

Heft 59

Was ist Evaluation ?

Dr. Wolfgang Meyer
Saarland University
Centrum für Evaluationsforschung - CEval

Gabriela Höhns
Bundesinstitut für Berufsbildung - BIBB

Die WISSENSCHAFTLICHEN DISKUSIONSPAPIERE des Bundesinstituts für Berufsbildung (BIBB) werden durch den Generalsekretär herausgegeben. Sie erscheinen als Namensbeiträge ihrer Verfasser und geben deren Meinung und nicht unbedingt die des Herausgebers wieder. Sie sind urheberrechtlich geschützt. Ihre Veröffentlichung dient der Diskussion mit der Fachöffentlichkeit.

Autoren:

Dr. Wolfgang Meyer,
Center for Evaluation Research (CEval)
Saarland University - FR 5.2
Box 151 150
D-66041 Saarbrücken
F. R. Germany
Phone: ++49-(0)681-302-3320
Fax: ++49-(0)681-302-3899
e-mail: w.meyer@rz.uni-sb.de

Gabriela Höhns
Bundesinstitut für Berufsbildung
D-53043 Bonn
Bundesrepublik Deutschland
Tel.: ++49 -(0)228-107-1931
Fax: ++49-(0)228-107-2971
E-Mail: hoehns@bibb.de

Übersetzung ins Russische: Paul Kremel, Neuss



Der Inhalt dieses Werkes steht unter einer Creative Commons Lizenz (Lizentyp: Namensnennung – Keine kommerzielle Nutzung – Keine Bearbeitung – 3.0 Deutschland).

Das Werk wird durch das Urheberrecht und/oder einschlägige Gesetze geschützt. Jede Nutzung, die durch diese Lizenz oder Urheberrecht nicht ausdrücklich gestattet ist, ist untersagt.

Weitere Informationen finden Sie im Internet auf unserer Creative Commons-Infoseite <http://www.bibb.de/cc-lizenz>

Vertriebsadresse:

Bundesinstitut für Berufsbildung
Fachabteilung 1.2 (VÖ)
53043 Bonn
Bestell-Nr.: 14.059

Copyright 2002 by Bundesinstitut für Berufsbildung, Bonn
Herausgeber: Bundesinstitut für Berufsbildung,
Der Generalsekretär, Bonn
Internet: www.bibb.de
E-Mail: zentrale@bibb.de
Umschlaggestaltung: Hoch Drei Berlin
Herstellung: Bundesinstitut für Berufsbildung, Bonn
Druck: Bundesinstitut für Berufsbildung, Bonn
Printed in Germany

ISBN 3-88555-718-5

Gedruckt auf Recyclingpapier, hergestellt aus 100 % Altpapier

Diese Netzpublikation wurde bei Der Deutschen Bibliothek angemeldet und archiviert.
URN: [urn:nbn:de:0035-0028-3](http://nbn-resolving.org/urn:nbn:de:0035-0028-3)

1. Einleitung

Die wissenschaftliche Evaluationsforschung gehört als Teilgebiet sozialwissenschaftlicher Forschungen weltweit seit vielen Jahren zu den am stärksten expandierenden Tätigkeitsfeldern von Akademikern. Immer mehr staatliche Institutionen in einer ebenfalls ständig wachsenden Zahl von Ländern greifen zur Vorbereitung, Durchführung und Aufbereitung der Ergebnisse ihrer politischen Programme auf die Unterstützung externer Experten – Evaluatoren – zurück. In einigen Anwendungsbereichen konnten sich Evaluationsroutinen herausbilden, und in manchen Ländern ist teilweise die Durchführung von Wirkungskontrollen sogar gesetzlich vorgeschrieben.

Über den gegenwärtigen Stand der Institutionalisierung von Evaluationen sowie den historischen Verlauf dieser Wissenschaftssparte in den westlichen Demokratien möchte dieser Beitrag informieren. Eine kurze Begriffsklärung ist diesen Ausführungen vorangestellt, die schließlich auch eine Beschreibung internationaler Standards und methodische Hinweise zur Durchführung von Evaluationen enthalten.

Ziel dieses Beitrags ist es, russischen Kollegen und Kolleginnen in der Berufsbildung eine kurze Übersicht über den Stand der internationalen Evaluationsforschung und -praxis und Anregungen für den Aufbau eines eigenen, den internationalen Standards entsprechenden, Evaluationskonzepts zu liefern. Schulleiter, Ausbilder, Lehrer und andere Praktiker der beruflichen Bildung soll dieser Beitrag über die Aufgaben und Ziele von Evaluation informieren.

2. Begriffsbestimmung

Der Begriff Evaluation stammt von dem lateinischen Wort „valor“, d. h. „Wert“, und der Vorsilbe e / ex, d. h.: aus, und bedeutet übersetzt „Bewertung“, „einen Wert aus etwas ziehen“. In seinem weitesten Wortsinn beschreibt der Begriff den „Prozess der Beurteilung des Wertes eines Produktes, Prozesses oder eines Programmes, was nicht notwendigerweise systematische Verfahren oder datengestützte Beweise zur Untermauerung einer Beurteilung erfordert“.¹ Somit ist z. B. bereits der morgendliche Blick aus dem Fenster zur Begutachtung des Wetters eine „Evaluation“. Obwohl der Begriff „Evaluation“ den meisten Menschen unbekannt sein dürfte, ist der damit beschriebene Vorgang so alt wie die Menschheit selbst.²

Für die Bezeichnung der wissenschaftlichen Herangehensweise wurde das Wort „Evaluation“ gerade wegen seiner vergleichsweise geringen Verbreitung gewählt, um sich von dem alltagssprachlichen Verständnis von „Bewertung“ oder „Beurteilung“

¹ Wottawa & Thierau 1990: 9

² vgl. Stockmann 2000: 11

abzuheben. Die allgemein anerkannten sozialwissenschaftlichen Definitionen von „Evaluation“ unterstreichen die explizite Verwendung wissenschaftlicher Forschungsmethoden³ und verstehen unter „Evaluationsforschung“ die „*systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme*“.⁴

Damit unterscheiden sich wissenschaftliche Evaluationen von Alltagsevaluationen ausschließlich durch die Nutzung empirischer Forschungsmethoden und die Einhaltung der hiermit verbundenen wissenschaftlichen Standards.⁵ Diese Differenzierung lässt sich nicht nur für sozialwissenschaftliche Fragestellungen vornehmen: im Unterschied zur eigenen Begutachtung der Wetterlage basiert das Urteil des Meteorologen auf wissenschaftlichen Analysen und dem Einsatz geeigneter standardisierter Messinstrumente. Dies bedeutet nicht zwangsläufig, dass sein Urteil zutreffender als die persönliche Einschätzung sein muss. Durch die wissenschaftliche Fundierung des Urteils werden aber die Bewertungskriterien offengelegt und dadurch die logischen Ableitungen von Bewertungen für Dritte nachvollziehbar. Wissenschaftliche Evaluationen leisten somit einen wesentlichen Beitrag zur Versachlichung von Diskussionen und zur Objektivierung des Urteilsprozesses.

3. Historische Entwicklung und Institutionen

Die Ursprünge wissenschaftlicher Evaluationen lassen sich in den USA bis in das neunzehnte Jahrhundert zurückverfolgen, als die Bundesregierung zum ersten Mal externe Inspektoren mit der Bewertung öffentlicher Programme beauftragte. Zu den damals untersuchten staatlichen Einrichtungen gehörten Waisenhäuser, Schulen, Krankenhäuser und Gefängnisse.⁶ Eine erste Expansionsphase von Evaluationen lässt sich in den Vereinigten Staaten für die dreißiger und vierziger Jahre des zwanzigsten Jahrhunderts festhalten. Im Rahmen des „New Deal“ wurden z. B. Reformprogramme zur Verminderung der Arbeitslosigkeit und zur Verbesserung der sozialen Sicherheit von Evaluationsstudien begleitet.⁷ Die meisten Autoren bestimmen allerdings erst die frühen sechziger Jahre als Ausgangspunkt professioneller Evaluationen, weil in der Folge umfangreicher staatlicher Sozial-, Bildungs-, Gesundheits-, Ernährungs- und Infrastrukturprogramme nicht nur die Anzahl der Evaluationsstudien sprunghaft anstieg, sondern auch ein Institutionalisierungsprozess einsetzte.⁸

Ein wesentliches Kennzeichen der fortschreitenden Professionalisierung der Evaluationstätigkeiten in den USA war die Gründung einer Reihe einschlägiger Fachzeitschriften, unter denen wohl das „American Journal of Evaluation“ und die „Evaluation

³ vgl. z.B. Bortz & Döring 2002: 102; Hadley & Mitchell 1995: 48; Mertens 2000: 41; Shadish 1994: 352

⁴ Rossi, Freeman & Hofmann 1988: 3

⁵ vgl. Stockmann 2000: 12

⁶ vgl. Madaus, Stufflebeam & Scriven 1983

⁷ vgl. Stockmann 2000: 21f.

