33. Kongress der Deutschen Gesellschaft für Soziologie
Universität Kassel, 9.-13.Okt. 2006
Sektion Methoden der empirischen Sozialforschung

Wissenschaftstheoretische Anforderungen an empirische Forschung und die Problematik ihrer Beachtung in der Evaluation

Oder: Wie sich die Evaluationsforschung um das Evaluieren drückt

Helmut Kromrey

Vorbemerkungen

Wenn von Evaluation − in welcher Bedeutung des Begriffs auch immer − gesprochen wird, ist damit eine bewertende Aussage über einen Sachverhalt gemeint. Kommt dabei als Erkenntnisinstrument (auch) empirische Forschung ins Spiel, lautet das Ziel von Evaluation: empirisch gestützte Gewinnung von Bewertungen mit intersubjektivem Geltungsanspruch. Zugespitzt auf das Thema dieser Veranstaltung („Die empirische Erfassung von Qualität"): Es geht um die forschungsgestützte Gewinnung von *Qualitäts*aussagen, d.h. um empirisch begründete normative Aussagen, um „Werturteile". Für die „herrschende Meinung" in Methodologie und Wissenschaftstheorie (der analytisch-nomologischen Richtung, auf deren Argumentation ich mich im Folgenden beziehe) manövriert sich die Forschung damit in ein offensichtliches Dilemma: Werturteile sind empirisch nicht begründbar.

Umso mehr überrascht es, dass dies – abgesehen von ganz seltenen Ausnahmen[1] – bei Evaluationen trotz ihrer quantitativ zunehmenden Bedeutung als Problem anscheinend nicht erkannt, zumindest jedenfalls nicht thematisiert wird. Die skizzierte Problematik wirft ein ganzes Spektrum von Fragen auf, die natürlich in einem einzigen Vortrag nicht abgehandelt werden können. Ich beschränke mich hier auf die Wertproblematik im Rahmen von *Evaluation durch Forschung* (ungenauer: „Evaluationsforschung").

Im Rahmen analytisch-nomologischer Wissenschaftstheorie ist − wie bereits angedeutet − der Anspruch, „empirisch gestützte Bewertungen mit intersubjektivem Geltungsanspruch" zu gewinnen, methodologisch nicht ohne Weiteres legitimierbar:

Die unmittelbare empirische Begründung von Bewertungen durch Forschung ist nicht möglich; auch aus korrekten *empirischen* Beschreibungen und Analysen sind *normative* Aussagen nicht ableitbar.

Die Geltungsbegründung empirischer Analysen folgt einer anderen Logik als die Geltungsbegründung normativer Aussagen. Für die ersteren gibt es in der Wissenschaftstheorie klare Regeln, für die letzteren nicht.

Möglich ist es lediglich, die Forschung möglichst genau auf den Zweck „Bewertung" auszurichten, indem wenn schon nicht direkt bewertende, so doch bewertungs*relevante* Informationen gesammelt und systematisiert werden.[2]

Soll also unmittelbar „durch Forschung" evaluiert werden, müssen Strategien gefunden werden, mit deren Hilfe erreichbar wird, dass die *empirischen* Daten einen quasi *normativen* Charakter erhalten, so dass sie „für sich selbst sprechen" können. Eine explizite Geltungsbegründung daraus abgeleiteter Wertaussagen durch die Forschung wäre dann nicht mehr notwendig. Solche Strategien gibt es in der Tat; und drei von ihnen sollen hier skizziert werden.

---

1 Beispielsweise Christian Lüders wie auch Wolfgang Beywl in Flick 2006.

2 Dies geschieht z.B. in der Hochschulevaluation und/oder bei Akkreditierungen nach dem bekannten mehrstufigen Modell des peer review. Die Forschung hat hier nur die Funktion des Informationszulieferers; das Evaluieren (d.h. das Fällen der Werturteile) geschieht durch Experten oder durch ein dazu legitimiertes Gremium oder durch Aushandeln zwischen den beteiligten Parteien.

