

Vorwort

Der hier vorgelegte Band zur Evaluation bietet eine Auswahl der wichtigsten Beiträge der amerikanischen und englischen Evaluationsliteratur der letzten Jahre (1967–1972). In diesem Zeitraum entstand in den USA auf dem Hintergrund der gewaltigen Anstrengungen zur Reform des Bildungswesens in den sechziger Jahren ein neues und komplexes Verständnis von Evaluation, das sich von den vorangegangenen Ansätzen durch eine größere Vielfalt der Zielsetzungen, Verfahren und Methoden unterscheidet. Im Rahmen dieser Entwicklung wurde Evaluation allmählich zu einem zentralen Bereich angewandter pädagogischer Forschung. Die führenden erziehungswissenschaftlichen Zeitschriften brachten Sonderhefte heraus, die sich mit Evaluation befaßten. Evaluation wurde zu einem integralen Bestandteil der gegenwärtigen Innovationen im Bereich der Curriculumentwicklung, in der Lehrerbildung und im Feld der Schulversuche. Das entscheidende Anliegen bestand darin, herauszufinden, inwieweit die unter großem Kostenaufwand durchgeführten Projekte zu einer tatsächlichen Verbesserung der Curricula und darüber hinaus der Schulwirklichkeit führen. Diese Frage wurde um so drängender, als die Bildungsreformer sich von seiten der Öffentlichkeit einem verstärkten Druck zur Rechtfertigung der Reformen ausgesetzt sahen. Die Rechtfertigung sollte sich einmal auf die Wirksamkeit der Reformen im Hinblick auf die intendierten Ziele, zum anderen wegen der knapper werdenden Ressourcen auch auf die Relation zwischen finanziellem und personellem Aufwand und der tatsächlich erreichten Verbesserung der Schulwirklichkeit beziehen. Diese Forderungen nach genauer Beschreibung der Innovationen und ihrer Bewertung fanden auch eine starke Resonanz, weil sich immer stärkere Zweifel an der lückenlosen Vorausplanbarkeit von Innovationen mit Hilfe vorgängig entwickelter Theorien einstellten – nicht zuletzt auf Grund der Erfahrungen mit den umfangreichen Reformen (vgl. z. B. Schwab 1970, 1971; Eisner 1971). Somit wurde Evaluation ein zentrales Element der Entwicklungsprozesse für Reformprojekte selbst.

Evaluation erfolgte in einer fortwährenden Überprüfung der Realisierung von Intentionen bzw. normativen Kriterien und als Entwicklung von Alternativen auf Grund unbefriedigender Ergebnisse, wobei die Modifikationen sich häufig auf die Mittel der Realisierung, oft genug aber auch auf die Intentionen selbst bezogen.

Im ersten Teil dieses Auswahlbandes (I) soll ein Überblick über den gegenwärtigen Diskussionsstand im Bereich der Evaluation gegeben werden (vgl. auch Wulf 1972c). Dabei wird sich zeigen, daß die wesentlichen Beiträge zur Konzeptualisierung von Evaluation im Rahmen der »Evaluation von Curricula« entstanden sind. Zum Teil erfolgt dabei eine erhebliche Ausweitung des Begriffs Curriculum, der zum Synonym für Bildungsprogramm wird und der damit als ein Oberbegriff für alle kodifizierten Bildungsintentionen dient. In diesem Teil wird zugleich eine Einführung in die Gesamtproblematik der Evaluation gegeben, die das Verständnis der weiteren Beiträge des Sammelbandes erleichtern soll.

Im zweiten Teil (II) sollen in erster Linie die Grundfragen erörtert werden, die Einfluß auf die Ziele, Formen und Verfahren der Evaluation haben und die z. T. ganz neue Aspekte für den Bereich liefern, der in der BRD häufig Begleitforschung genannt wird. Im Unterschied zur Beurteilung individueller Schülerleistungen zielt Evaluation auf die Beschreibung und Beurteilung von Bildungsprogrammen, deren Verbesserung eine entscheidende Voraussetzung für relevante Schülerleistungen ist. Die starke Konzentration auf die Fragen der Evaluationsforschung hat in den USA dazu geführt, daß Evaluation bereits in der Planungsphase größerer Curriculumprojekte und Schulversuche als »Evaluation der Voraussetzungen« bzw. als »Kontextevaluation« eine wichtige Rolle spielt. Dabei ist es Aufgabe der Evaluatoren, die Bedürfnisse der Adressaten zu erheben, damit eine bessere Planung von Innovationen stattfinden kann. Eine weitere wichtige Funktion kommt der Evaluation während der Entwicklung von Reformprojekten als »formative Evaluation« zu; auch ihr wurde in letzter Zeit verstärkte Aufmerksamkeit zugewandt. Sie soll die Mängel der Reformprojekte (Curricula) so rechtzeitig aufzudecken versuchen, daß sie noch vor Abschluß des Projekts beseitigt werden können. Schließlich gilt es, neue Funktionen und neue Verfahren für die »Evaluation der Endprodukte der Curricula bzw. Schulversuche« zu entwerfen und zu entwickeln. Allein im Zusammenhang mit diesen drei Funktionen der Evaluation

- in der Planungsphase,
- in der Realisierungsphase,
- in der Überprüfungsphase der Ergebnisse

tauchen zahlreiche Fragen auf, die oft schwer lösbar sind. So ist es z. B.

schwierig zu klären, welches die speziellen Aufgaben der »Evaluation der Voraussetzungen« sind oder welche Untersuchungsverfahren man für die »formative Evaluation« wählen soll. Dabei spielt eine zentrale Rolle die Frage, ob man sich für die Vergleichsgruppenuntersuchungen oder Detailuntersuchungen einzelner Gruppen entscheiden soll. In diesem Zusammenhang wird immer wieder das Problem erörtert, ob nicht die Funktion und die Verfahren gegenwärtiger pädagogischer Forschung im Rahmen der Durchführung von Bildungsreformen neu definiert werden müssen, damit man den komplexen Bedingungen pädagogischer Innovationen gerecht werden kann. Mit der Frage nach der Funktion der Evaluation erhebt sich zugleich das zentrale Problem, inwieweit Evaluation eine *Beschreibung* von pädagogisch relevanten Sachverhalten ist oder in welchem Maß sie zur *Bewertung* der Reformen und ihrer Ergebnisse beiträgt oder inwieweit Evaluation als *Entscheidungsvorbereitung* und Entscheidungshilfe für Entscheidungsträger dienen kann.

Es wird sich zeigen, daß das »Konzept der Evaluation« in hohem Maße ideologieanfällig ist. Auf der einen Seite läuft Evaluation Gefahr, als *Auftragsforschung* so verwendet zu werden, daß bildungspolitische oder pädagogische Entscheidungen im nachhinein durch wissenschaftlich gewonnene Ergebnisse gerechtfertigt werden können. Auf der anderen Seite kann man durch »Evaluation« radikal jedes Ergebnis unter Ideologieverdacht stellen und somit fragwürdig machen.

Man wird die Rolle der Evaluation in jedem Fall zwischen diesen Extremen klar definieren müssen. Dementsprechend werden auch die für die Evaluation verwendeten Verfahren empirisch bzw. hermeneutisch-ideologiekritisch sein müssen; hermeneutische bzw. ideologiekritische Verfahren sind im Rahmen der amerikanischen Evaluationsliteratur bisher wenig ausgearbeitet worden.

Die Frage nach den Verfahren und Methoden der Evaluation wirft weiter das Problem auf, welche Form und welches Ausmaß der Evaluation im Hinblick auf die beteiligten Personen und den Anteil der Kosten im Verhältnis zum Gesamtprojekt zukommen soll. In einigen Fällen kann es – wie z. B. in den USA und in England – sinnvoll sein, im Rahmen eines Großprojekts ein Evaluationsteam auszugliedern; es kann aber auch Evaluation ohne organisatorische Ausgliederung im Rahmen der Entwicklung des Projekts durchgeführt werden (vgl. z. B. Meyer 1972); in einigen Fällen wird Evaluation möglicherweise sogar von allen am Versuch Beteiligten zur Verbesserung ihrer eigenen Arbeit durchgeführt; dann ist ein Ziel dabei, die beteiligten Lehrer und Wissenschaftler zur *Selbstevaluation* zu bringen. So wichtig diese Form der Evaluation ist, so begrenzt ist hier die Möglichkeit, Erkenntnisse zu generalisieren.