⁸ vgl. Mertens 2000: 42ff.

Review“ zu den international am meisten rezipierten gehören dürften. 1986 formierte sich durch den Zusammenschluss zweier kleinerer Evaluationsverbände die American Evaluation Association (AEA), die gegenwärtig mehr als 3.000 Mitglieder vertritt. Ziel dieser Organisation ist die Verbesserung von Evaluationsmethoden und ihres praktischen Einsatzes sowie die Förderung wissenschaftlicher Evaluationsforschung in der amerikanischen Öffentlichkeit (zu den Aufgaben und Zielen der AEA siehe die Website der Organisation „www.eval.org“). Durch die Verbreitung der Anfang der neunziger Jahre entwickelten Evaluationsstandards des „Joint Committee“ sowie der „Guiding Principles for Evaluators“ hat die AEA einen wesentlichen Beitrag zur Qualitätssicherung von Evaluationen geleistet.

Nach Schätzungen wurden in den USA bereits 1976 etwa 600 Mio. \$ für die Evaluation von sozialen Dienstleistungsprogrammen aufgewendet.⁹ In vielen Ressorts sind Evaluationen gesetzlich vorgeschrieben und eigene Haushaltsmittel werden dafür bereitgestellt. In den achtziger Jahren wurden sogenannte „Sunset-Gesetze“ erlassen, die bei fehlendem Nachweis der Wirksamkeit von Programmen deren automatische Beendigung nach einer festgelegten Zeitspanne vorsahen. 1993 verabschiedete der US-Congress den „Government Performance and Results Act“, welcher die regelmäßige Durchführung von Evaluationen für alle Bundesprogramme gesetzlich verankerte.¹⁰ Insgesamt ist somit festzuhalten, dass in den USA wissenschaftliche Evaluationen zu einem zentralen Steuerungsinstrument politischen Handelns geworden sind und sich entsprechende Institutionen zur Qualitätssicherung der Evaluations-tätigkeit herausgebildet haben.

Im Vergleich zu den USA setzte in Europa eine Professionalisierung der Evaluationsforschung mit zehnjähriger Verspätung Ende der sechziger Jahre ein. Zu den Vorreitern dieser Entwicklung gehörten Schweden, Großbritannien und die Bundesrepublik Deutschland. Ähnlich wie in den USA wurden Evaluationen zunächst hauptsächlich zum Nachweis der Effektivität von Programmen in der Bildungs-, Gesundheits- und Sozialpolitik eingesetzt, um dann Ende der siebziger Jahre mit einer Prioritätsverlagerung auf Kosten-Nutzen-Erwägungen konfrontiert zu werden. Im Unterschied zu den USA stagnierte aber die Entwicklung der Evaluationsforschung in Folge der Einsparungen öffentlicher Haushalte, und die noch in den Anfängen befindlichen Institutionalierungsversuche scheiterten weitgehend. Erst Anfang der 90er Jahre ergab sich für die Evaluationsforschung im Zuge der Reform- und Modernisierungsdebatte staatlicher Institutionen, die unter dem Stichwort „New Public Management“ zuerst in Großbritannien, Holland und den skandinavischen Staaten Konturen gewann, ein neuer Aufschwung.¹¹

Beginnend in den späten 80er Jahren in Großbritannien - und auf europäischer Ebene mit Formierung der European Evaluation Society (EES) - werden besonders seit Mitte

⁹ vgl. Stockmann 2000: 21

¹⁰ vgl. Mertens 2000: 49

¹¹ vgl. Stockmann 2000: 23f.

der 90er Jahre in immer mehr Ländern Evaluationsgesellschaften ähnlich der American Evaluation Association gegründet (1997 z. B. die Deutsche Gesellschaft für Evaluation). Gegenwärtig existieren Evaluationsgesellschaften in Belgien, Deutschland, Finnland, Frankreich, Großbritannien, Italien, den Niederlanden, Schweden und der Schweiz.¹² In weiteren Ländern (z. B. Österreich) wurden vergleichbare Initiativen gestartet.

Trotz erheblicher Entwicklungsrückstände lässt sich jedoch auch für Westeuropa feststellen, dass sich die Evaluationsforschung etabliert hat und derzeit weiterhin stark expandiert. Die Bemühungen zur Professionalisierung befinden sich, verglichen mit den USA, noch in einer Anfangsphase, es lassen sich allerdings in vielen Ländern rasche Fortschritte erkennen. Ohne systematische wissenschaftliche Beobachtungen und Beurteilungen der Wirkungen politischen Handelns kommen moderne Demokratien nicht mehr aus. Ein aktuelles Beispiel für die politischen Konsequenzen wissenschaftlicher Evaluationen stellt die PISA-Studie zur vergleichenden Evaluierung von Schulleistungen in einer Vielzahl von Ländern dar¹³: Die Ergebnisse für die Bundesrepublik Deutschland wiesen zum einen auf Grund des schwachen Abschneidens der Schüler auf einen akuten und seit Jahren vernachlässigten Handlungsbedarf hin, und zum anderen konnten einige langjährige ideologische Grabenkämpfe der deutschen Bildungspolitik durch Bereitstellung sachlicher Informationen beendet werden.

Generell kann in Deutschland eine Reihe von Bereichen aufgezählt werden, für die bereits regelmäßig und standardisiert Evaluationen durchgeführt werden. In der Entwicklungszusammenarbeit gehören wissenschaftliche Evaluationen schon seit mehr als zwanzig Jahren zu den Routinen der Projektdurchführung, wobei sich in letzter Zeit vor allem die Zahl der ex-post Evaluationen stark erhöht hat.¹⁴ Die Diskussionen über Formen der Hochschulevaluation sind in einigen Bundesländern bereits soweit fortgeschritten, dass es zur Etablierung standardisierter Evaluationsinstitutionen (z. B. der Zentralen Akkreditierungsstelle ZEVA in Hannover) gekommen ist. Auch im Arbeitsmarktbereich, bei sozialen Einrichtungen oder in der Schulpolitik sind entsprechende Ansätze erkennbar.

Für die berufliche Bildung ist festzuhalten, dass durch das Bundesinstitut für Berufsbildung (BIBB) eine Vielzahl von Evaluationen selbst durchgeführt oder als Auftrag an externe Experten vergeben wird. Eine standardisierte Vorgehensweise konnte sich jedoch angesichts der Vielfalt unterschiedlicher Aufgabenstellungen noch nicht herausbilden. Evaluiert werden zum Beispiel alle neu geordneten Berufe, Modellversuche, neu eingeführte Lehrinhalte wie Umweltbildung, neue Prüfungsformen, Programme zur Ausbildungsförderung und zur Reduzierung der Jugendarbeitslosigkeit und vieles andere mehr.¹⁵

Ziel der Evaluationen ist es in jedem Fall, einen Beitrag zur Optimierung eines Programms oder einer Neuerung zu leisten.

¹² vgl. Leeuw 2000: 59 f.

