsevaluation dadurch, dass nicht „irgend etwas“ beurteilt wird, sondern dass genau benannte und empirisch beschreibbare Sachverhalte (Programme, Maßnahmen, Organisationen) *und* auf diese gerichtete präzise und operationalisierbare Fragestellungen ihr „Gegenstand“ sind.

Beurteilt wird nicht von „irgend jemandem“, sondern durch Personen, die zur Bewertung besonders befähigt erscheinen (das *können* die beauftragten Forscher oder Experten sein oder Beteiligte und/oder Betroffene oder zum Fällen der Urteile speziell legitimierte Gremien wie der intern/extern besetzte Beirat eines Projekts oder ein Fachausschuss in der Kommunalverwaltung oder die Lehrkommission in einem universitären Fachbereich).

Evaluert wird auch nicht „in irgend einer Weise“ und „nach irgend welchen Kriterien“, sondern in einem objektivierten Verfahren auf der Basis empirisch abgesicherter Informationen sowie nach explizit auf den Sachverhalt bezogenen und begründeten Beurteilungskriterien und –maßstäben.

Versuch einer Konkretisierung: Begleitforschung und Evaluation

Die damit aufgeworfenen Fragen möchte ich abschließend für den durch das Tagungsthema bezeichneten Gegenstandsbereich „Förderung von Innovationen“ (durch Modellversuche) und „Sicherung von Transfer“ etwas konkreter stellen und versuchsweise beantworten.

Die erste Frage muss natürlich sein: *Was soll evaluiert werden?* Aufgabe einer wissenschaftlichen Begleitung von Modellvorhaben könnte die Analyse von *Planung, Implementation und Vollzug des Vorhabens* sein mit dem Ziel, die Beurteilung des Projektmanagements, der Effektivität der Arbeitsprozesse, der Effizienz des Ressourceneinsatzes etc. zu ermöglichen. Gegenstand der Evaluation könnte auch das im Pilotprojekt zu entwickelnde *Produkt* sein: eine „transferfähige“ Innovation. In diesem Fall wäre Maßstab der Güte des zu begleitenden Projekts eine hypothetische (nicht durch empirische Daten zu

messende) Eigenschaft eines noch nicht existierenden, sondern erst im Projektverlauf zu entwickelnden Produkts, nämlich dessen Qualitätsmerkmal „Transferpotential“. Und falls die Forderung „Transfer sichern“ ernst genommen werden soll, wäre der „Gegenstand“ der Evaluation sogar noch weiter zu fassen: Nicht lediglich die *Eignung* zum Transfer, sondern der *Grad des realisierten Transfers* müsste als Erfolgskriterium herangezogen werden. Allerdings wäre dies insofern ein unfairer Beurteilungsmaßstab, als nicht der tatsächliche Transfer, sondern die Entwicklung einer *transferfähigen* Innovation das Ziel eines Pilotvorhabens ist.

Die bisherigen Überlegungen werfen zwangsläufig zwei weitere Fragen auf: *Wann wird evaluiert? Und wie kann dies geschehen?* Für den *Projektprozess* als „Gegenstand“ sind diese Fragen relativ einfach und eindeutig beantwortbar. Die empirische Informationserhebung kann projektbegleitend in Realzeit erfolgen und ist methodisch einlösbar über ein möglichst vollständiges „monitoring“ aller relevanten Daten. Die „Evaluation“ kann in zweierlei Weise erfolgen: zum einen „formativ“ als beratende Rückkoppelung in das Entwicklungsvorhaben, zum anderen „summativ“, indem zum Ende des Projekts die begleitend erhobenen und dokumentierten Informationen zusammenfassend ausgewertet werden. Der Nutzen der formativen Verfahrenslogik besteht darin, einen Beitrag zur Optimierung von Projektablauf und –produkt zu leisten. Die summative Evaluationslogik erfüllt dagegen Funktionen der Rechenschaftslegung und/oder der Erfolgskontrolle und/oder der Generierung von Handlungswissen, das für künftige Projekte nutzbar gemacht werden könnte. Weniger leicht sind die Fragen im Hinblick auf die Beurteilung der *Produktqualität* (Kriterium in unserem Fall: Transferfähigkeit) zu beantworten. Die empirischen Daten über das Projektgeschehen bieten dafür keine geeignete Beurteilungsgrundlage; und auch die Beschreibung „objektiver“ Eigenschaften der Zwischenprodukte sowie des Endprodukts sagen für sich genommen wenig darüber aus, ob die in Entwicklung befindliche Innovation später „am Markt“ (bzw. „in der Gesellschaft“) auf Akzeptanz stoßen wird und wie sie sich gegenüber vorhandenen oder erst noch entstehenden Alternativen behaupten können. Hier könnten allenfalls empirische Informationen vom Typ Marktforschungsdaten weiterhelfen; diese sind allerdings nicht durch die wissenschaftliche Begleitung des Projektverlaufs zu gewinnen, sondern müssten ergänzend erhoben werden. Als Alternative bietet sich an, das Transferpotential durch ausgewählte Experten einschätzen zu lassen (übliche Methoden: Delphi-Befragung oder Experten-Workshop in der Schlussphase des Projekts). Die *Beurteilung des tatsächlichen Transfers* des Produkts schließlich wirft methodisch zwar keine besonderen Probleme auf; sie erfordert allerdings eine Follow-up-Studie in einem angemessenen Zeitraum nach Abschluss des Entwicklungsprojekts.