In der Frage der *Generalisierbarkeit* liegt ein weiteres Problem der Evaluationsforschung, das wissenschaftstheoretisch schwierige Fragen aufwirft, sobald der klassische Versuchsplan mit Versuchs- und Kontaktgruppen aufgegeben wird. Dies sind nur einige der zahlreichen Probleme, die im Rahmen der Grundfragen der Evaluation behandelt werden und die einen wichtigen Beitrag zu einem komplexen Verständnis für die Funktionen von Begleitforschung bilden.

Im dritten Teil (III) wird mit der Evaluation von *Unterrichtsprozessen* bzw. *-interaktionen* ein weiteres Feld pädagogischer Evaluation behandelt. Es ist bisher unter evaluativer Fragestellung nur z. T. erforscht worden; dennoch gibt es zahlreiche Unterrichtsbeobachtungssysteme und -untersuchungen, die für eine Evaluation von Unterrichtsprozessen relevante Forschungsergebnisse erbracht haben. Dieser Bereich ist besonders wichtig, da zahlreiche Forschungen ergeben haben, daß viele Innovationsversuche bereits infolge traditioneller »Interaktionspattern« an enge Grenzen der Wirksamkeit stoßen, die, wenn sie durch eine Evaluationsuntersuchung festgestellt werden, systematisch abgebaut oder verändert werden können. Die Evaluation von Unterrichtsprozessen ist in letzter Zeit in stärkerem Maße Teil der Evaluationsforschung geworden, die zuvor auf formative oder summative Evaluation im Sinne von Leistungsüberprüfung eines Bildungsprogramms begrenzt war. Das wurde durch die Tradition der verhaltensorientierten Ansätze verursacht, nach denen nur das meßbare Ergebnis von Unterrichtsprozessen, nicht aber ihr Prozeßcharakter selbst Gegenstand der Evaluation war. Unter dem Einfluß verstärkter Kritik im Rahmen der Evaluationsliteratur an rigiden behavioristischen Modellen, bei denen die Überprüfung der Lernerfolge (verstanden als das Erreichen der vorher formulierten Lernziele) oft mit Evaluation verwechselt wird, hat die Prozeßdimension im Unterricht und damit auch in der Evaluation mehr Beachtung gefunden. Dennoch sind gerade in diesem Bereich der Evaluation noch viele Fragen offen, die weiterer Klärung bedürfen.

Schließlich galt es, einige ausgewählte Beispiele für Evaluationsuntersuchungen in den Band mitaufzunehmen. Im Teil IV sollen konkrete Fälle für die Bewältigung einiger der angesprochenen Probleme der Evaluation in einem bestimmten pädagogischen Kontext geboten werden, bei denen auch manche Unzulänglichkeiten dieser Untersuchungen sichtbar werden, die zur Kritik und Verbesserung herausfordern können. Zudem werden bei einer solchen Konkretisierung Probleme sichtbar, die in die theoretischen Reflexionen noch nicht eingehen oder eingegangen sind, die jedoch für die Planung von Evaluationsuntersuchungen von außerordentlicher Wichtigkeit sind. Das dürfte bei den unterschiedlichen methodi-

schen Ansätzen der einzelnen Evaluationsuntersuchungen besonders deutlich werden.

Von deutscher Seite aus gesehen, verkürzen die angelsächsischen Evaluationsmodelle und -untersuchungen die politische Dimension der Evaluation und kommen daher in Gefahr, die technologische Dimension zu sehr in den Vordergrund zu rücken, inhaltlich aber vor einer eindeutigen Wert- und Kriteriensetzung zurückzusehen. Daher wurde für Teil V ein Beitrag aus der BRD ausgewählt, der in knapper Form neben vielen anderen zum Teil bereits angesprochenen Fragen die *gesellschaftspolitische Dimension* der Evaluation an Hand eines Beispiels aus der Vorschulerziehung durch das Engagement für die Emanzipation von Unterschichtkindern deutlich macht.

Für alle, die sich eingehender mit Evaluation, vor allem im curricularen Bereich, befassen wollen, dürfte die speziell für die deutschen Verhältnisse überarbeitete klassifizierte Bibliographie von Wert sein (VI).

Um das Lese-Verständnis zu erleichtern, wurde ein Glossar (VII) erarbeitet, das dem Leser bei schwierigen Begriffen eine erste Orientierung geben soll.

Schließlich sei den zahlreichen amerikanischen Kollegen gedankt, die mich im Verlauf von zwei längeren Forschungsaufenthalten in den USA bei der Gestaltung dieses Bandes beraten haben. Mein Dank gilt auch den zahlreichen Mitarbeitern des Deutschen Instituts für Internationale Pädagogische Forschung, die bei der Übersetzung der Beiträge mitgearbeitet haben und die mich bei Überwindung der manchmal erheblichen sprachlichen Schwierigkeiten unterstützt haben. Ferner möchte ich G. Skoupil und H. Kohl für das wiederholte Schreiben der Manuskripte danken.

Christoph Wulf

Anmerkung:

Im allgemeinen sind schwierige amerikanische Begriffe, wenn sie einen bestimmten Eigenwert haben, bei der ersten Verwendung in Klammern hinzugefügt worden.

Die kursiv gedruckten Seitenzahlen in den bibliographischen Angaben beziehen sich auf diesen Auswahlband.

I Einführung in die Problematik der Evaluation, dargestellt am Beispiel der Curriculumevaluation

CHRISTOPH WULF

Curriculumevaluation

I. Die Notwendigkeit der Curriculumevaluation¹

In der Bundesrepublik findet heute Curriculumentwicklung an vielen Orten, zugleich aber auf verschiedenen Ebenen statt:

- Im Rahmen der *Lehrplanrevisionen und Richtlinienarbeit*, die von den Ländern betrieben wird²,
- in vielen unmittelbar auf die Schule bezogenen *Initiativgruppen*, die an traditionellen Schulen und Gesamtschulen gebildet werden³,
- als sorgfältige Entwicklung einiger *Teilcurricula*, wie sie an mehreren Universitäten und Forschungsinstituten geplant oder bereits erfolgt ist⁴.

Diese Bemühungen sind unterschiedlich zu bewerten, was ihren jeweiligen theoretischen Ansatz, den erreichten Grad der theoretischen Reflexion und die Qualität der verschiedenen Ergebnisse betrifft. Bereits in diesem für die BRD noch frühen Stadium der Curriculumentwicklung sollte ein weiterer Schritt vollzogen werden. Man sollte die Aufgaben und Ziele der Curriculumevaluation reflektieren und für ihre Realisierung ein methodisches Instrumentarium entwickeln. Die systematische Erweiterung der Curriculararbeit in dieser Hinsicht ist notwendig: ein neues Curriculum, über dessen Wert nichts zu erfahren ist, erscheint wertlos. Es bedarf aber nicht nur der kritischen Bewertung dessen, was als Produkt der Curriculumentwicklung entsteht, sondern auch der kritischen Beurteilung des ganzen Prozesses, in dem es entstand. So gesehen, ist Evaluation integraler Bestandteil aller curricularen Arbeit überhaupt. Daher muß Evaluation von Anfang an in der ganzen Breite ihrer Aufgaben, ihrer Bedeutung, ihrer Verfahren begriffen werden. Es besteht aber durchaus die Gefahr, daß der Begriff der Evaluation zu eng und oberflächlich verstanden wird, z. B. wenn Evaluation nur auf individuelle Leistungsmessung begrenzt wird (vgl. z. B. Bloom/Hastings/Madaus 1971), wodurch ihre wichtige Funktion für den Prozeß der Curriculumentwicklung übersehen wird. Dieser Gefahr sind tatsächlich große Teile der amerikanischen Curriculumentwicklung seit dem Anfang der sechziger Jahre

erlegen. Sollen ähnliche Mißerfolge in der Bundesrepublik vermieden werden, dürfte zunächst ein Überblick über den augenblicklichen Stand der Evaluationsforschung in den USA angebracht sein; dadurch läßt sich ein guter Ausgangspunkt für die Entwicklung von Modellen, Methoden und einer Technologie der Curriculumevaluation gewinnen.