¹³ vgl. Baumert et al. 2000, 2001

¹⁴ vgl. Stockmann 1997; 2000a

¹⁵ Berger, Walden 2001, Härtel, Stockmann, Gaus 2000, Ebbinghaus, Görmar, Stöhr 2001, Stiller, Stöhr

4. Evaluationsstandards und Durchführungsprinzipien

Wie werden nun wissenschaftliche Evaluationen durchgeführt? Hierfür wurden im Jahr 1994 von einem eigens eingerichteten „Joint Committee on Standards of Educational Evaluation“ in den USA Standards entwickelt, die in vier Hauptgruppen zusammengefasst wurden.¹⁶ Dabei sollen *Nützlichkeitsstandards* sicherstellen, dass sich eine Evaluation an den Informationsbedürfnissen der vorgesehenen Evaluationsnutzer ausrichtet. Zu ihnen gehört z. B. die Offenlegung der eingesetzten Verfahren und Vorgehensweisen gegenüber den Auftraggebern und gegenüber den Evaluierten. Mit *Durchführbarkeitsstandards* soll gewährleistet werden, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst ausgeführt wird. Ein wesentlicher Aspekt ist z. B. die angemessene Berücksichtigung der Interessen und Werte der an einer Evaluation beteiligten Interessengruppen und Personen. Über die Festlegung von *Korrektheitsstandards* wird geregelt, dass eine Evaluation rechtlich und ethisch korrekt durchgeführt sowie dem Wohlergehen der in die Evaluation einbezogenen und durch die Ergebnisse betroffenen Personen gebührende Aufmerksamkeit gewidmet wird. Zu den wichtigsten Regeln ist hier die Offenlegung der Ergebnisse gegenüber allen an der Evaluation beteiligten Personen zu zählen. Mit den *Genauigkeitsstandards* schließlich soll garantiert werden, dass eine Evaluation fachlich angemessene Informationen über den Nutzen des evaluierten Programms hervorbringt und vermittelt. Insbesondere die Verpflichtung auf die adäquate Nutzung allgemein anerkannter Verfahren zur Informationsgewinnung ist dabei als wichtiger Gesichtspunkt hervorzuheben.

Die Standards wurden mittlerweile von Evaluationsgesellschaften in vielen verschiedenen Ländern (u. a. in Deutschland und der Schweiz) übersetzt und in leicht modifizierter Form übernommen. Die Deutsche Gesellschaft für Evaluation (DeGEval) arbeitet gegenwärtig an einer russischen Übersetzung, die noch im Laufe dieses Jahres verfügbar sein soll. Aufgrund ihrer weltweiten Verbreitung und der großen Zustimmung unterschiedlichster Akteure können die „Joint Committee“-Standards als allgemein anerkannter Konsens der Evaluationsforschung zur Sicherung der Qualität von Evaluationen bezeichnet werden. Auch auf der Ebene der EU, die ebenfalls Evaluationen in den verschiedensten Politikbereichen durchführt bzw. in Auftrag gibt, u. a. auch im Bereich der Berufsbildung, wird gegenwärtig an einer Anpassung der DeGEval-Standards auf Evaluationen und Wirkungsforschungen speziell im europäischen Berufsbildungswesen gearbeitet.

In den USA wurden diese Standards von der American Evaluation Association (AEA) durch Leitlinien für Evaluatoren¹⁷ ergänzt. Die „Guiding Principles for Evaluators“ fordern von den Evaluatoren

¹⁶ ausführlich siehe Joint Committee on Standards of Educational Education 2000; eine Kurzzusammenfassung ist bei Widmer 2000 zu finden

¹⁷ Shadish et al. 1995

-
- die systematische, datenbasierte Untersuchung des zu evaluierenden Gegenstands,
 - den Beleg ausreichender Fachkompetenz zur Durchführung von Evaluationsstudien,
 - die Gewährleistung eines fairen und integren Evaluationsprozesses,
 - den angemessenen Respekt gegenüber den Persönlichkeitsrechten aller Beteiligten sowie
 - ein allgemeines Verantwortungsgefühl für die durch Programme beeinflusste öffentliche Wohlfahrt.

Den Auftraggebern wird empfohlen, Aufträge nur an Evaluatoren zu vergeben, die sich diesen Prinzipien verpflichtet haben.

In den europäischen Evaluationsgesellschaften ist es bisher noch nicht gelungen, einen vergleichbaren Konsens zur Formulierung von Leitlinien zu erzielen. Es gibt dabei allerdings keine grundsätzlichen Widerstände gegen die von der AEA vorgelegten und hier kurz vorgestellten Prinzipien.

Das wesentliche Problem einer einheitlichen Festlegung von Leitlinien für Evaluatoren liegt vielmehr in der Vielfalt der bearbeiteten Aufgaben und Themenfeldern begründet. Es würde hier zu weit führen, auf die Vielzahl der bearbeiteten Themenstellungen (zum Beispiel in den bereits oben erwähnten Politikbereichen Gesundheitswesen, Sozialwesen, Bildungswesen und vielen anderen mehr) im einzelnen einzugehen und ihre Besonderheiten herauszustellen. Die vielfältigen Anforderungen an die Durchführung einer Evaluation hängen beispielsweise vom politischen Prozess des zu evaluierenden Programms ab, der sehr grob in drei verschiedene Phasen unterteilt werden kann¹⁸. Während der Programmentwicklung werden zur Untersuchung der Ausgangs- und Rahmenbedingungen *ex-ante Evaluationen* (Machbarkeitsstudien) mit dem Ziel einer Einschätzung möglicher Durchführungsprobleme sowie der Erfolgchancen geplanter Maßnahmen in Auftrag gegeben. Der Charakter dieser prognostisch orientierten Studien unterscheidet sich stark von dem *formativer Evaluationen*, die als Begleitforschung möglichst aktuelle Informationen über den Programmverlauf sammeln und ihre Bewertungen als Entscheidungshilfen für die Steuerung zur Verfügung stellen möchten. Summative oder *ex-post Evaluationen* schließlich erfassen und beurteilen die langfristigen Programmwirkungen, wobei stärker als bei den anderen Vorgehensweisen die Identifikation kausaler Zusammenhänge im Vordergrund steht.

Als ein Beispiel kann das von Reinhard Stockmann entwickelte Evaluationskonzept angeführt werden, welches vom Bundesinstitut für Berufsbildung (BIBB) und vom Institut für die Entwicklung der Berufsbildung (IRPO) zur *ex-post Evaluation* von

¹⁸ vgl. Stockmann 2000: 13 ff.

ausgewählten deutsch-russischen Modellprojekten aus dem TRANSFORM-Programm genutzt wird. Es handelt sich hierbei um ein theoriegeleitetes Konzept mit einem integrativen methodischen Ansatz, dessen Grundlagen wie folgt kurz zu skizzieren sind:¹⁹

Zur Bestimmung der Analysefelder werden drei unterschiedliche theoretische Modelle zusammengeführt.

1. Durch Nutzung von Erkenntnissen der Lebensverlaufsforschung (das zu evaluierende Projekt wird, grob vergleichbar mit dem menschlichen Lebensverlauf, in eine Entstehungsphase, eine Verlaufsphase und eine Phase der Nachwirkungen aufgeteilt) ergibt sich eine zeitliche Perspektive, die die historische Projektentwicklung und die jeweiligen Anforderungen zueinander in Beziehung setzt und so für die einzelnen Projektphasen eine Bewertung der Entscheidungen zur Projektplanung und -steuerung ermöglicht.

2. Ein organisationstheoretisch gestütztes Wirkungsmodell lenkt den Blick auf die Leistungsfähigkeit der jeweiligen Trägerorganisation eines Projekts (z. B. eine berufliche Lehranstalt, an der eine Neuerung des Berufsbildungssystems modellhaft erprobt wurde) und hilft die Frage zu beantworten, welche Auswirkungen im weitesten Sinn ein Projekt auf die Trägerorganisation hatte, was sich im Zusammenhang mit dem Projekt verändert hat.

3. Die Leistungsfähigkeit der Trägerorganisation ist auch für die Verbreitung der eingeführten Neuerungen über die Grenzen der Organisation hinaus von entscheidender Bedeutung. In der beruflichen Bildung sollen beispielsweise von Modellprojekten Wirkungen sowohl innerhalb des Berufsbildungssystems (z. B. die Verbreitung des Modellversuchs) als auch innerhalb des Erwerbssystems (z. B. die Verbesserung der Anpassung von Qualifikationen an die Erfordernisse des Arbeitsmarktes) erzielt werden. Derartige Verbreitung oder, um den Fachterminus zu gebrauchen, Diffusion von Projektwirkungen ist in dem Evaluationskonzept von Stockmann ebenfalls Gegenstand der Untersuchungen.