Exkurs: Die Argumentationslogik analytisch-nomologischer, „wertfreier“ Forschung

Für *wissenschaft*stheoretische Argumentationen kann das von Hempel und Oppenheim konzipierte Schema wissenschaftlicher Erklärung (Hempel/Oppenheim 1948) als verbindliches Gerüst gelten:

Explanans:	(1) Es gilt (mindestens) ein nomologisches Gesetz z.B.: „Wenn A und B, dann C“) (2) Die in der Wenn-Komponente genannten Randbedingungen sind empirisch erfüllt (z.B.: „A und B liegen vor“)
Explanandum:	(3) Singulärer Satz, der den zu erklärenden Sachverhalt beschreibt (z.B. „C liegt vor“).

Gegeben ist das zu erklärende „singuläre Ereignis“(3), *gesucht* ist das „Explanans“ (1 und 2). Bei dieser Art von Erklärung muss (3) deduktiv-logisch aus (1) und (2) folgen, wobei (2) aus der Wenn-Komponente und (3) aus der Dann-Komponente des nomologischen Gesetzes abgeleitet wird. In der *alltäglichen empirischen Forschung*, die sich auf die analytisch-nomologische Wissenschaftstheorie beruft, scheint das H-O-Schema allerdings wenig praktische Bedeutung zu haben. Dieser Eindruck täuscht jedoch.[3]

Zum einen gilt die *Logik der Erklärung* auch für die Konstruktion des Forschungsdesigns zum empirischen *Test von Theorien und Hypothesen*. Lediglich das Erkenntnisinteresse ist ein anderes: Es sind nicht singuläre Ereignisse (3) „zu erklären“, sondern es sind nomologische Hypothesen (1) auf ihre empirische Geltung „zu prüfen“. Hierfür dürfte die skizzierte Logik auch von Forschungspraktikern leicht einzusehen sein. In anderen Forschungszusammenhängen dagegen werden seit Poppers Fokussierung der Methodologie auf die Logik des Testens wissenschaftstheoretische Grundlagen eher stiefmütterlich behandelt.

Bei genauerem Hinsehen können die drei Komponenten des H-O-Schemas aber z.B. auch im deskriptiven Survey-Modell − also bei empirischen Forschungen zum Zwecke deskriptiver Diagnose sozialer Problemfelder − wiedergefunden werden. Als methodologisches Gerüst ist in diesem Design nach Formulierung einer präzisen Fragestellung ein deskriptives Modell des zu untersuchenden Gegenstands auszuarbeiten. Dieses wird idealtypischerweise auf der Basis „empirisch bewährter“ Theorien (1) entwickelt, im Realfall faktischer Forschung ergänzt um Hypothesen möglichst hoher Plausibilität. Dieses Modell hat eine forschungsleitende Funktion,

3 Ausführlicher dazu Kromrey 2006, S. 87 ff.

dient sozusagen als „Wegweiser" im Forschungsprozess, ist also Erkenntnis*basis* und nicht Gegenstand der Überprüfung. Da jedoch die erhobenen Daten der empirischen „Diagnose" sozialer Probleme dienen sollen, müssen sie „Erklärungswert" haben und sowohl die Problemdimensionen differenziert beschreiben (3) als auch die relevanten Randbedingungen (2) für das Auftreten der Probleme erfassen.

Zu bedenken ist, dass das H-O-Erklärungsschema ebenso wie die Übernahme seiner Struktur für andere Erkenntniszwecke als *erkenntnistheoretische* Basis den erkenntnistheoretischen Realismus impliziert: Unterstellt wird auf der *Gegenstandsseite* eine „real existierende" Welt, gekennzeichnet durch Merkmale wie Ordnung, Struktur und Tatsachenautonomie, die Geltung von Regelhaftigkeiten bzw. Gesetzmäßigkeiten und Kausalität. Unterstellt wird zudem auf der *Seite des erkennenden Subjekts* die prinzipielle, wenn auch möglicherweise unvollständige und teilweise fehlerbehaftete Erkennbarkeit dieser Realität durch Wahrnehmungssinne sowie unterstützende Instrumente.