In den letzten Jahren ist Evaluation zu einem zentralen Thema der amerikanischen Curriculumforschung geworden⁵. Die Kontroversen über die Funktion der Evaluation haben wesentliche Beiträge zur Curriculumentwicklung geleistet (vgl. Review of Educational Research 1970). Kennzeichnet ist die Entwicklung der letzten Jahre durch eine zunehmende Vielfalt der Aufgaben, die der Evaluation zufallen. Dabei ist es zu grundlegenden Veränderungen in der Konzeptualisierung und Methodologie gekommen, ohne daß es jedoch gelungen wäre, die dabei aufgeworfenen Probleme zu lösen. Summarisch läßt sich sagen: *Curriculum-evaluation zielt auf die Sammlung, Verarbeitung und Interpretation von Daten mit dem Ziel, Entscheidungen über ein Curriculum zu fällen. Das impliziert: (1) objektive Beschreibungen von Zielen, Umwelt, Personal, Methoden und Inhalt und Ergebnissen; und (2) persönliche Urteile über die Qualität und Angemessenheit dieser Ziele, der Umwelt usw.* (vgl. Stake 1967 b).

Scriven hat auf die Unterscheidung zwischen Ziel und Rolle der Evaluation aufmerksam gemacht (vgl. Scriven S. 60 ff). Ziel ist die Sammlung von Informationen zum Zweck rational begründeter Entscheidungen über »etwas«. Die Rolle der Evaluation hängt davon ab, was dieses »etwas« ist und von *wem* und *welche* Maßstäbe dabei angelegt werden. Aufgabe der Evaluation ist es z. B., zur Konstruktion eines Curriculum, zur Antizipation seines Erfolges oder zu seiner Verbesserung beizutragen. In der Regel erfolgt Evaluation für verschiedene Adressaten, z. B. Politiker, Fachleute, Lehrer, Eltern, Schüler und Curriculumentwickler. Mit Rücksicht auf ihr unterschiedliches Erkenntnisinteresse ist die *Form* und die *Sprache*, in der die Ergebnisse eines Evaluationsprozesses dargestellt werden, von Fall zu Fall zu variieren.

Um zu einer angemessenen Konzeptualisierung von Evaluation beizutragen, muß man

- ein Vorstellungsschema liefern, das die Gebiete und Probleme klassifiziert, die evaluiert werden sollen,
- Strategien und Methoden der Evaluation entwickeln und
- ein System von Generalisationen für den Gebrauch der verschiedenen Verfahren und Techniken entwerfen (Alkin 1969, 2) ⁶.

Das ist in den USA im Laufe der letzten Jahre in verstärktem Maße geschehen. So entwarf Scriven bereits 1967 eine Methodologie der Eval-

uation und Stake 1967 ein Evaluationsmodell. 1969 folgten Alkins Evaluationstheorie, Maguires Evaluationsmethodologie, Provus' Diskrepanzmodell und Stufflebeams CIPP-Evaluationsmodell (Context Input Process Product).

In mehreren dieser Evaluationsmodelle wird im Rahmen der Theoriebildung als Hauptaufgabe der Evaluation die *Bereitstellung von Daten für Entscheidungen* bezeichnet. Sie finden in Situationen statt, in denen zwischen mehreren Alternativen auf Grund unterschiedlicher Bewertung gewählt wird. Im Prozeß der Entwicklung und Einführung eines Curriculum in die Schule lassen sich mehrere Phasen kennzeichnen, in denen Entscheidungen von Entscheidungsträgern auf verschiedenen Ebenen zu fällen sind. Das bedeutet: es werden je nach Phase mehrere Funktionen der Entscheidung angesprochen und dementsprechend unterschiedliche Informationen benötigt; es müssen daher mehrere Evaluationsfelder unterschieden werden (Alkin 1969 a, 2):

1. Erhebung der *Werte und Bedürfnisse* der Schüler und der Gesellschaft⁷, um die Bildungs- und Lernziele auszuwählen, die in einem Curriculum in einer bestimmten historisch-gesellschaftlichen Situation realisiert werden sollen.

2. *Programmplanung* mit dem Ziel der Auswahl von Curricula. Hier gilt es, denen, die Entscheidungsbefugnis haben, möglichst objektive Informationen über verschiedene Curricula zu liefern, damit sie entscheiden können, welche speziellen Curricula ihren besonderen Bedürfnissen entsprechen.

3. *Programmimplementation*: Die Einführung curricularer Programme in die Schule und die Evaluation des Ausmaßes, in dem die Intentionen der eingeführten Curricula verwirklicht werden, und die Feststellung, inwieweit sie den Bedürfnissen entsprechen, die als Voraussetzung für die Entscheidung über die Einführung formuliert worden sind.

4. Evaluation der betreffenden Curricula mit dem Ziel, *sie zu verbessern, während sie noch im Prozeß der Entwicklung sind*.

5. Evaluation der Ergebnisse der betreffenden Curricula mit dem Ziel der *Programmbestätigung* (programm certification). Hier gilt es, eine abschließende Evaluation der Curricula vorzunehmen, um den generellen Wert und ihre Verwendbarkeit bestimmen zu können.

Auf die Evaluation im fünften Evaluationsfeld folgt wieder die Evaluation im ersten Feld, die auf z. T. veränderte Bedürfnisse und Interessen stoßen dürfte, wodurch ein neuer Ausgangspunkt für die Curriculumrevision gebildet wird.

Im ganzen bedeutet das, daß in jeder Phase der curricularen Arbeit evaluiert werden muß. Daraus folgt wieder, daß die Rückmeldung der

Ergebnisse Einfluß haben muß auf die Gestaltung der jeweils nächsten Phase oder, um dieses gedankliche Modell zu benutzen, daß Evaluation in einem Regelkreis mit allen anderen Verfahren im Curriculumgefüge zusammengekoppelt ist. Ihre Funktion ist es, die Optimierung des Curriculum zu steuern.

Am Beispiel eines Teils der Evaluation, der Datenerhebung, exemplifiziert, heißt das:

1. Daten über Bedürfnisse haben Einfluß auf die Bestimmung der Ziele.
2. Daten über die Gebiete, auf denen die Ziele liegen, haben Einfluß auf die Planung des Programms.
3. Daten über die schulische Einführung des Programms haben Einfluß auf die Einschätzung der Bedürfnisse, Ziele, Gebiete, Programmplanungen.
4. Die veränderte Einschätzung hat Einfluß auf die Umgestaltung des bisherigen Programms.
5. Daten über die Gesamtergebnisse des Programms haben Einfluß auf die vorläufig abschließende Einschätzung aller vorausliegenden Phasen und auf das Gesamturteil über das Programm und so fort.

Um die angestrebten Informationen optimal zu erhalten, bedarf es auch der Erarbeitung einer *Technologie der Evaluation* (Stake 1967 b), die die Konzeptualisierung und Methodologie der Evaluation ergänzen muß. Diese müßte Elemente der psychometrischen Testtechnologie (vgl. Buros 1965)⁸, der sozialwissenschaftlichen Erhebungstechnologie, der Kommunikationstechnologie und der Unterrichtstechnologie einschließen. Das Fehlen einer solchen Technologie macht sich um so stärker bemerkbar, als die vorhandenen Tests sich selten für die Zwecke der Evaluation eignen. Die meisten von ihnen sind standardisierte Leistungstests, die zwar zur Feststellung von individuellen Leistungsunterschieden bei Schülern geeignet sind, nicht aber zur Evaluation eines Curriculum.