Bei der methodischen Vorgehensweise berücksichtigt das Konzept von Stockmann einige Besonderheiten von Evaluationen. So werden durch den partizipativen Ansatz die Fachkenntnisse und Erfahrungen von Evaluationsexperten mit dem Insiderwissen der Projektbeteiligten zusammengeführt. Grundlage ist der offene Austausch von Informationen und die gegenseitige faire Behandlung während der Evaluierung.

Besondere methodische Schwierigkeiten könnten sich durch die Mängel einzelner Datenerhebungsmethoden ergeben, die während einer Evaluation nicht wie im Experiment durch entsprechende Kontrollmaßnahmen ergänzt werden können. Um die

¹⁹ Ausführlich hierzu siehe: Stockmann 1997.

Schwächen der einzelnen Verfahren auszugleichen und dabei ihre Stärken möglichst optimal zu nutzen, werden in dem Evaluationskonzept unterschiedliche Datenerhebungstechniken in einem Multimethodenansatz zusammengefasst.

Obwohl auf diesem Weg wesentliche Verbesserungen der Datenqualität erzielt werden können, bleibt ein weiteres Problem davon unberührt: in der Regel werden bei einer Evaluation einige wenige Projekte untersucht, zu denen dann sehr viele Informationen vorliegen, während die Evaluation letztendlich eine Vielzahl unterschiedlicher Forschungsfragen beantworten soll. Letztlich lässt sich nur durch geeignete Vergleiche bestimmen, inwieweit die in einem Einzelfall gefundenen Zusammenhänge rein zufällig entstanden und damit einmalig oder die Folge systematischer und dementsprechend regelmäßig wieder auftretender kausaler Ursachen sind. Die Integration eines komparativen Ansatzes ist dementsprechend zur Identifikation von Problemfeldern und zur Erarbeitung von Lösungsvorschlägen unabdingbar notwendig.

5. Ausblick

Was kann Russland aus der Geschichte der Evaluation lernen?

In Russland finden seit Beginn der Perestrojka Mitte der 80er Jahre und sehr intensiv seit Beginn der 90er Jahre tiefgreifende Umwälzungen in allen Bereichen der Politik statt, insbesondere auch im Bildungswesen. Es gibt ein neues „Gesetz über die Bildung“, ein neues Verzeichnis der Ausbildungsberufe, neue Berufsbilder, neue Forderungen des Arbeitsmarkts an die Absolventen und entsprechend neue Ausbildungsmethoden, neue Schulformen. Berater aus verschiedenen ausländischen Staaten wurden eingeladen, damit Russland die jeweiligen herausragenden internationalen Erfahrungen näher kennen lernen und gegebenenfalls für sich übernehmen konnte. In zahlreichen nationalen und internationalen Modellversuchen wurden und werden Neuerungen im Bildungswesen erprobt. Allein das Bundesinstitut für Berufsbildung führte zusammen mit dem russischen Partner IRPO neun teilweise sehr komplexe Modellprojekte durch, in denen berufsbildende Ausbildungsgänge modernisiert oder völlig neu konzipiert wurden.

Zum gegenwärtigen Zeitpunkt ist in Russland eine Art Besinnungspause eingetreten. Der Effekt vieler Veränderungen wird überdacht. Welche Wirkungen hat diese oder jene Neuerung gebracht, welche Beratung war effektiv und angemessen? Welche Bedingungen führten zum Erfolg eines Modellversuchs, und welche Modellversuche lohnt es zu verbreiten?

Eine systematische, wissenschaftlich begründete und damit für alle Beteiligten transparente und nachvollziehbare Antwort auf solche und ähnliche Fragen erhält man mit der Durchführung von Evaluationen.

Wie bereits aus dem historischen Abriss deutlich wurde, hat sich die wissenschaftliche Evaluation in allen modernen westlichen Ländern als wichtiges Instrument zur rationalen Steuerung politischen Handelns und zur öffentlichen Legitimierung von Entscheidungen, zur Überprüfung der Wirkungen staatlicher Programme und zur Kontrolle der Effektivität eingesetzter finanzieller und sonstiger Ressourcen bewährt. Entscheidend ist hierbei die Sicherstellung der Professionalität der Evaluationsdurchführung, wozu die Einhaltung internationaler Standards ebenso gehört wie die Vermeidung von Beeinflussung der Evaluationsergebnisse durch einzelne Interessengruppen. Der Einsatz wissenschaftlicher Evaluationen nach internationalem Standard und die rückhaltlose Offenlegung der Ergebnisse, auch wenn diese möglicherweise nicht den Wünschen und Erwartungen entsprechen, trägt letzten Endes auch zu einem Klima der Offenheit, des Vertrauens, des Dialogs in der Gesellschaft bei, von dem alle profitieren.

Erste Kontakte zwischen russischen Sozialwissenschaftlern und deutschen Evaluationsforschern bestehen bereits. Diese Kontakte verdienen es, vertieft zu werden. Bei der Durchführung von Evaluationen und auch bei der Einrichtung eigener Strukturen der Evaluationsforschung in Russland können die russischen Sozialwissenschaftler auf das Erfahrungswissen ihrer Kollegen und Kolleginnen insbesondere aus den westeuropäischen Ländern zurückgreifen.

Das IRPO, das bereits seit langem die Umgestaltungen im russischen Berufsbildungssystem wissenschaftlich begleitet, sammelt durch die Zusammenarbeit mit dem BIBB in einem gemeinsamen Forschungsprojekt erste Erfahrungen mit der Durchführung von Evaluationen, die durch die Vertiefung spezifischer sozialwissenschaftlicher Qualifikationen und durch eine systematische Erweiterung der Anwendungsbereiche ausgebaut werden können. Mit diesem Ziel könnten entsprechende Fort- und Weiterbildungskurse für junge russische Akademiker durch erfahrene Evaluationsexperten aus westlichen Ländern in Russland durchgeführt oder entsprechende Angebote wie z. B. in Deutschland die regelmäßig vom Centrum für Evaluation der Universität des Saarlandes (CEval) in Zusammenarbeit mit der Deutschen Stiftung für Internationale Entwicklung (DSE) in Mannheim veranstalteten Schulungen genutzt werden.

Bereits durchgeführt wurde in Zusammenarbeit des Bundesinstituts für Berufsbildung, Bonn, mit der Universität des Saarlandes und dem Institut für die Entwicklung der Berufsbildung, Moskau, eine erste wissenschaftliche Evaluation von einem deutsch-russischen Beratungsprojekt, deren Ergebnisse in der evaluierten russischen Lehranstalt auf großes Interesse stießen und wiederum neue Impulse für die Verbesserung der Unterrichtstätigkeit gaben. Eine zweite ähnlich angelegte Evaluation eines in deutsch-russischer Kooperation modernisierten oder völlig neu konzipierten kaufmännischen Berufs steht in der Anfangsphase. Es ist zu erwarten, dass die Ergebnisse der Evaluationen die Diskussionen in der russischen Berufsbildung beleben und bereichern werden.