Zwei weitere, für die Projektorganisation zentrale Fragen lauten: *Wo ist die Evaluation / die Begleitforschung angesiedelt? Und: Wie ist das Begleitforschungs- bzw. Evaluationsprojekt in das Entwicklungsprojekt / das Modellvorhaben eingebunden?* Einige Konstellationen und Zuständigkeitszuschreibungen zwischen Entwicklungsprojekt und Begleitforschung / Evaluation habe ich zu Beginn in den Erläuterungen zur Skizze „Wissenschaftliche Begleitung bei der Programmentwicklung“ bereits angeführt. Zugunsten der Vergabe an Programm-Externe spricht, dass die Begleitforscher nach Kriterien der Professionalität und Unabhängigkeit ausgewählt werden können. Probleme können allerdings entstehen, wenn die Kompetenzen der beteiligten Parteien nicht hinreichend präzise und verbindlich vor Beginn der Projektarbeit festgeschrieben wurden und wenn die Ziele von Entwicklungsprojekt und Begleitforschung / Evaluation nicht für alle Beteiligten hinreichend transparent sind. Für den Einbezug der Begleitforschung / Evaluation als Bestandteil der Projektentwicklung (bis hin zu personeller Identität) spricht, dass der damit sichergestellte Zugang zu allen projektrelevanten Informationen die Datenerhebung erleichtert und dass sich der Transfer der aufbereiteten empirischen Informationen in die Projektarbeit problemlos gestaltet. Allerdings dürfte von

Außenstehenden die „Objektivität“ der Begleitforschungsaussagen (insbesondere von „positiven“ Evaluationsurteilen) angezweifelt werden, und zwar um so eher, je stärker die Begleitung auch personell in die Projektarbeit eingebunden ist.