Will man sich einen Überblick über das gesamte Problemfeld der Evaluation verschaffen, so lassen sich drei Problemkreise der Evaluationsforschung, die im folgenden zu behandeln sind, abgrenzen:

1. die verschiedenen Formen und Rollen der Evaluation,
2. die Versuche, ein Evaluationsmodell zu erstellen, das einen Rahmen für weitere Bemühungen um eine Konzeptualisierung und Methodologie abgibt,
3. die Wert- und Entscheidungsprobleme der Evaluation und Methoden zu ihrer Lösung.

II. Rollen und Formen der Evaluation

Die erste und einfachste Unterscheidung zur Begriffs- und Funktionsbestimmung der Evaluation ist bereits getroffen. Es ist Scrivens Distinktion zwischen Zielen und Rollen. Dabei liegt für unsere Darstellung das Hauptgewicht zunächst auf den möglichen verschiedenen Rollen; denn als Ziel kann einheitlich die Absicht gesehen werden, Antworten auf bestimmte Fragen zu erhalten, die bestimmte Entscheidungen ermöglichen. Befragt werden muß alles, was in dem Gesamtprozeß der Curriculumentwicklung relevant ist, z. B.

- die Personen, die mitarbeiten,
- die Ziele, die für ihre Arbeit gesetzt werden oder gelten,
- die Verfahren, mit deren Hilfe man die Ziele anstrebt,
- die Programme, von denen man sich dabei leiten läßt,
- die Instrumente, die dabei benutzt werden,
- die Urteile und Wertschätzungen, die die Arbeit steuern und die sich naturgemäß auf Personen, Ziele, Themen, Verfahren usw. beziehen können.

In Frage gestellt werden kann schließlich (auf einer weiteren Reflexionsstufe) die Evaluation selbst (und was aus ihr als Verfahren folgt oder gefolgt ist). Einige Klärungen innerhalb dieses weiten Problem-, Sach- und Begriffsfeldes sind bereits erfolgt. Von Belang sind m. E. die Distinktionen zwischen

- formativer und summativer Evaluation,
- Mikro- und Makroevaluation,
- innerer und äußerer Evaluation,
- vergleichender und nicht vergleichender Evaluation,
- intrinsischer und Ergebnisevaluation.

Sie werden im folgenden kurz dargestellt (vgl. Flehsig 1970).

1. Formative und summative Evaluation

Scrivens Unterscheidung zwischen formativer und summativer Evaluation zur Kennzeichnung zweier Rollen von Evaluation hat Verbreitung gefunden. Formative Evaluation erfolgt, wenn das Curriculum sich noch im Prozeß der Entstehung (Formung) befindet. Summative Evaluation wird nach der Beendigung der Entwicklung eines Curriculum benötigt, wenn eine (vorläufig) abschließende Beschreibung und Wertung gewünscht wird. Entsprechend diesen beiden Rollen der Evaluation sind das Erkenntnisinteresse und die Methoden der Evaluation verschieden. Cronbach hatte 1963 in seinem bekannten Aufsatz »Evaluation for Cour-

se Improvement« die große Bedeutung der formativen Evaluation für die Curriculumentwicklung deutlich gemacht, ohne jedoch terminologisch zwischen formativer und summativer Evaluation zu unterscheiden. Die Aufgabe formativer Evaluation ist es, Schwächen der in der Entwicklung befindlichen Curricula aufzudecken und zu beseitigen (Scriven 1967). Sie ist also als *der* Teil der Curriculumentwicklung zu begreifen, der Auskünfte über die Qualität des in der Entwicklung begriffenen Curriculum bietet. Auf Grund formativer Evaluation kann das Curriculum oder können einige seiner Teile revidiert werden, bis es seine Endfassung findet, die dann summativer Evaluation unterliegt. Grobman hat zu Recht darauf aufmerksam gemacht, daß »es keine scharf abgrenzbare Unterscheidung zwischen diesen beiden Phasen gibt; die formative Evaluation hört nicht auf, bevor die summative Evaluation anfangen kann« (1968, 14). Es ist die Aufgabe formativer Evaluation, auf jede nur erdenkliche Weise Informationen über die Wirkungen des Curriculum zu erhalten; dabei sollten nicht nur die strengen Maßstäbe eines Forschungsplans (research design) angelegt werden, vielmehr gilt es, einen Evaluationsplan zu entwerfen. Um möglichst vielfältige Informationen über das Curriculum zu erhalten, schlägt Cronbach statt der Verwendung von Tests, die auf individuelle Unterschiede zielen, die Einrichtung einer Gesamtheit von Items vor. Aus einer beispielsweise 600 Items umfassenden Gesamtheit sollten besser je 50 Items 12 Gruppen von 50 Schülern gegeben werden als nur 50 Items 600 Schülern.

Die Entwicklung einer Gesamtheit von Items, die es erlauben, qualifizierte Aussagen innerhalb der formativen Evaluation zu erhalten, sollte mit der Formulierung von Lernzielen Hand in Hand gehen. Die Operationalisierung der Lernziele in Items legt ja obendrein eine permanente Revision der Lernziele und Items nahe. Es gilt die Übereinstimmung zwischen den formulierten Lernzielen, den impliziten und denen, die durch Items getestet werden, zu erreichen. Dabei ist das Validitätsproblem schwierig zu lösen. Es ist notwendig, Lernziele und Curriculuminhalte, Lernziele und Inhalte der Prüfung, Curriculuminhalt und Prüfungsinhalt so zusammenzubringen, daß sie wechselseitige Identifikation zulassen (Scriven 1967). Die Herstellung einer derartigen Entsprechung ist eine wichtige Aufgabe des Evaluators, der er sich unterziehen muß, bevor er das Curriculum im Rahmen der formativen Evaluation hinsichtlich seiner Wirkungen auf Schüler mit dem Ziel der Verbesserung evaluieren kann.

Lindvall und Cox formulieren vier Fragen, die die Funktion von Kategorien für formative Evaluation haben; jede Kategorie enthält mehrere Unterkategorien, die ebenfalls in Form von Fragen entworfen wer-

den. Mit der Verwendung dieses *Kategoriensystems* kann formative Evaluation beginnen. Es wurde von den Autoren vor allem im Hinblick auf die formative Evaluation der *Individually Prescribed Instruction* entwickelt, beansprucht aber zugleich allgemeine Verwendungsfähigkeit:

1. Welche Ziele soll das Programm erreichen?
 - a) Sind die formulierten Ziele wirklich Lernziele?
 - b) Sind die formulierten Lernziele die wirklichen Ziele des Programms?
 - c) Sind die Ziele wertvoll?
 - d) Sind die Ziele erreichbar?
2. Wie ist der Plan, diese Lernziele zu erreichen?
 - a) Verspricht der Plan, zur Erreichung der Lernziele beizutragen?
 - b) Ist der Plan genügend detailliert entwickelt?
 - c) Können Plan und Verfahren ohne Schwierigkeiten von den Leuten verstanden werden, die sie verwirklichen sollen?
 - d) Ist es wahrscheinlich, daß der Plan verwirklicht werden kann?
3. Stellt das funktionierende Programm eine richtige Einführung des Plans dar?
 - a) Welches sind die spezifischen Punkte, die in einer Analyse der Funktion zu beobachten sind?
 - b) Werden die Aktivitäten wirklich entsprechend dem Plan ausgetragen?
 - c) Wie kann erreicht werden, daß Plan und Verwirklichung korrespondieren?
 - d) Führt die Untersuchung der wirklichen Funktion zu irgendwelchen Veränderungsvorschlägen?
4. Erreicht das Programm, wenn es entwickelt und verwirklicht ist, sein erwünschtes Ziel?
 - a) Sehen die Pläne die Evaluation aller Programmziele vor?
 - b) Sind die Evaluationsverfahren reliabel?
 - c) Ist der gesamte Evaluationsprozeß umfassend genug, um das benötigte Gesamtbild der Programmresultate zu liefern?
 - d) Welches sind die Folgerungen aus den Ergebnissen für die Modifikation des Programms? (Lindvall/Cox, 1970, 5-11) 9.