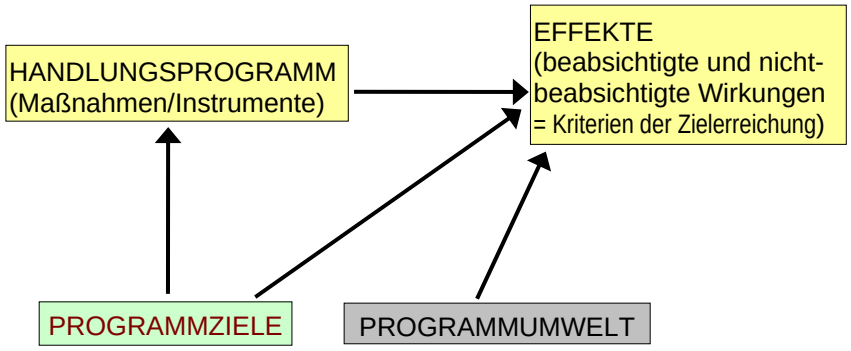
Das *Ziel* empirischer Wissenschaft ist hier die Erkenntnis der „wahren" Strukturen und Gesetzmäßigkeiten der Realität sowie ihre Dokumentation in Theorien. Erreicht werden soll dies durch eine *Strategie* des „kontrollierten Ratens" über das Aufstellen erkenntnisleitender ex-ante-Hypothesen und deren Konfrontation mit der (*objektiven*) Realität, abgebildet in (*subjektiven*) Wahrnehmungsdaten. Die darin implizierte Subjektivität wird kontrolliert durch strenge methodologische Regeln („Objektivierung" der Verfahren, intersubjektive Nachprüfbarkeit).

Eine besondere Bedeutung kommt hier der Trennung deskriptiver (und damit „objektivierbarer") von normativen Aussagen zu, deren subjektiver Charakter methodologisch nicht aufhebbar und deren intersubjektive Geltung daher mit empirischen Mitteln nicht begründbar ist. Die normativen Elemente der Erkenntnisgewinnung werden daher aufgeteilt in solche, die die *normative Basis* der Forschung bilden (der Forschung vorgelagerte *wissenschaftsimmanente* Werte), und die übrigen, d.h. die *nichtwissenschaftlichen* Interessen und Werte, die aus dem wissenschaftlichen Begründungskontext ausgelagert und in den Entstehungs- und Verwertungskontext des Forschungsprojekts verwiesen werden.

In genau dieser pragmatischen Strategie der „herkömmlichen empirischen Forschung" zur Handhabung des Werturteilsproblems wird auch in der Evaluationsforschung die Lösung des Bewertungsdilemmas gesucht, wenn auch in unterschiedlichen Varianten.

„Wertneutrale Evaluation“ im Design der Programmevaluation: der methodologische Idealtypus

So ist die skizzierte Forschungslogik einer auf Wertfreiheit verpflichteten Wissenschaft auf die Evaluationsforschung bruchlos anwendbar, sofern es sich bei ihrem Gegenstand um ein ausgearbeitetes „Programm“ handelt, das in Form explizit ausformulierter Programmziele die für eine „empirische Bewertung“ notwendige normative Basis bereits in die Evaluation mitbringt. Das Design der dieses Evaluationstyps berücksichtigt drei Dimensionen des zu bewertenden Gegenstands − Ziele, Maßnahmen, Effekte − sowie die programmexterne Umwelt als Quelle möglicher „Störvariablen“:



Man kann leicht erkennen, dass auch dieses Konzept die *Hempel-Oppenheimsche* Logik verwendet.

Zum einen hat schon das „*Programm*“ als „technologische Aussage“ formal die gleiche Struktur wie eine „Erklärung“: Programmziele als angestrebte künftige Situation (Punkt 3), „Maßnahmen“ als Eingriffe in die gegenwärtigen „Randbedingungen“ (Punkt 2) sowie die theoretische Basis für die Art und Weise des Eingreifens (Punkt 1 im H-O-Schema).

Aber auch das Design der Evaluation orientiert sich an diesen Komponenten:

- Sowohl die existierenden Randbedingungen (2) als auch der Ist-Zustand der Zielvariablen (3) sind *vor* Programmbeginn – t_0 – empirisch zu beschreiben.
- Während der Programmlaufzeit sind die Veränderungen der Randbedingungen (2) zu erfassen („monitoring“ sowohl der Maßnahmen als auch der Veränderungen in der Programmumwelt).
- Schließlich ist sicherzustellen, dass der Zustand der Zielvariablen (3) nach Programmdurchführung – t_1 – wiederum empirisch beschrieben wird, so dass Art und Ausmaß der Veränderungen feststellbar sind.