Literatur

- BAUMERT, J. & BOS, W. & LEHMANN, R. (Hrsg., 2000), TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn (2 Bde.), Opladen: Leske+Budrich.
- BAUMERT, J. & KLIEME, E. & NEUBRAND, M. (Hrsg., 2001), PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich, Opladen: Leske+Budrich.
- BERGER, K. & WALDEN, G. (2001), Evaluierung der Bund-Länder-Programme zur Ausbildungsförderung in den neuen Bundesländern 1996 – 1999, Bonn: Bundesinstitut für Berufsbildung
- BORTZ, J. & DÖRING, N. (2002), Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, Berlin u.a.: Springer (3. Auflage).
- EBBINGHAUS, M. & GÖRMAR, G. & STÖHR, A., Evaluert: Projektarbeit und Ganzheitliche Aufgaben: Ergebnisbericht zur Evaluation der Abschlussprüfungen in den vier IT-Berufen, Bielefeld: Bertelsmann
- HADLEY, R. G. & MITCHELL, L. K. (1995), Counseling research and program evaluation, Pacific Groce, CA: Brooks/Cole.
- HÄRTEL, M. & STOCKMANN, R. & GAUS, H. (2000), Berufliche Umweltbildung und Umweltberatung – Grundlagen, Konzepte und Wirkungsmessung, Bielefeld: Bertelsmann
- JOINT COMMITTEE ON STANDARDS FOR EDUCATION EVALUATION, J.R. SANDERS (eds., 2000), Handbuch der Evaluationsstandards. Die Standards des 'Joint Committee on Standards for Educational Evaluation'. Aus dem Amerikanischen übersetzt von Wolfgang Beywl und Thomas Widmer. Deutsche Ausgabe bearbeitet und ergänzt von Wolfgang Beywl, Thomas Widmer und James R. Sanders, Opladen: Leske + Budrich.
- LEEuw, F. L. (2000), Evaluation in Europe, in: Stockmann, R. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske+Budrich, S. 57-76.
- MADDAUS, G. F. & STUFFLEBEAM, D. L. & SCRIVEN, M. S. (1983), Program evaluation: A historical overview, in: Maddaus, G. F., M. S. Scriven & D. L. Stufflebeam (eds.), Evaluation Models, Boston: Kluwer-Nijhoff, pp. 3-22.
- MERTENS, D. M. (2000), Institutionalizing Evaluation in the United States of America, in: Stockmann, R. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske + Budrich, S. 41-56.
- ROSSI, P. H. & FREEMAN, H. E. & HOFMANN, G. (1988), Programm Evaluation. Einführung in die Methoden angewandter Sozialforschung, Stuttgart: Enke. Neueste Englische Auflage: Rossi, P. H., H. E. Freeman & M. W. Lipsey (1999), Evaluation. A Systematic Approach, Thousand Oaks et al.: Sage (6th ed.).
- SHADISCH, W. R. (1994), Need-based evaluation: good evaluation and what you need to know about it, in: Evaluation Practice, 15 (3), pp. 347-358.
- SHADISH, W. R. & NEWMAN, D. & SCHEIRER, M. A. & WYE, C. (eds., 1995), The American Evaluation Association's Guiding Principles, San Francisco: Jossey-Bass.
- STILLER, I. & STÖHR, A. (2001), Ergebnisse aus der Evaluation der Büroberufe, Bielefeld: Bertelsmann.
- STOCKMANN, R. (1997), The Sustainability of Development Cooperation, Baden-Baden: Nomos.
- STOCKMANN, R. (2000), Evaluation in Deutschland, in: ders. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske + Budrich, S. 11-40.
- STOCKMANN, R. (2000a), Evaluation staatlicher Entwicklungspolitik, in: ders. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske + Budrich, S. 375-408.
- WIDMER, T. (2000), Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird, in: Stockmann, R. (Hrsg., 2000), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske+Budrich, S. 77-102.
- WOTTAWA, H. & THIERAU, H. (1990), Lehrbuch Evaluation, Bern u.a.: Huber Verlag.

ПУБЛИКАЦИИ ДЛЯ НАУЧНОГО ОБСУЖДЕНИЯ

Издание № 59

Что такое эвалюация?

Доктор Вольфганг Майер
Саарский университет
Центр эвалюационных исследований - SEval

Габриела Хёнс
Федеральный институт профессионального
образования - BIBB

Серия публикаций
Федерального института
профессионального образования
Генеральный секретарь . Бонн

Федеральный институт
профессионального образования

BIBB ▶

- ▶ исследования
- ▶ консультация
- ▶ созидание будущего

ПУБЛИКАЦИИ ДЛЯ НАУЧНОГО ОБСУЖДЕНИЯ федерального института профессионального образования (БИББ) издаются Генеральным секретарем. Они выходят в свет как именные статьи своих авторов и отражают мнение авторов, которое не обязательно совпадает с мнением издателя. Они защищаются авторским правом. Их публикация служит импульсом для дискуссий в научной общественности.

Авторы:

Д-р Вольфганг Майер,
Центр эвалюационных исследований (CEval)
Саарский университет- FR 5.2
п/я 151 150
Д-66041 Саарбрюкен
ФРГ
Тел: ++49-(0)681-302-3320
Факс: ++49-(0)681-302-3899
e-mail: w.meyer@rz.uni-sb.de

Габриэла Хёнс,
Федеральный институт профессионального образования
Д-53043 Бонн
ФРГ
Тел: ++49 -(0)228-107-1931
Факс: ++49-(0)228-107-2971
E-Mail: hoehns@bibb.de

Перевод на русский язык: Пауль Кремель, г. Нойсс

Адрес для заказов:

Федеральный институт профессионального образования
Область деятельности 1.2 (публикации)
Д-53043 г. Бонн
Номер для заказа: 14.059

(с) 2002 by Федеральный институт профессионального образования
[Bundesinstitut für Berufsbildung], г. Бонн [Bonn]
Издание: Федеральный институт профессионального образования,
Генеральный секретарь, 53043 г. Бонн [53043 Bonn]
сайт в интернете: www.bibb.de
адрес электронной почты: zentrale@bibb.de
Оформление: «Хох драй» [Hoch Drei], г. Берлин [Berlin]
Типография: Федеральный институт профессионального образования
[Bundesinstitut für Berufsbildung], г. Бонн [Bonn]

ISBN 3-88555-718-5

Напечатано на бумаге, изготовленной из 100 % макулатуры.

1. Введение

Научная эвалюация как раздел общественно-научных исследований уже на протяжении многих лет является одной из наиболее интенсивно развивающихся областей науки во всём мире. Всё больше государственных учреждений в нарастающем числе стран мира обращается за поддержкой к сторонним экспертам-эвалюаторам при подготовке, проведении и анализе результатов реализации политических программ. В некоторых областях эвалюация уже стала повседневной практикой, а в некоторых странах проведение контроля эффективности предписывается действующим законодательством.

Данная статья информирует читателя о современном уровне институционализации эвалюации, а также об историческом развитии данной области науки в демократических странах Запада. В начале статьи даётся краткое определение термина «эвалюация». Статья включает в себя также характеристику международных норм и методических рекомендаций по проведению эвалюаций.

Цель настоящей статьи состоит в кратком ознакомлении российских коллег, занятых в сфере профессионального образования, с современным состоянием теории и практики эвалюации в мире. Статья призвана дать толчок к разработке собственной концепции эвалюации, соответствующей международным нормам и стандартам. Директоров школ, мастеров производственного обучения, инструкторов, преподавателей и других практиков сферы профессионального образования данная статья знакомит с задачами и целями эвалюации.

2. Определение термина «Эвалюация»

Термин «эвалюация» происходит от латинского «valor» т. е. «стоимость» / «ценность», и приставки «e/ex» и означает в переводе «оценка», «извлечение ценности чего-то». В широком смысле слова термин означает «процесс оценки стоимости или ценности продукции, процесса или программы, не требующий для своего обоснования обязательного использования систематических методов или доводов». ¹ Таким образом, «эвалюация» имеет место даже тогда, когда мы просто выглядываем утром из окна, чтобы убедиться, какая на улице погода. Несмотря на то, что термин «эвалюация» не знаком, пожалуй, большинству людей, процесс этот на самом деле стар как само человечество. ²

¹ Wottawa & Thierau 1990: 9

² ср.. Stockmann 2000: 11

Выбор термина «эвалюация» для научного обозначения этого процесса был продиктован как раз его сравнительно малой распространённостью и стремлением оттенить научное понимание этого процесса от бытового («оценка», «характеристика»). Общепризнанные определения «эвалюации» в социологии подчёркивают эксплицитное использование научных методов исследования³ и трактуют исследования в данной области как *«систематическое использование общественно-научных методов исследования с целью оценки концепции, оформления, реализации и эффекта социальных интервенционных программ»*.⁴

Таким образом, научная оценка, т. е. эвалюация, отличается от обычной оценки исключительно использованием эмпирических методов исследования и соблюдением связанных с этим научных стандартов.⁵ Это различие применимо не только к общественно-научным вопросам: в отличие от индивидуальной оценки состояния погоды суждение метеоролога основывается на научном анализе и использовании адекватных стандартизированных средств измерения. Это не означает, однако, что выносимое им суждение обязательно должно быть правильнее личной, субъективной оценки. В то же время благодаря научному обоснованию такого суждения становятся ясными и понятными критерии оценки и, следовательно, основы следующих из этой оценки логических выводов для окружающих. Таким образом, эвалюация, т. е. научная оценка, вносит важный вклад в дело объективизации дискуссий и процессов формирования суждений.