Summative Evaluation, auch Ergebnisevaluation oder Produktevaluation genannt, ist die abschließende Evaluation eines Curriculum. Scriven hat ihre prinzipielle Gleichwertigkeit mit formativer Evaluation betont. Ihre Rolle ist es, nach der Evaluation des Curriculum Daten zur Verfügung zu stellen, die seinen Wert erkennen lassen und die »den Schulsystemen helfen, Entscheidungen über die Adaptation und den Gebrauch von Materialien zu treffen« (Grobman 1968, 14). Evaluation kann dabei einmal auf das Curriculummaterial selbst bezogen werden und seine Validität im Hinblick auf seine Gestaltung und auf seine Lernziele prüfen.

Sie kann aber auch das Verhalten der Schüler untersuchen, die mit diesem Curriculummaterial arbeiten, um festzustellen, welche Fähigkeiten und welche Veränderungen im Verhalten sie erreicht haben.

2. Mikro- und Makroevaluation

Innerhalb einer abschließenden Evaluation sind zwei Schwerpunkte möglich. Den einen bildet die *Mikroevaluation*, also die Evaluation von Einzelteilen unter einer bestimmten Fragestellung, den anderen die *Makroevaluation*, die auf eine Beantwortung von allgemeinen Fragen, z. B. über die Benutzbarkeit des Materials, zielt. Summative Evaluation wird im allgemeinen auf die Kombination von Mikroevaluation und Makroevaluation zielen, um zu einer fundierten abschließenden Beschreibung und Wertung des Curriculum zu kommen.

3. Innere und äußere Evaluation

Formative Evaluation erfolgt als Teil der Entwicklung eines Curriculum. Sie wird am besten von den an der Entwicklung Beteiligten ausgeführt (innere Evaluation). Da sie in der Entwicklungsarbeit stehen, haben sie Kenntnis vom Gegenstand der Evaluation. In einem stärkeren Maße mit dem Projekt identifiziert als Evaluatoren, die nicht mit der Projektherstellung verbunden sind, werden sie nicht so leicht durch formalen Dogmatismus den kreativen Schwung der übrigen Curriculumhersteller bremsen. Auch können sie eher bei der rechtzeitigen deutlichen Formulierung der Lernziele helfen. Der dynamische Charakter der Curriculumentwicklung fordert eine schnelle Anpassungsbereitschaft an die sich wandelnden Fragen und Probleme der Curriculumhersteller, die nur ein mit dem Projekt mitarbeitender Evaluator leisten kann. Bei formativer Evaluation ist der Zeitpunkt ihres Beginns und die Art ihrer Durchführung zu bedenken. Atkin hat auf diesbezügliche Probleme hingewiesen (1963, 129–132). So kann z. B. das Testen eines schwierigen Vorstellungszusammenhangs zu einem zu frühen Zeitpunkt negative Folgen im Hinblick auf ein weiteres vertieftes Verständnis dieses Zusammenhangs haben.

Summative Evaluation kann von Evaluatoren geleistet werden, die nicht in Verbindung mit Curriculumentwicklung oder -realisierung stehen (äußere Evaluation). Das hat den Vorteil, daß die Evaluationskriterien oft objektiver gewählt werden, mithin auch größere Objektivität bei umfassenden Wertungen über das Curriculum gegeben ist. Auch ist es für nicht an der Projektentwicklung beteiligte Evaluatoren leichter, Vergleichsuntersuchungen verschiedener Curricula vorzunehmen. Ein solches Eval-

ationsteam kann sich aus verschiedenen Spezialisten zusammensetzen, so daß ein großes Spektrum von methodischen Ansätzen gewährleistet wird.

4. Vergleichende und nicht vergleichende Evaluation

Betrachtet man die Ergebnisse der Evaluation neuer Curricula in den USA, die im Vergleich zu Kontrollgruppen erfolgt ist, die mit altem Unterrichtsmaterial arbeiten, fällt im allgemeinen das gute Abschneiden der neuen Curricula und die Unzulänglichkeit der alten Materialien ins Auge; zuweilen allerdings ergeben sich keine signifikanten Unterschiede. Das hat u. a. dazu geführt, daß die Zulänglichkeit der Kriterien des Vergleichs angezweifelt wurde, da sie vor allem im Hinblick auf die Lernziele der neuen Curricula erstellt wurden und für vergleichende Evaluation nicht angemessen waren. Infolgedessen hat man oft dem Urteil von Fachleuten über Inhalt und Ziele und Unterrichtsprozesse mehr Wert zugemessen als empirisch festgestellten geringen Unterschieden von Schülern in Experimentier- und Kontrollgruppen. Cronbach schlägt die genaue Untersuchung der »Leistungen einer genau umschriebenen Gruppe am Ende eines Curriculum im Hinblick auf die wesentlichen Ziele und Nebenwirkungen« (1963, 43, 48) vor. Durch eine genaue Evaluation des Verhaltens *einer* Gruppe glaubt Cronbach zuverlässigere Daten zu erhalten als durch den methodisch schwierigeren Vergleich verschiedenartig unterrichteter Gruppen. Scriven wendet sich gegen diese Annahme, indem er darauf hinweist, daß die Evaluation eines Curriculum nicht nur die Beschreibung impliziert, sondern auch seine Wertung im Hinblick auf andere einschließt. Ob vergleichende oder nicht vergleichende Evaluation gewählt wird, richtet sich nach der Ausgangssituation und dem Erkenntnisinteresse. Evaluation mit Hilfe des Vergleichs empfiehlt sich dann, wenn die allgemeinen Zielsetzungen gleich sind, d. h. wenn z. B. zwischen zwei Curricula oder zwei Teilen eines Curriculum gewählt werden soll, die auf verschiedenen Wegen das gleiche Ziel erreichen wollen. Eine nicht vergleichende Evaluation ist dann angebracht, wenn die Unterschiede zwischen den Curricula auf Grund verschiedener Zielsetzungen so groß sind, daß die Summe der Gemeinsamkeiten zu klein ist.

5. Intrinsische Evaluation und Ergebnisevaluation

Unter intrinsischer Evaluation wird die Evaluation von beispielsweise Inhalten, Lernzielen, Zensierungsverfahren verstanden, ohne daß ihre Auswirkungen auf die Schüler untersucht werden. Ihre Kriterien sind im

allgemeinen nicht operationalisiert und beziehen sich auf den zu evaluierenden Gegenstand selbst. Die Schwierigkeiten bei der intrinsischen Evaluation liegen darin, daß sie die Kriterien und Ziele formulieren muß, durch die Inhalt, Lehrerverhalten usw. evaluiert werden, zumal die von den Curriculumentwicklern formulierten Lernziele und Kriterien oft nicht gleich den impliziten Lernzielen und Kriterien sind ¹⁰.

Unter der reinen Ergebnisevaluation wird dagegen die Untersuchung der Effekte des Curriculum auf die Schüler verstanden. Sie werden z. B. durch Vor- und Nachtest in der Experimentiergruppe festgestellt. Dabei wird oft davon ausgegangen, daß »die Berücksichtigung von Ziel- und Inhaltbewertung oder einer anderen sekundären Bewertung für die Curriculumevaluation nicht nur irrelevant, sondern auch äußerst unzuverlässig« ist (Scriven 1967, 59, 79). Für diese Art von Evaluation ist es weniger wichtig, was die Lehrer und Schüler als ihr Tun *bezeichnen*, wichtig ist, was sie *wirklich tun*. Die Schwierigkeiten bei der reinen Ergebnisevaluation liegen darin, *mit einem sinnvollen Grad an Allgemeinheit* zu beschreiben, was der Schüler gelernt hat.