Der Gegenstand, den es zu bewerten gilt, ist natürlich *als Gegenstand* nicht wertneutral oder zweckfrei. Ganz im Gegenteil: Das Programm soll etwas erreichen. Damit wird auch das Konzept „Evaluation als Programmwirkungsforschung" mit dem Wertproblem konfrontiert. Für die Forschung ist es aber dadurch gelöst, dass es in den „Entstehungskontext" verlagert wurde (Programmziele als normative Basis), wodurch die eigentliche „Evaluation" einen deskriptiven (und somit „wertneutralen") Charakter erhält: Der Bewertungsprozess reduziert sich damit auf einen Vergleich der vom Programm gesetzten Sollwerte (Zielerreichungskriterien) mit den gemessenen (und den Maßnahmen zurechenbaren) Effekten im Wirkungsfeld des Programms. Solche Aussagen lassen sich in vollem Umfang auf die Logik der Geltungsbegründung empirischer Faktenbehauptungen stützen.

Dass die Realisierung dieser Aufgaben riesige Schwierigkeiten bereitet, tut der überzeugenden *Logik* des Modells keinen Abbruch.

Anders fällt das Urteil aus, wenn es um die praktische Bedeutung für den „alltäglichen" Evaluationsbedarf geht. Die methodologischen Schwierigkeiten sind nämlich so riesig, dass diese überzeugende Logik nur unter sehr einschränkenden, ganz selten erfüllbaren Bedingungen praktisch einsetzbar ist. Überwiegend muss die Programmevaluationsforschung mit mehr oder weniger guten Annäherungen an das idealtypische Modell zufrieden geben. Es liegt also nahe, nach komplett andersartigen Ersatzlösungen zu suchen.

Eine solche Ersatzlösung folgt aus der Überlegung, „Qualität" nicht erst anhand von „Effekten", also als *Folge* der Eigenschaften des zu bewertenden Gegenstands oder Sachverhalts zu interpretieren, sondern *unmittelbar* zu messen.

Evaluation als Qualitätsmessung: das methodologische Problemkind

Wenn es nämlich gelänge, am zu evaluierenden Gegenstand Qualitätsmerkmale zu bestimmen und präzise zu definieren, hätten wir einen direkten Weg, die Evaluation methodologisch zu „okjektivieren", d.h. am zu bewertenden „Objekt" festzumachen. Wir könnten uns damit zugleich den schwierigen Umweg über die Messung von Outcome-Variablen und die methodisch äußerst schwierige Zurechnung ihrer Veränderungen als Zielerreichungs-Kriterien ersparen. Die implizite Annahme bei dieser Überlegung ist: Wenn die Qualität des zu bewertenden Sachverhalts hoch ist, dann werden auch seine Wirkungen positiv sein; bzw.: dann werden auch die mit ihm verknüpften Ziele erreicht werden. In diesem Fall hätte die Forschung lediglich die – forschungsmethodisch zur Alltagsroutine

zählende – Aufgabe zu erfüllen, „Qualität" durch einen Satz qualitätsrelevanter Merkmale auszudifferenzieren und durch geeignete Indikatoren so zu operationalisieren, dass an ihnen situationsunabhängige „Qualitätsmesswerte" abgelesen werden können.

Auch bei dieser Strategie wird die normative Basis der Evaluation in den Entstehungskontext ausgelagert, in dem der normative Begriff „Qualität" von dazu legitimierter Seite festzulegen ist.

Methodologisch ist die o.g. Aufgabe allerdings gar nicht so simpel, wie es auf den ersten Blick erscheinen mag. Die „Sozialindikatorenbewegung" in den 1970er Jahren hat sich damit intensiv auseinandergesetzt und eine Liste von Anforderungen an Indikatorensysteme formuliert, [4] die auch in unserem Fall Geltung beanspruchen kann.

Grundlegender ist jedoch eine andere Problematik, die in dieser Argumentation häufig übersehen wird: Neben dem Zutreffen der o.g. impliziten Annahme eines direkten Zusammenhangs zwischen Qualitätsmerkmalen und Zielerreichung (was empirisch geprüft werden kann) muss nämlich eine weitere, eine erkenntnistheoretische (und damit *axiomatische*) Voraussetzung als erfüllt angesehen werden: Qualität muss als direkte Eigenschaft des Objekts verstanden werden können (wie etwa Größe, Gewicht, Farbe usw.); bzw. methodologisch: Das Konstrukt „Qualität" ist so zu definieren, dass seine Dimensionen als Merkmale des Gegenstands erscheinen. Schon eine oberflächliche semantische Analyse lässt erkennen, dass „Qualität" eben nicht als direktes Merkmal(sbündel) des zu bewertenden Objekts zu verstehen ist, sondern als ein *relationales* Merkmal: als Eignung, Brauchbarkeit, Güte in Bezug auf bestimmte Ziele und Zwecke sowie auf bestimmte Nutzer- und Klientengruppen.