3. Историческое развитие и институты

Истоки научной эвалюации восходят к девятнадцатому веку, когда Федеральное правительство США впервые поручило сторонним инспекторам осуществить оценку некоторых государственных программ. Объектом проведённых тогда исследований были такие государственные учреждения, как детские приюты, школы, больницы и тюрьмы.⁶ Первый этап распространения эвалюации имел место в Соединённых Штатах в 30-40 годах двадцатого столетия. В рамках программы «New Deal» эксперты-эвалюаторы сопровождали процессы реализации таких программ, как реформы, направленные на снижение безработицы и

³ ср. напр. Bortz & Döring 2002: 102; Hadley & Mitchell 1995: 48; Mertens 2000: 41; Shadish 1994: 352

⁴ Rossi, Freeman & Hofmann 1988: 3

⁵ ср. Stockmann 2000: 12

⁶ ср. Madaus, Stufflebeam & Scriven 1983

улучшение социальной защищённости населения.⁷ Большинство авторов придерживается, однако, мнения о том, что началом профессиональной эвалюации следует считать период начала шестидесятых годов, поскольку именно в это время вследствие реализации широких государственных программ в сфере социального развития, образования, здравоохранения, питания и инфраструктуры резко возросло не только число эвалюационных исследований, но и начался процесс соответствующей институционализации.⁸

Признаком все продолжающейся профессионализации деятельности в области эвалюации в США явилось создание целого ряда специализированных журналов, среди которых «American Journal of Evaluation» и «Evaluation Review», наверное, в мире самые широко распространенные издания. В 1986 году в результате слияния двух небольших объединений экспертов-эвалюаторов образовалась Американская ассоциация эвалюации («American Evaluation Association», сокр. АЕА) в состав которой сегодня входит более трёх тысяч членов. Цель этой организации состоит в совершенствовании методов эвалюации и их практического применения, а также в распространении знаний о научной эвалюации в американской общественности (задачи и цели ассоциации приводятся на сайте организации www.eval.org). В результате распространения разработанных в начале девяностых годов стандартов эвалюации Объединённой комиссии («Joint Committee»), а также «Руководящих принципов для экспертов-эвалюаторов» («Guiding Principles for Evaluators») ассоциация АЕА внесла важный вклад в дело обеспечения качества эвалюации.

По имеющимся оценкам, уже в 1976 году в США было затрачено примерно 600 млн. долларов на эвалюацию социальных программ в сфере услуг.⁹ Во многих областях проведение эвалюаций предписано действующим законодательством. На эти цели выделяются соответствующие бюджетные средства. В восьмидесятых годах были приняты законы, предусматривающие автоматическое прекращение реализации программ по истечении определённого периода времени в случае отсутствия документального подтверждения их эффективности. В 1993 году конгресс США принял Закон «Об эффективности работы правительства» (Government Performance and Results Act), в котором зафиксирована

⁷ ср. Stockmann 2000: 21f.

⁸ ср. Stockmann 2000: 21f.

⁹ ср. Mertens 2000: 42ff.

необходимость регулярного проведения эвалюации всех федеральных программ.¹⁰ Таким образом, следует отметить, что в США научная эвалюация стала одним из центральных инструментов управления в сфере политической жизни, и что в стране сформировались институты и учреждения, обеспечивающие качество эвалюационных исследований.

По сравнению с США процесс профессионализации исследований в области эвалюации начался в Европе с десятилетним опозданием, в конце шестидесятых годов. Флагманами этого процесса были Швеция, Великобритания и ФРГ. Также как и в США, методы эвалюации использовались сначала, главным образом, для определения эффективности программ в области образования, здравоохранения и социальной политики. В конце семидесятых годов приоритет в эвалюационных исследованиях перемещался в направлении сопоставления затрат и результатов. Между тем в Европе (в отличие от США) наступил период застоя в области научной эвалюации вследствие сокращения государственных бюджетов, и только что начатые первые попытки институционализации этого научного направления чаще всего не увенчались успехом. Лишь в начале девяностых годов произошел новый подъем исследований в области эвалюационных методов в ходе дискуссий, связанных с реформированием и модернизацией государственных учреждений, сначала в Великобритании, а затем в Голландии и скандинавских странах (кампания «New Public Management», англ. «Новое государственное управление»).

В нарастающем числе стран, начиная с конца восьмидесятых годов в Великобритании, а также на общеевропейском уровне, где было создано Европейское общество эвалюации (European Evaluation Society, EES), создаются организации наподобие Американской ассоциации. Этот процесс заметно усиливается с середины девяностых годов. В 1997 году, например, учреждается Германское общество эвалюации («Deutsche Gesellschaft für Evaluation»). На сегодняшний день такие организации существуют в Бельгии, Германии, Финляндии, Франции, Великобритании, Италии, Нидерландах, Швеции и Швейцарии.¹² Аналогичные инициативы получили развитие и в других странах, например в Австрии.

Несмотря на значительное отставание в Западной Европе тоже можно констатировать наличие сформировавшихся и динамично развивающихся

¹⁰ ср. Mertens 2000: 49

¹¹ ср. Stockmann 2000: 23 f.

¹² ср. Leeuw 2000: 59 f.

структур, занимающихся научными исследованиями в области методов эвалюации. По сравнению с США работа по профессионализации этой деятельности находится пока на начальном этапе, однако во многих странах отмечается уже значительный прогресс в данной области. Современные демократии уже не могут обойтись без систематического научного наблюдения и оценки последствий политических действий. Примером политической значимости научной эвалюации может служить проведённое недавно исследование «PISA», предметом которого был сравнительный анализ и оценка успеваемости школьников во многих странах мира.¹³ Результаты исследования применительно к Германии указали, во-первых, на острую и давно назревшую необходимость срочных мер ввиду низкого уровня подготовки немецких школьников, а во-вторых, позволили положить конец некоторым многолетним идеологическим стычкам и политическим распрям в сфере образования, предоставляя объективную информацию.

В Германии существует целый ряд областей, в которых регулярно проводятся стандартизированные эвалюации. В сфере сотрудничества с развивающимися странами научная эвалюация уже на протяжении более двадцати лет является повседневной практикой при реализации проектов. В последнее время отмечается резкое увеличение числа эвалюаций „ex post“, т.е. предметом которых являются уже реализованные проекты.¹⁴ Дискуссия о формах эвалюации в сфере высшей школы достигла в некоторых федеральных землях ФРГ уже такого уровня, что в результате был создан целый ряд стандартизированных институтов, например центральное управление по вопросам аккредитации в Ганновере («Zentrale Akkreditierungsstelle ZEVA»). Аналогичные процессы намечаются также в области политики по проблемам рынка труда, в социальных учреждениях и в сфере школьной политики.

В сфере профессионального образования, активную работу по эвалюации осуществляет Федеральный институт профессионального образования (BIBB), который провёл и проводит уже множество исследований подобного рода, как своими силами, так и с привлечением сторонних экспертов. Стандартизированного подхода пока не удалось разработать ввиду большого разнообразия рассматриваемых задач. Предметами эвалюации являются, например, все новые профессии, пилотные проекты, новые содержания в учебных планах, например, в области экологии,

¹³ ср. Baumert и др. 2000/20001

¹⁴ ср. Stockmann 1997; 2000a

новые формы проведения экзаменов, государственные программы поддержки при создании учебных мест на предприятиях и программы сокращения безработицы среди молодёжи, а также многое другое.¹⁵

Цель любой эвалюации состоит в оптимизации анализируемой программы или нововведения.