Aufgabe der Evaluation ist es jedoch nicht nur, die Ergebnisse eines Curriculum festzustellen; Evaluation muß auch die Frage beantworten, wie diese im Verhältnis zu anderen Curricula zu werten sind. Dazu bedarf es der Spezifizierung des Lernerfolges, seiner Begründung und der Formulierung von Zielen und Kriterien. In der Ergebnisevaluation muß entschieden werden, »welches Verhalten zu werten ist als adäquates Verständnis für Schüler auf einem bestimmten Niveau für ein bestimmtes Fach, und es bedarf der Anwendung dieser Entscheidung auf Daten über das spätere Verhalten der Schüler, um eine Gesamtevaluation zu ermöglichen« (Scriven 1967, 61). Es muß also das Verhalten der Schüler analysiert werden, um zu bestimmen, wo Unzulänglichkeiten im Hinblick auf bestimmte Kriterien oder Lernziele liegen. Im allgemeinen verbindet man intrinsische und Ergebnisevaluation, um zu möglichst umfassenden Ergebnissen zu kommen.

III. Modelle der Curriculumevaluation

1. Das Stakesche Evaluationsmodell

Einen Schritt weiter in die Problematik der Konzeptualisierung von Evaluation führt der Entwurf von Modellen für die Curriculumevaluation. Sie liegen auf einer höheren begrifflichen Stufe als die Distinktionen zur Beschreibung unterschiedlicher Evaluationsformen und Evaluationsfunktionen. Geht man wieder davon aus, daß die Verfahren der Evaluation

beabsichtigen, relevante Informationen für die Verbesserung von Curricula zu erhalten, so sind vor allem drei Felder wichtig:

- das Vorfeld des Unterrichts,
- das Aktionsfeld des Unterrichts,
- das Feld seiner Wirkungen.

Etwas detaillierter beschrieben, heißt das: Das erste Evaluationsfeld enthält alle Bedingungen, die vor dem Unterricht bestehen und Einfluß auf seine Ergebnisse haben. Zu diesen gehört die Erhebung der sozialen Situation der Schüler, ihrer Bedürfnisse und Interessen und ihrer Umwelt mit ihren Wert- und Zielvorstellungen.

Zum zweiten Evaluationsfeld gehören alle Bedingungen innerhalb der Schule, die für die Erreichung der Ergebnisse wichtig sind. Es schließt den Prozeß der Einbringung des Curriculum in die Schule ein, die Einführung des Lehrers in die Arbeit mit den curricularen Materialien und vor allem die Unterrichtsprozesse, die Interaktionen der Lehrer mit den Schülern, der Schüler untereinander und der Schüler mit dem Unterrichtsmaterial. Das dritte Evaluationsfeld beinhaltet die augenblicklichen, mittelfristigen und langfristigen Ergebnisse des gesamten Unterrichts (vgl. Sjo- gren 1970).

Stake (1967 a) hat die drei Evaluationsfelder *Voraussetzungen, Prozesse, Ergebnisse* genannt und eine verstärkte Evaluation der Voraussetzungen und der Prozesse gefordert. Zu diesem Zweck entwickelt er eine Eval-

Evaluationsmatrix nach Stake

Rationale
Begrün-
dung

| | Intentionen | Beobachtungen | | Normen | Urteile |
|--|---------------------|---------------|-----------------|--------------|---------|
| | | | Voraussetzungen | | |
| | | | Prozesse | | |
| | | | Ergebnisse | | |
| | Beschreibungsmatrix | | | Urteilmatrix | |

uationsmatrix, indem er über die drei Evaluationsfelder zwei Beschreibungskategorien, »Intentionen« und »Beobachtungen«, und zwei Urteilkategorien, »Normen« und »Urteile«, legt. Dadurch entsteht eine zwölf-feldige Evaluationsmatrix. Ergänzende Informationen für ihre Anwendung bietet die Kenntnis der rationalen Begründung des Curriculum, dessen Zielsetzungen mit ihrer Hilfe besser zu erfassen sind.

Unter »Intentionen« werden z. B. die geplante Erarbeitung von Sachgebieten oder das erwartete Schülerverhalten verstanden. Sie sollten möglichst präzise verbalisiert werden. Die Diskrepanz zwischen Lernzielen und Lernergebnissen ist nicht ohne weiteres Zeugnis für einen negativen Verlauf des Unterrichts. Es können unvorhergesehene Ergebnisse erzielt worden sein, die durchaus positiv zu werten sind. Ob Lernziele *immer* vor dem Unterricht festgelegt und ob sie *immer* in Verhaltenstermini bestimmt werden müssen (vgl. Wulf 1972 b), ist in der amerikanischen Curriculumforschung umstritten¹¹. »Beobachtungen« bilden die zweite Beschreibungskategorie der Matrix. Hier geht es um die Beobachtung dessen, was in den drei Evaluationsfeldern beschreibenswert erscheint. Es setzt die Auswahl von Kriterien voraus, unter denen zu beobachten ist. Diese ist eine Entscheidung des Evaluators, bei der es im Curriculum formulierte Kriterien und von außen herangetragene zu berücksichtigen gilt. Es bieten sich zwei Wege an, die Daten der beiden Beschreibungskategorien zu verarbeiten: »Man muß die Kontingenzen zwischen den Voraussetzungen, Prozessen und Ergebnissen und die Kongruenz zwischen den Intentionen und Beobachtungen finden« (Stake 1967 a, 104). Die Kontingenzen zwischen den einzelnen Variablen der drei Evaluationsfelder bedürfen der Beachtung: »Insofern, als Evaluation die Suche nach Beziehungen ist, die die Verbesserung der Erziehung ermöglichen, ist es Aufgabe des Evaluators, Ergebnisse zu identifizieren, die mit bestimmten Voraussetzungen und Unterrichtsprozessen kontingent sind« (a. a. O.).

Zwischen den drei Evaluationsfeldern besteht in der ersten Beschreibungskategorie, »Intentionen«, eine logische Kontingenz, im Bereich der zweiten, »Beobachtungen«, eine empirische Kontingenz. Vollständige Kongruenz der Daten eines Curriculum besteht, wenn alle beabsichtigten Voraussetzungen, Prozesse und Ergebnisse eintreten. Das ist selten der Fall. Im allgemeinen wird der Evaluator Kongruenz lediglich bei einigen Teilzielen feststellen können. Kongruenz besagt nur, daß das, was beabsichtigt war, eintrat, nicht jedoch, daß die Ergebnisse *reliabel* oder *valide* sind.

Die Urteilmatrix besteht aus zwei Kategorien, den »Normen« und den »Urteilen«. Mit welchen Methoden und Strategien die Normen zu erstellen sind, ist ein zentrales Problem der Evaluation. Normen sind im allgemeinen gebunden an die historisch-gesellschaftliche Situation. Mit

den Lernzielen können sie nicht gleichgesetzt werden, wie es in einigen Evaluationsprojekten geschehen ist, da dann Evaluation nur Überprüfung der Erfüllung von Lernzielen wäre. Vielmehr müssen auch die Lernziele an Normen gemessen werden. Zwei Arten von Normen sind denkbar, *absolute* und *relative*. Letztere sind beispielsweise in einem alternativ vorliegenden Curriculum enthalten. Wenigstens an einem der beiden Normengefüge muß das Curriculum durch Urteilen gemessen werden, um seinen Wert zu bestimmen.

2. Das Provussche Diskrepanzmodell

Provus macht darauf aufmerksam, daß die entscheidenden evaluativen Informationen durch die Feststellung von Diskrepanz gewonnen werden. Danach besteht der Prozeß der Evaluation darin,

- Normen festzulegen,
- die Diskrepanz zwischen dem Curriculum und den Normen festzustellen und
- die Diskrepanzinformation dazu zu benutzen, die Schwächen des Curriculum zu identifizieren.