Damit (und wenn wir die methodologischen Anforderungen der Sozialindikatorenbewegung einbeziehen) haben wir aber wieder das Anforderungsniveau des Modells der Programmevaluation erreicht − ohne allerdings über deren Problemlösung zu verfügen, nämlich: vom Programm vordefinierte Ziele und Maßnahmen. Anders als zu Beginn postuliert, bleibt − wenn sich das Konstrukt „Qualität" nicht aus dem Objekt selbst herleiten lässt − damit die Wert-Entscheidungs-Frage unbeantwortet. Die *normative Basis* für die Bewertung durch Qualitätsindikatoren setzt (und das heißt: die eigentliche Evaluation betreibt) diejenige Instanz, die festlegt, was als „Qualität" gelten soll und welche Qualitätskriterien und –standards anzulegen sind. Und wenn diese Instanz Nebenziele verfolgt wie: a) das ganze Evaluationsverfahren solle möglichst ohne großen Ressourceneinsatz zu bewerkstelligen sein und/oder b) die Resultate dürften nicht noch interpretationsbedürftig, sondern sollten selbsterklärend sein (d.h.: höherer

4 s. z.B. Werner 1975.

Messwert = „mehr Qualität"), dann wird leicht in den Rang von Qualitätsindikatoren das erhoben, worüber Daten zur Verfügung stehen (d.h.: Qualität ist, was leicht messbar ist). Und dann werden vor allem quantitativ messbare Merkmale herangezogen (d.h.: Qualität wird operationalisiert durch Quantität).

Angesichts dieser erneuten Problematik verwundert es nicht, wenn die Forschung sich aus diesen Dilemmata zu befreien versucht, indem Sie Evaluation auf das reduziert, was sie unbestritten kann: Befragungen durchführen.

Evaluation durch Befragung: das erkenntnistheoretische Problemkind

Statt aufwändige, methodisch kontrollierte Evaluation durch Programmforschung zu betreiben (deren Anwendungsvoraussetzungen selten erfüllbar sind) oder Qualitätsindikatoren zu messen (deren Gültigkeit fragwürdig ist), wird die Bewertung von Maßnahmen (Sachverhalten, Dienstleistungen) per „Betroffenenbefragung" ermittelt. Die Adressaten und Nutzer, die Kunden und Klienten sind – so wird argumentiert – die von den zu evaluierenden Leistungen ganz konkret „Betroffenen" und daher in der Lage, aus eigener Erfahrung auch deren Qualität sachverständig und zuverlässig zu beurteilen. Befragungen erscheinen erheblich weniger anspruchsvoll – sowohl hinsichtlich des Aufwands der Durchführung als auch hinsichtlich der Strategie der Objektivierung: Sind die erbrachten Dienstleistungen „schlecht", so werden auch die Beurteilungen auf einer vorgegebenen Skala negativ ausfallen und umgekehrt. Befragt man eine hinreichend große Zahl von „Betroffenen" und berechnet pro Skala statistische Kennziffern (etwa Mittelwerte oder Prozentanteile), dann kommen – so die weitere Argumentation – individuelle Abweichungen der einzelnen Urteilenden darin nicht mehr zur Geltung.

Sofern dies zuträfe,[5] wäre die Evaluation per Befragung der Königsweg zur Lösung aller Probleme der Evaluationsforschung − auch der Werturteilsproblematik, denn die Bewertungen nehmen hier die „per Betroffenheit dazu Legitimierten" vor. Die Forschung selbst bliebe neutral; denn sie erhebt, systematisiert und analysiert lediglich.

Zwar sind „Messungen" per Befragung nicht so problemlos wie dies dem Laien häufig erscheint (s. Kromrey 2006, S. 257 ff.). Doch sofern systematische

5 Leider ist dieser Optimismus bei Lehrevaluationen nicht gerechtfertigt, wie differenzierte statistische Analysen von Daten aus Veranstaltungsbefragungen belegen (s. z.B. Kromrey 1994, 1995).