Schließlich muss noch die wichtigste Frage beantwortet werden: *Wer beurteilt nach welchen Kriterien?* Anders ausgedrückt: *Wer ist die Evaluierungs-Instanz, und im Hinblick auf welche / wessen Ziele wird die Bewertung vorgenommen?* Im Evaluationsdesign der empirischen *Programmwirkungsforschung* fällt die Antwort leicht. Die Beurteilungskriterien ergeben sich aus den operationalisierten Zielen des zu evaluierenden Programms. Die ermittelten Zielerreichungsgrade gelten als die Erfolgswerte des Programms. Die Evaluierung besteht darin, die gemessenen Grade der Zielerreichung mit den vom Programm angestrebten Sollwerten zu vergleichen. Sie ist somit kein eigentliches „Werturteil“, sondern eine technologische Vergleichsaussage¹. Forschung und Evaluierung fallen in diesem Konzept definitionsgemäß zusammen. Auch für die *Begleitforschung* von Modellvorhaben gilt: Jedes Evaluieren ist ein zielbezogenes Bewerten. Allerdings existieren im Allgemeinen noch keine so präzisen Zielvorstellungen, dass sie bis in operationalisierbare Zielerreichungsgrade „übersetzbar“ wären. Was sich als innovatives Produkt des Entwicklungsprozesses herauskristallieren wird, ist so konkret zu Beginn des Projekts noch nicht vorhersehbar. Außerdem kann es durchaus strittig sein, ob die Ideenlieferanten eines Entwicklungsprojekts, aufgrund deren Antrag ein Modellvorhaben gefördert wird, wirklich die legitime Instanz sind, durch Beschreibung der beabsichtigten Innovation zugleich die Wertbasis für die Evaluation ihres eigenen Projekts festzulegen. Mindestens ebenso plausibel dürfte es sein, die Ziele der potentiellen Nutzer der Innovation oder der von ihr Betroffenen zum Bezugspunkt zu nehmen. In jedem Fall aber besteht expliziter Entscheidungsbedarf. Es muss zu jedem Zeitpunkt eindeutig geklärt sein oder geklärt werden, was jeweils das zu Erreichende ist. Dies ist eine notwendige Voraussetzung nicht nur jeder Evaluierung, die nicht lediglich individuell subjektive, sondern intersubjektive Geltung beanspruchen will. Es ist auch Bedingung für ein Erfolg versprechendes Projektmanagement („by objectives“). Bei der Präzisierung – und je nach Projektfortschritt auch Fortschreibung – der Ziele sowie der Entscheidung über die Festlegung geeigneter Messgrößen (Zielerreichungs-Indikatoren) kann und muss die wissenschaftliche Begleitung mitwirken, sie kann und darf aber diese Entscheidung nicht stellvertretend für das Projekt treffen. Sobald diese Voraussetzungen gegeben sind, *kann* die Begleitforschung – ähnlich wie die Programmwirkungsforschung – auch die Aufgabe des Evaluierens übernehmen. Es kann aber auch in diesem Fall pragmatisch geboten sein, das eigentliche Bewerten der erreichten Zwischen- und Abschlussresultate an eine nicht in den laufenden Entwicklungsprozess eingebundene Instanz zu delegieren (Experten, Nutzer, Betroffene, durch legitimierte Vertreter der Beteiligten und Betroffenen gebildeter Projektbeirat).

Um abschließend noch einmal auf die beobachtete Engführung des Blickwinkels am ersten Tagungsnachmittag (s.o.) zurückzukommen: wissenschaftliche Begleitung > *Begleitforschung* > Evaluation. Wissenschaftliche Begleitung muss sich *nicht* auf *Begleitforschung* *beschränken*, sie sollte zumindest auch wissenschaftliche Beratung umfassen, sie sollte die Interessen und Perspektiven der Betroffenen und Beteiligten einbringen, sie kann die Rolle des Dolmetschers zwischen unterschiedlichen (Alltags-)Welten spielen, die im Modellvorhaben aufeinander treffen. Wissenschaftliche Begleitung muss aber auf jeden Fall *auch* *Begleitforschung* *beinhalten*. Sie muss methodisch gut begründete empirische Informationen sammeln und rechtzeitig und zielgenau bereitstellen, damit auf ihrer Basis evaluiert werden

¹ Das eigentliche „Werturteil“ wurde bereits bei der Entscheidung über die zu erreichenden Zielgrößen gefällt.

kann. Diese Informationen müssen den traditionellen Gütekriterien empirischer (standardisierter und/oder qualitativer) Forschung gerecht werden. Begleitforschung *muss* aber *nicht* zugleich auch Evaluation sein, sie muss nicht die Aufgabe des Evaluierens haben. In manchen Fällen wird ihr Beitrag zur Qualitätsentwicklung des Modellversuchs-Produkts höher sein, wenn sie sich auf die Aufgaben Datenbeschaffung und Informationsmanagement sowie erläuternde und engagierte Beratung beschränkt und die Funktion des Evaluierens an eine von der Forschung unabhängige Instanz delegiert wird. Gerade unter dem Gesichtspunkt der intersubjektiven Geltungsbegründung von Evaluationsbefunden und der Sicherung ihrer Akzeptanz kann eine solche Trennung sinnvoll sein.

Lektürehinweis. Wesentliche Aspekte dieses Beitrags werden ausführlicher abgehandelt in:

KROMREY, Helmut: Evaluation - ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: SuB Sozialwissenschaften und Berufspraxis, Jg. 24 (2001), Heft 2, S. 105-132

Autor:

Prof. Dr. Helmut Kromrey, Kufsteiner Str. 12, 10825Berlin,
Tel. 030 8536070 / eMail mail@hkromrey.de / URL: www.profkromrey.de