Die Evaluation eines Curriculum soll vier Entwicklungsstadien durchlaufen und dabei drei größere Inhaltskategorien einschließen. Das erste Entwicklungsstadium ist die *Definition* des Curriculum, das zweite seine *Einführung* in die Schule (installation), das dritte die curricularen *Prozesse*, das vierte seine Fertigstellung als *Produkt*. Nach diesen vier kann als fünftes Entwicklungsstadium die *Kosten-Effektivitäts-Berechnung* erfolgen. In den vier bzw. fünf Entwicklungsstadien »bleiben die inhaltlichen Spezifikationen dieselben: *Eingabe, Prozeß* und *Ergebnis*, relativ zu Zeit und Geld« (Provus 1969, 244).

Unter *Definition* eines Curriculum wird seine genaue Beschreibung im Hinblick auf seine Lernziele, die Schüler, Lehrer, Medien und den Prozeß verstanden, der zur Erreichung der Lernziele führt. Sie dient als Norm zur Evaluation des Curriculum. Werden im (zweiten) Entwicklungsstadium, der *Einführung* des Curriculum, Informationen gewonnen, die eine Diskrepanz zwischen den Erwartungen hinsichtlich der Einführung des als Norm definierten Curriculum und dem tatsächlichen Lehrer- und Schülerverhalten anzeigen, muß entweder das Verhalten so korrigiert werden, daß es der Norm entspricht, oder es muß die Norm verändert werden. In dem (dritten) Entwicklungsstadium, dem *Prozeß*, bildet der Teil der Curriculumdefinition die Norm, der die Intentionen hinsichtlich des Unterrichtsprozesses enthält. Wird eine Diskrepanzinformation zwischen Intentionen und tatsächlichem Prozeß erhalten, muß sie zur Neudefini-

tion der Intentionen oder zu einer besseren Durchführung des Unterrichtsprozesses führen. Im (vierten) Entwicklungsstadium, in dem es um das *Produkt* des Unterrichts geht, gilt es festzustellen, ob das Curriculum seine in der Definition als Norm gesetzten End- und Gesamtlernziele erreicht hat oder ob Diskrepanz festgestellt werden muß.

Mit diesem Evaluationsmodell wird die Evaluationskategorie der *Diskrepanz* gewonnen. Darüber hinaus bietet Provus brauchbare Hinweise auf die methodischen Probleme seiner Anwendung und die spezifischen Aufgaben von Evaluationsgruppen, die hier nicht mehr berücksichtigt werden können.

Da es ein Modell für die Evaluation laufender curricularer Programme ist, fehlt die Reflexion auf die Aufgaben der Evaluation zur »Erhebung der Werte und Bedürfnisse«. Es fehlt auch die Reflexion auf die Notwendigkeit einer auf den Ergebnissen der Produktevaluation fußenden Revision des Curriculum. Außerdem werden die in den einzelnen Entwicklungsphasen notwendigen Entscheidungen nicht als Probleme der Wertung innerhalb der Evaluation gesehen.

3. Das *Stufflebeamsche* Evaluationsmodell

Gerade die Transparenz von Entscheidungsprozessen bildet das Schwergewicht der Bemühungen *Stufflebeams* (1969; vgl. auch: Phi Delta Kappa National Study Committee on Evaluation 1971). In diesem Rahmen werden sieben Entscheidungsvariablen unterschieden, die in den verschiedenen Entscheidungssituationen heranzuziehen sind:

- die Ebene der Entscheidungen,
- der Schwerpunkt der Entscheidungen,
- die Inhalte der Entscheidungen,
- die Funktion der Entscheidungen,
- die Gegenstände der Entscheidungen,
- der Zeitpunkt der Entscheidungen,
- der Grad an kritischer Reflektiertheit der Entscheidungen.

Diese Variablen gilt es auf den drei von *Stufflebeam* unterschiedenen interdependenten Ebenen der Evaluation (Gemeinde, Bundesstaat, Nation) zu bedenken. In diesem Modell werden vier Situationen und die entsprechenden Entscheidungsfunktionen (Planung, Programmentwicklung, Implementation, Modifikation) genannt, für die vier Strategien entwickelt werden. Die erste zielt auf die Evaluation des *Kontexts*. Hierbei gilt es, – ähnlich der ersten der fünf oben unterschiedenen Phasen der Curriculumentwicklung – die Bedürfnisse im Vorfeld zu erheben. Die zweite Strategie, die auf die *Eingabe* (input) gerichtet ist und deren

Aufgabe die Analyse alternativer Verfahrensentwürfe ist, will Daten für Entscheidungen über die Spezifizierung von Lernzielen, Verfahren, Materialien usw. sammeln. Die dritte beabsichtigt die Erhebung von Daten, die über den *Prozeß* gewonnen werden und die zur Verbesserung des Curriculum helfen sollen. Die vierte Evaluationsstrategie zielt darauf ab, die *Gesamtergebnisse* zu messen.

Darüber hinaus hat Stufflebeam (1969, 70, 143) versucht, die Struktur eines Evaluationsentwurfs zu entwickeln, der sechs Kategorien und jeweils einige Unterkategorien enthält. Er dürfte wegen seines formalen Charakters als Leitlinie für Evaluationsprojekte Beachtung verdienen, obwohl er nicht die unter den einzelnen Kategorien subsumierten Probleme der Wertung und immanenten Entscheidungen reflektiert.

A. Evaluationsschwerpunkt

1. Identifikation der wichtigsten Entscheidungsebenen (z. B. örtliche, einzelstaatliche und/oder bundesstaatliche)
2. Planung und Beschreibung aller Entscheidungssituationen auf jeder Entscheidungsebene in bezug auf ihren Schwerpunkt, die kritische Reflektiertheit, den Zeitpunkt und die Komposition der Alternativen
3. Bestimmung der Kriterien für jede Entscheidungssituation durch Spezifikation der Variablen für die Messungen und der Normen für die Beurteilung von Alternativen
4. Definition der Grundsätze und Richtlinien, innerhalb derer die Evaluation erfolgen soll

B. Informationssammlung

1. Spezifikation des Ursprungs der zu sammelnden Informationen
2. Bestimmung der Instrumente und Methoden für die Sammlung der erforderlichen Informationen
3. Spezifikation des anzuwendenden Stichprobenverfahrens
4. Spezifikation der Bedingungen und des Zeitplans für die Informationssammlung

C. Informationsorganisation

1. Erstellung eines Plans für die Informationen, die gesammelt werden sollen
2. Bestimmung der Mittel zur Kodierung, Organisation, Speicherung und zum Wiederabruf der Informationen

D. Informationsanalyse

1. Auswahl der analytischen Verfahren, die angewendet werden sollen

2. Bestimmung der Mittel zur Durchführung der Analyse
3. Spezifikation des Ausmaßes und der Form der Evaluationsberichte
4. Zeitplan des Informationsberichts

E. Informationsbericht

1. Definition der Adressatengruppe
2. Bestimmen der Mittel der Informationsvermittlung
3. Festlegen des Formats des Evaluationsberichts
4. Planung der Elemente für die Darstellung der Information

F. Administration der Evaluation

1. Zusammenfassung des Evaluationsplans
2. Bestimmung der für die Evaluation erforderlichen Mitarbeiterstellen und Finanzen
3. Spezifikation der Mittel, um die Evaluation gemäß ihren Grundsätzen und Richtlinien durchzuführen
4. Evaluation der Möglichkeiten des Evaluationsplans, valide, reliable, zuverlässige, aktuelle und überzeugende Informationen zu liefern
5. Spezifikation und zeitliche Planung der Mittel, um den Evaluationsplan regelmäßig auf den neuesten Stand zu bringen
6. Breitstellung eines Etats für das ganze Evaluationsprogramm.

IV. Das Wert- und Entscheidungsproblem in der Evaluation

Das dritte Problemfeld der Evaluationsforschung führt in das auch wissenschaftstheoretisch strittige Feld der Wertung (vs Wertfreiheit bzw. Wertneutralität). Damit kommt auch die in der BRD geläufige Opposition von deskriptiven und präskriptiven Verfahren ins Spiel.