Verzerrungen vermieden werden können, lassen sich in der Tat bei hinreichend großer Befragtenzahl und bei repräsentativer Datenbasis individuelle Unterschiede − wie im obigen Zitat behauptet − „herausmitteln". Einzulösen sind hierfür lediglich durch das Erhebungsinstrument und in der Befragungssituation einige formale methodologische Voraussetzungen − die allerdings nicht ohne weiteres als erfüllt gelten können:

- der „Gegenstand" (das Objekt) der Beurteilung ist eindeutig definiert,
- das zu messende „Merkmal" ist eindeutig definiert und operationalisiert,
- eine „Mess-Skala" existiert und ist eindeutig definiert,
- die Befragten sind in der Lage, den „Gegenstand" intersubjektiv übereinstimmend zu identifizieren, das zu messende „Merkmal" intersubjektiv übereinstimmend zu erkennen und die „Mess-Skala" in intersubjektiv übereinstimmender Weise anzuwenden.

Im Falle der Erhebung von Evaluationen wird die Situation zusätzlich dadurch schwieriger, dass es sich bei den zu messenden Merkmalen um die oben genannten „qualitätsrelevanten Merkmale" (oder „Qualitätskriterien") handelt, durch die der Begriff „Qualität" operationalisiert wird. Und die hierauf anzuwendende Mess-Skala ist die Bezugsgröße, auf der das „Ausmaß" von Qualität angebbar ist (also der „Qualitätsstandard"). Damit sind wir aber auch bei der *indirekten* Qualitätsmessung per Befragung wieder mit dem gleichen Problem konfrontiert wie beim Ansatz der *direkten* Messung von Objektqualität durch Indikatoren.

Das statistische „Ausmitteln" von Messungenauigkeiten setzt bekanntlich die Existenz eines „wahren Wertes" voraus, von dem die einzelnen Messwerte lediglich „zufällig" abweichen. Bezogen auf die Beurteilungsvariation zwischen den einzelnen Befragten heißt dies: Um auf diese Weise zu einem gültigen Qualitätsmaß zu kommen, muss die Annahme gerechtfertigt sein, dass es einen „wahren" Qualitätswert für den zu beurteilenden Sachverhalt gibt, um den die einzelnen Antworten „zufällig" streuen. Diese Annahme wäre aber nur dann haltbar, wenn eines der beiden folgenden erkenntnistheoretischen Axiome zuträfe:

Alternative 1: Qualität ist ein „objektives" Merkmal eines Sachverhalts,[6] dessen Ausprägung durch abbildende subjektive Wahrnehmung ohne systematische Verzerrung „gemessen" werden kann, so dass bei hinreichend großer Zahl von Messungen der Erwartungswert dem „*wahren objektiven Wert*" entspricht. 0der:

Alternative 2: Qualität ist ein „intersubjektiv gültiges" Konzept, über das alle Menschen in gleicher Weise verfügen. Anders formuliert: Alle Menschen bewerten nach gleichen Kriterien und Standards in gleicher Weise; in konkreten Situationen

6 Dies ist die identische Voraussetzung, die auch im Konzept der direkten Qualitätsmessung durch Indikatoren erfüllt sein muss.

auftretende Unterschiede zwischen Bewertern sind als Zufallsvariation anzusehen, so dass bei hinreichend großer Zahl von Messungen der Erwartungswert dem „*wahren subjektiven Wert*" entspricht.

Grundlage für die erste Alternative ist die im erkenntnistheoretischen Realismus (vom frühen Empirismus bis zum Gründer des Wiener Kreises, Moritz Schlick) vertretene Überzeugung von der Möglichkeit abbildender Wahrnehmung der Realität: Das Wahrgenommene steht in einem genauen Entsprechungsverhältnis zum Wirklichen. In diesem Fall wäre „Qualität" durch standardisierte Befragung „objektiv" messbar.