4. Стандарты эвалюации и принципы их осуществления

Каким же образом осуществляется научная эвалюация? Специально созданная Объединённая комиссия по стандартам эвалюации в сфере образования («Joint Committee on Standards of Educational Evaluation») в 1994 году в США разработала целый ряд стандартов, которые можно подразделить на четыре основные группы.¹⁶ *Стандарты полезности* призваны обеспечить ориентацию эвалюации на информационные запросы предполагаемых пользователей. К ним относится, например, ознакомление заказчиков и эвалюируемых с применяемыми методами и подходами эвалюации. *Стандарты реализации* призваны обеспечить реалистичность, продуманность, дипломатичность и экономичность эвалюации. При этом важным аспектом является, например, адекватный учёт интересов и ценностей участвующих в эвалюации заинтересованных групп и лиц. *Стандарты корректности* предполагают соблюдение правовых и этических принципов при осуществлении эвалюации, а также должный учёт благосостояния лиц, вовлекаемых в процесс эвалюации и испытывающих на себе его результаты. Одним из наиболее важных правил в этой связи является гласность результатов для всех участвующих в эвалюации лиц. И, наконец, соблюдение *стандартов точности* должно обеспечить возможность получения адекватной информации о пользе/эффекте эвалюируемой программы. В этой связи следует особо отметить необходимость адекватного использования общепризнанных методов сбора данных.

Эти стандарты уже переведены и приняты в слегка видоизменённой форме объединениями по вопросам эвалюации во многих странах мира (в частности, в Германии и Швейцарии). Германское общество эвалюации („DeGEval“) работает в настоящее время над русским переводом вышеуказанных стандартов. Русский текст должен быть готов ещё в этом

¹⁵ Berger, Walden 2001, Härtel, Stockmann, Gaus 2000, Ebbinghaus, Görmar, Stöhr 2001, Stiller, Stöhr 2001

¹⁶ подробнее см. Joint Committee on Standards of Educational Education 2000; краткое изложение см. Widmer 2000

году. В силу того, что нормы Объединённой комиссии получили распространение во всём мире и snискали себе одобрение самых разных кругов людей, их вполне можно рассматривать как общепризнанный результат научного консенсуса, нацеленный на обеспечение качества эвалюаций. На уровне Европейского Союза, который тоже, кстати, проводит эвалюации в различных областях политической жизни (своими силами и с привлечением сторонних экспертов, в частности в сфере профессионального образования), также ведётся работа по модификации норм «DeGEval» для проведения эвалюаций и исследования эффективности программ в сфере профессионального образования на европейском уровне.

В США вышеуказанные стандарты были дополнены Американской ассоциацией эвалюации (American Evaluation Association, AEA) принятием так называемых «Руководящих принципов работы экспертов-эвалюаторов»¹⁷ («Guiding Principles for Evaluators»). Эти принципы требуют от экспертов-эвалюаторов:

- систематического изучения предмета эвалюации на основе соответствующей базы данных,
- документального подтверждения своей профессиональной компетенции для проведения эвалюационных анализов,
- обеспечения корректного и целостного процесса эвалюации,
- должного уважения прав личности всех участников эвалюации,
- общее осознание ответственности за общественное благосостояние, на которое влияют эвалюируемые программы.

Заказчикам рекомендуется выдавать заказы на проведение исследований только тем экспертам-эвалюаторам, которые приняли эти принципы и обязались действовать сообразно им.

Европейским обществам эвалюации пока не удалось достичь аналогичного консенсуса по формулировке руководящих принципов. Однако в Европе нет принципиальных возражений к вышеизложенным принципам AEA.

Основная проблема определения единых руководящих принципов работы экспертов-эвалюаторов состоит, главным образом, во многообразии решаемых задач и тематики. Подробный анализ многочисленных прикладных областей эвалюации (например, вышеупомянутых областей политической жизни, здравоохранения, социальной сферы, сферы

¹⁷ Shadish и др.. 1995

образования и др.) выходит за рамки настоящей статьи. Различные требования, предъявляемые к осуществлению эвалюации, зависят, в частности, от политического процесса, лежащего в основе эвалюируемой программы. Этот процесс можно грубо подразделить на три этапа.¹⁸ На этапе разработки программ выдаются заказы на проведение *эвалюационных исследований «ex ante»*, т. е. с целью изучения исходных общих условий, оценки возможных проблем при реализации мероприятий и их потенциальной результативности (анализы осуществимости). По своему характеру эти прогностические исследования существенно отличаются от так называемых *формирующих эвалюаций*, которые сопровождают процесс проведения программ и суть которых состоит в получении максимально актуальной информации о ходе осуществления программы и оценке этой информации с целью выработки рекомендаций для принятия решений и управления соответствующими процессами. И, наконец, *суммативные эвалюации* или *эвалюации «ex post»* служат для учёта и оценки долгосрочных последствий реализации программ. В этих исследованиях больше внимания уделяется выявлению каузальных зависимостей.

В качестве примера можно привести концепцию эвалюации, разработанную Райнхардом Штокманном, которая используется Федеральным институтом профессионального образования (BIBB) и Институтом развития профессионального образования (ИРПО) для эвалюации «ex post» ряда реализованных германо-российских пилотных проектов в рамках программы «TRANSFORM». Речь здесь идёт об обоснованной теории концепции, в основе которой лежит интегративный методический подход, суть которого можно кратко изложить следующим образом.¹⁹

Для определения сфер анализа используются три различные теоретические модели.

1. На базе исследований течения человеческой жизни (оцениваемый проект грубо сравнивается с ходом человеческой жизни и подразделяется на фазу рождения, фазу реализации и фазу последствий) определяется временная перспектива, которая позволяет устанавливать связи между хронологией развития проекта и предъявляемыми к нему требованиями и, следовательно, дать оценку решениям, принимаемым при планировании и управлении проектом на его различных этапах.

¹⁸ ср. Stockmann 2000 г.: 13 и далее

¹⁹ Подробно см.: Stockmann 1997.

2. Модель эффектов, разработанная на основе теории организации, фокусирует внимание на эффективности организации, являющейся носителем проекта (например, учебного заведения, в котором прошло апробацию какое-то нововведение в системе профессионального образования) и помогает ответить на вопрос о том, как данный проект повлиял на данную организацию в широком смысле слова, какие изменения и последствия повлекла за собой реализация данного проекта.

3. Эффективность организации, являющейся носителем проекта, имеет также решающее значение для дальнейшего распространения внедрённого нововведения / инновации за пределами данной организации. В сфере профессионального образования целью опытно-экспериментальных (пилотных) проектов является достижение определённых эффектов как в самой системе профобразования (например, распространение нововведений из пилотного проекта), так и в системе трудоустройства (например, совершенствование адаптации квалификации к требованиям рынка труда). Такое распространение эффектов проекта, или выражаясь терминологически, диффузия эффектов, также является предметом изучения в рамках концепции эвалюации Штокманна.

Методология концепции Штокманна учитывает целый ряд особенностей эвалюаций. Так, например, за счёт реализации подхода партиципации (соучастия) обеспечивается сочетание опыта и знаний экспертов-эвалюаторов и инсайдерской информации участников проекта. Предпосылкой этого является откровенный обмен информацией и корректное отношение друг к другу в процессе эвалюации

Особые методологические сложности могут быть вызваны недостатками отдельных методов сбора информации, которые в отличие от условий обычного эксперимента не поддаются контролю при проведении эвалюации. В целях нивелирования недостатков определённых методов и наиболее оптимального использования их преимуществ данная концепция эвалюации предлагает применение целого ряда различных способов сбора данных (так называемый мультиметодический подход).

При таком подходе достигается существенное повышение качества используемых данных. Нерешённой, однако, остаётся следующая проблема. Как правило, при проведении эвалюации исследуется лишь небольшое число проектов, по которым имеется довольно большой объём информации. При этом цель эвалюации состоит, в конечном итоге, в получении ответов на множество различных научных вопросов. Ответ на

вопрос о том, являются ли выявленные в данном конкретном случае взаимосвязи чистой случайностью или же следствием систематических и регулярно повторяющихся каузальных причин, возможно дать только на базе адекватных сравнений. Таким образом, интеграция компаративного подхода (подхода сравнений) абсолютно необходима для адекватного выявления проблем и разработки вариантов их решения.