Grundlage für die zweite Alternative wäre der erkenntnistheoretische Idealismus, am kompromisslosesten konzipiert in Platons „Ideenlehre": Hinter der sinnlich wahrnehmbaren Welt stehen (als das „in Wahrheit Seiende") die „Ideen", die zwar der direkten Wahrnehmung nicht zugänglich, aber der unsterblichen Seele des Menschen von Anfang an mitgegeben sind. Erkenntnis besteht nach dieser Vorstellung im *Wiedererkennen* der allgemeingültigen Konzepte (der „Ideen") in den empirischen (Einzel-)„Erscheinungen". Sofern „Qualität" der Status einer solchen „Idee" zukäme (analog zu Gerechtigkeit, Gleichheit, Heldentum, Liebe etc.), wäre sie durch standardisierte Befragung „intersubjektiv" messbar.

Fazit

Die Versuche, das Wertproblem der Evaluationsforschung dadurch zu entschärfen, dass man die für das „wissenschaftliche Evaluieren" erforderliche Wertbasis aus dem Begründungskontext empirischer Forschung hinausverlagert, erscheinen mir wenig überzeugend. Aus meiner Sicht bieten sich in dieser Situation zwei Alternativen für eine „wissenschaftliche Evaluation" an.

Die eine bestünde darin, die Evaluation als einen „Sonderfall" aus dem Aufgabengebiet einer wertneutral verfahrenden empirischen Forschung auszusondern und ihr die zusätzliche Aufgabe der (nach wissenschaftlicher Methodologie verfahrenden, intersubjektiv nachprüfbaren) Ableitung von Wertaussagen zuzuschreiben. Überlegungen in dieser Richtung werden von Christian Lüders (2006) angestellt. Eine überzeugende Methodologie ist allerdings derzeit nicht erkennbar.

Die andere Alternative wäre − und diese halte ich für die angemessenere Variante −, Evaluieren und Forschen klar zu trennen. Der Forschung ist die Aufgabe zuzuschreiben, alle für die Bewertung von Programmen, Maßnahmen etc. relevanten Informationen unter Einsatz des bewährten empirischen Instrumentariums zu erheben, zu analysieren und für Bewertungs- und Entscheidungsprozesse aufzubereiten. Die Funktion des Evaluierens sowie der Ableitung

möglicher Konsequenzen für das Evaluationsobjekt sollte dagegen einem dafür explizit legitimierten Gremium zugewiesen werden. Dass dieses Modell realisierbar ist und die Akzeptanz von Evaluation erhöht, zeigt das bereits angesprochene Modell der mehrstufigen Hochschulevaluation, beispielsweise in der vom Verbund Norddeutscher Universitäten praktizierten Variante: Selbstbeschreibung / Selbstevaluation – peer review − Auswertende Konferenz (Nordverbund 2004).

Literatur:

Beywl, Wolfgang (2006), »Evaluationsmodelle und qualitative Methoden«, in: Uwe Flick (Hg.), *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*, Reinbek bei Hamburg, S. 92-116

Hempel, Carl G.; Oppenheim, Paul C. (1948): »Studies in the Logic of Explanation«; in: *Philosophy and Science*, Vol. 15, S. 135-175

Kromrey, Helmut (1994), »Wie erkennt man "gute Lehre"? Was studentische Vorlesungsbefragungen (nicht) aussagen«, in: *Empirische Pädagogik*, 1994/2, S. 153-168

ders. (1995), »Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern«. In: Peter Ph. Mohler. (Hg.): *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*, Münster, S. 105-128

ders. (2006), *Empirische Sozialforschung. Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung*, 11. Aufl., Stuttgart, utb 1040

Lüders, Christian (2006), »Qualitative Evaluationsfiorschung – Was heißt hier Forschung?«, in: Uwe Flick (Hg.), *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*, Reinbek bei Hamburg, S. 33-62

Verbund Norddeutscher Universitäten (Hg.) (2004), *10 Jahre Evaluation von Studium und Lehre.* Verbund-Materialien Band 16, Hamburg

Werner, Rudolf (1975), *Soziale Indikatoren und politische Planung. Einführung in Anwendungen der Makrosoziologie*, Reinbek: rororo

Helmut Kromrey, Prof. Dr. rer.pol., Universitätsprofessor i.R. (Soziologie und Empirische Sozialforschung) an der Freien Universität Berlin und Adjunct Professor of Sociology der Graduate School of Management der Universität Educatis (Altdorf/Schweiz), Arbeitsschwerpunkte: Methoden/Methodologie der empirischen Sozialforschung, Evaluation. URL: www.hkromrey.de