5. Перспективы

Что нового может извлечь для себя Россия из истории эвалюации?

С момента перестройки (с середины восьмидесятых годов) и ещё более интенсивно с начала девяностых годов в России происходят глубокие преобразования во всех областях политической жизни, что затрагивает естественно и сферу образования. Принят новый Закон об образовании, новый Перечень профессий, сформулированы новые характеристики профессий, изменились требования рынка труда к выпускникам учебных заведений, в результате чего возникли новые методы обучения и новые типы школ. В страну были приглашены консультанты из различных стран мира с целью ознакомления российских специалистов с международным опытом в данной области и изучения возможностей его использования. В рамках реализации многочисленных национальных и международных пилотных проектов исследовались и исследуются различные инновации в сфере образования. Один только Федеральный институт профессионального образования совместно с российским партнёром ИРПО осуществил девять частично весьма сложных опытно-экспериментальных проектов, результатом которых явилась модернизация и коренное изменение некоторых курсов профессионального образования.

В настоящий момент в России наступил этап осмысления и переосмысления последствий многих преобразований. Возникла необходимость ответа на вопрос о результативности внедрённых нововведений, а также о том, какие именно консультации оказались полезными, какие именно условия способствовали успеху реализации пилотных проектов, и какие экспериментальные проекты имеет смысл распространять.

Систематический, научно обоснованный и тем самым ясный и понятный для всех участников ответ на подобные или аналогичные вопросы можно получить посредством проведения эвалюаций.

Как уже отмечалось в кратком историческом экскурсе, эвалюация оправдала себя во всех развитых странах Запада как важный инструмент рационального управления в политической сфере и как средство общественной легитимации политических решений, оценки эффектов государственных программ и контроля эффективности использованных финансовых и других ресурсов. Решающее значение при этом имеет обеспечение высокого профессионального уровня проведения эвалюаций, предполагающее как соблюдение международных стандартов так и предотвращение влияний тех или иных заинтересованных групп лиц на результаты эвалюационных исследований. Проведение эвалюаций в соответствии с международными стандартами и безоговорочное раскрытие информации о результатах исследований (даже в том случае, если они не соответствуют желаниям и ожиданиям заинтересованных лиц) способствуют, в конечном итоге, формированию благоприятного для всех климата и атмосферы гласности, доверия и диалога в обществе.

Между российскими учёными в области социальных наук и немецкими экспертами-эвалюаторами уже установлены первые контакты, которые следует углублять и расширять. Российские учёные могут использовать опыт и знания своих зарубежных коллег, в частности, из западноевропейских стран при проведении эвалюаций и создании собственных структур в сфере эвалюационных исследований в России.

Институт развития профессионального образования, уже давно осуществляющий научное сопровождение процессов преобразования в российской системе профессионального образования, в сотрудничестве с BIBB в рамках совместного научно-исследовательского проекта начал накапливать опыт проведения эвалюаций, который может быть расширен путём повышения квалификации кадров в соответствующих общественно-научных областях, а также путём систематического расширения сфер применения эвалюаций. Опытные эксперты-эвалюаторы из стран Запада могли бы провести с этой целью в России соответствующие курсы переподготовки и повышения профессиональной квалификации молодых российских специалистов. Кроме того, могла бы быть использована возможность участия в курсах и семинарах за рубежом, например, в Германии в рамках программы обучения, предлагаемой Центром эвалюационных исследований Саарского университета (CEval) в сотрудничестве с Германским фондом международного развития (DSE) в Маннхайме.

В рамках кооперации Федерального института профессионального образования (г. Бонн), Саарского университета и Института развития профессионального образования (г. Москва) уже проведена первая эвалюация одного германо-российского консультативного проекта, результаты которой вызвали большой интерес эвалюируемого российского учебного заведения и дали новые импульсы совершенствованию его работы. На начальном этапе находится аналогичная эвалюация модернизированной или переработанной в рамках германо-российского сотрудничества коммерческой специальности. Результаты этих эвалюаций приведут, по всей видимости, к оживлению и обогащению дискуссии в сфере профессионального образования в России.

Библиография

- BAUMERT, J. & BOS, W. & LEHMANN, R. (Hrsg., 2000), TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn (2 Bde.), Opladen: Leske+Budrich.
- BAUMERT, J. & KLIEME, E. & NEUBRAND, M. (Hrsg., 2001), PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich, Opladen: Leske+Budrich.
- BERGER, K. & WALDEN, G. (2001), Evaluierung der Bund-Länder-Programme zur Ausbildungsförderung in den neuen Bundesländern 1996 – 1999, Bonn: Bundesinstitut für Berufsbildung
- BORTZ, J. & DÖRING, N. (2002), Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, Berlin u.a.: Springer (3. Auflage).
- EBBINGHAUS, M. & GÖRMAR, G. & STÖHR, A., Evaluert: Projektarbeit und Ganzheitliche Aufgaben: Ergebnisbericht zur Evaluation der Abschlussprüfungen in den vier IT-Berufen, Bielefeld: Bertelsmann
- HADLEY, R. G. & MITCHELL, L. K. (1995), Counseling research and program evaluation, Pacific Groce, CA: Brooks/Cole.
- HÄRTEL, M. & STOCKMANN, R. & GAUS, H. (2000), Berufliche Umweltbildung und Umweltberatung – Grundlagen, Konzepte und Wirkungsmessung, Bielefeld: Bertelsmann
- JOINT COMMITTEE ON STANDARDS FOR EDUCATION EVALUATION, J.R. SANDERS (eds., 2000), Handbuch der Evaluationsstandards. Die Standards des 'Joint Committee on Standards for Educational Evaluation'. Aus dem Amerikanischen übersetzt von Wolfgang Beywl und Thomas Widmer. Deutsche Ausgabe bearbeitet und ergänzt von Wolfgang Beywl, Thomas Widmer und James R. Sanders, Opladen: Leske + Budrich.
- LEEUW, F. L. (2000), Evaluation in Europe, in: Stockmann, R. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske+Budrich, S. 57-76.
- MADDAUS, G. F. & STUFFLEBEAM, D. L. & SCRIVEN, M. S. (1983), Program evaluation: A historical overview, in: Maddaus, G. F., M. S. Scriven & D. L. Stufflebeam (eds.), Evaluation Models, Boston: Kluwer-Nijhoff, pp. 3-22.
- MERTENS, D. M. (2000), Institutionalizing Evaluation in the United States of America, in: Stockmann, R. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske + Budrich, S. 41-56.
- ROSSI, P. H. & FREEMAN, H. E. & HOFMANN, G. (1988), Programm Evaluation. Einführung in die Methoden angewandter Sozialforschung, Stuttgart: Enke. Neueste Englische Auflage: Rossi, P. H., H. E. Freeman & M. W. Lipsey (1999), Evaluation. A Systematic Approach, Thousand Oaks et al.: Sage (6th ed.).
- SHADISCH, W. R. (1994), Need-based evaluation: good evaluation and what you need to know about it, in: Evaluation Practice, 15 (3), pp. 347-358.
- SHADISH, W. R. & NEWMAN, D. & SCHEIRER, M. A. & WYE, C. (eds., 1995), The American Evaluation Association's Guiding Principles, San Francisco: Jossey-Bass.
- STILLER, I. & STÖHR, A. (2001), Ergebnisse aus der Evaluation der Büroberufe, Bielefeld: Bertelsmann.
- STOCKMANN, R. (1997), The Sustainability of Development Cooperation, Baden-Baden: Nomos.
- STOCKMANN, R. (2000), Evaluation in Deutschland, in: ders. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske + Budrich, S. 11-40.
- STOCKMANN, R. (2000a), Evaluation staatlicher Entwicklungspolitik, in: ders. (Hrsg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske + Budrich, S. 375-408.
- WIDMER, T. (2000), Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird, in: Stockmann, R. (Hrsg., 2000), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske+Budrich, S. 77-102.
- WOTTAWA, H. & THIERAU, H. (1990), Lehrbuch Evaluation, Bern u.a.: Huber Verlag.