In der Curriculumevaluation befaßt man sich erst seit kurzem mit der Problematik der Wertimplikationen beim Urteilen und bei Entscheidungsprozessen. So kann Stake (1970, 186) urteilen: »Die meisten Arbeiten über die Methodologie der Evaluation erwähnen kein Verfahren, Urteilsdaten zu sammeln und zu analysieren«.

Westbury kennzeichnet das Verhältnis von Evaluation und Beschreibung: »Evaluation kann (und muß wahrscheinlich) Beschreibung einschließen, aber Beschreibung schließt nicht notwendigerweise Evaluation ein« (1970, 241). Nach Larkins/Shaver liegt die Aufgabe der Evaluation neben Datensammlung und Analyse in der Berücksichtigung von Werten, Wertinhalten und -kriterien zur Beurteilung eines Curriculum.

»Ein adäquater Evaluationsentwurf fordert sowohl Verfahren, um ad-

äquate Schätzungsdaten (assessment data) – Schätzungsentwurf – zu erhalten, als auch Verfahren, um den Wert der Schätzungsdaten abzuwägen. Diese letzteren Verfahren sollen von nun an als *Wertanalyse* bezeichnet werden. Ein Modell für den Wertanalyseteil eines Evaluationsentwurfs schließt substantielle und Verfahrensüberlegungen ein. Die grundlegenden substantiellen Überlegungen sind der Satz von Werten, der benutzt wird, das Curriculum zu beurteilen. Die Verfahrensüberlegungen schließen ein, wie relevante Werte zu entwickeln, wie Wertkonflikte zu lösen, wie Entscheidungskriterien festzusetzen sind, um zu bestimmen, ob ein Curriculum einem Wert angemessen genügt« (Larkins/Shaver 1969, 6).

Jensen (1950) entwickelt ein Verfahren, die Rationalität der Auswahl von Lernzielen unter Zugrundelegung einer bestimmten Wertposition zu belegen. Auf jeden Fall sollte der Evaluator Kenntnis von verschiedenen Schemata zur Kategorisierung von Werten haben¹², weil dadurch die Spezifizierung von Lernzielen vereinfacht wird.

1. Werte, Prioritäten, Normen

Unterschiede in den strittigen öffentlichen Fragen beruhen auch auf unterschiedlichen Auffassungen von ethischen Werten. Neben den Wertvorstellungen, die aus direkten Interessen entstehen, bilden die ethischen Werte eine wichtige Basis für die Aufstellung von Bildungszielen und Lernzielen. Auch das ist bisher in der Curriculumforschung kaum reflektiert worden, so daß relativ wenig von den Wertvorstellungen der an der Erziehung Beteiligten bekannt ist. Der verbreitete Tylersche Weg der Lernzielbestimmung berücksichtigt kaum die den Lernzielen vorgegebenen Werte und ihre kritische Sichtung (vgl. Smith/Tyler 1942 u. Tyler 1969 b). Die Hinzuziehung von Sozialwissenschaftlern für die Erforschung der Wertvorstellungen wird wiederholt gefordert. In diesem Zusammenhang beginnt die Pädagogik auch die lange vernachlässigten Beziehungen zwischen Erziehungswissenschaft und Ethik wieder aufzunehmen¹³. Unter dem Gesichtspunkt der Sammlung von Informationen für die Evaluation lassen sich verschiedene Arten von Wertdaten unterscheiden. Außer *Lernzielen*, *summativen (Wert-)Urteilen* sind noch *Prioritäten* und *Normen* als Wertdaten zu nennen. Die Bestimmung der Prioritäten hat im Vergleich zu der genauen Lernzielbestimmung wenig Beachtung gefunden. Es liegt bisher kein Modell vor, um begründet festlegen zu können, was in einer Folge von Lernzielen oder Ereignissen den Vorrang hat. Hier bedarf es weiterer Theorienbildung und Forschung.

Maguire (1968) hat versucht, empirisch die Beziehungen zwischen Wer-

ten, Lernzielen und Prioritäten zu untersuchen. Er erhielt Wertesinschätzungen von einem heterogenen Satz von Lernzielen durch Lehrer und sodann von jeweils denselben Lehrern verschiedene Urteile darüber, welchen Lernzielen sie die Priorität geben wollten. Dabei stellten sich vor allem vier verschiedene Wert-Dimensionen der Lernziele in den Augen der Lehrer heraus: fachwissenschaftlicher Wert, motivationale Qualität, Einfachheit der Implementation und semantische Eigenschaften. Eine andere Kategorie von Beurteilungsdaten bilden »Standards«, die von Stake »als ein erwünschtes Qualitätsniveau für etwas, wie es von einer Autorität angegeben wird« (1970, 185), begriffen werden. Ihnen entspricht im deutschen Sprachgebrauch am ehesten der Begriff Norm.

Im Anschluß an Stake lassen sich drei Situationen identifizieren, in denen der Evaluator Daten über Werte, Prioritäten, Normen und Lernziele sammeln kann:

- die Befragung von Personen nach einem standardisierten Protokoll,
- die Beobachtung des Personenverhaltens,
- die Analyse von Curricula durch Experten.

Je nach Zielsetzung der Evaluation kann sich eine Kombination der Situationen empfehlen.

Vier empirische Methoden bieten sich an, um die Wertvorstellungen von Befragten zu ermitteln:

- Survey,
- Skalierung,
- Q-Technik,
- Semantisches Differential (Polaritätsprofil).

Survey-Befragungen empfehlen sich dort, wo gute Daten durch direkte Befragung erhalten werden können¹⁴. Wenn jedoch die einzelnen Fragen mehrdeutig sind oder das Ziel der Frage unklar ist, empfehlen sich stärker redundante Verfahren wie Skalierung¹⁵. Eine besondere Art ihrer Verarbeitung ist die 1953 von Stephenson entwickelte Q-Technik, die sich zur Ermittlung von Werten eignet¹⁶. Das semantische Differential ist ebenfalls ein spezielles Skalierungsverfahren. Mit ihm kann festgestellt werden, welche Bewertung Menschen bestimmten Begriffen und Vorstellungen, wie z. B. »Gewaltenteilung«, zuordnen. Es wurde 1957 von Osgood, Suci und Tannebaum entwickelt. Von Geis wurde dieses Verfahren 1968 zur Evaluation eines Projekts zur Verbesserung eines Curriculums verwendet. Taylor und Maguire benutzten es, um Biologielernziele zu untersuchen (1967).

Um Personenverhalten im Bereich der Schule zu beschreiben und zu bewerten, sind in den letzten Jahren viele Beobachtungssysteme entworfen und erprobt worden. Sie haben verschiedene Ansatzpunkte und beto-

nen unterschiedliche Aspekte. Mehrere versuchen, aus der Beobachtung des Verhaltens die ihm immanenten Werte, Normen und Ziele zu erfassen; andere begnügen sich mit der Beschreibung der Interaktionen zwischen Lehrer und Schülern oder untersuchen die Logik des Lehrens. Eine wertvolle Sammlung von Instrumenten zur Beobachtung, Beschreibung und Bewertung von Verhalten (vor allem im Unterrichtsprozeß) enthalten die Bände »Mirrors for Behavior« (Simon/Boyer 1967, 1970). Sie geben eine detaillierte Übersicht über 79 Beobachtungsverfahren, so daß auf eine Darstellung einzelner Systeme hier verzichtet wird.

Für die dritte Situation empfehlen sich die Techniken der Inhaltsanalyse, mit deren Hilfe sowohl Aufschluß über den Inhalt selbst als auch über seine Beziehung zum Produzenten und zum Konsumenten geliefert werden soll ¹⁷.

2. Der programmatische und der öffentlich-politische Bereich

Ein Kategorisierungsschema für Wertvorstellungen, Lern- und Bildungsziele, für Prioritäten und vor allem Ergebnisse von Curricula im Hinblick auf ihren politischen Charakter liefert Berlak (1970). Er unterscheidet zwischen *programmatischen* (programmatic) und *öffentlich-politischen* (public-policy) Ergebnissen (outcomes) von Curricula. Diese Bereichsunterscheidung geschieht im Bewußtsein des eminent politischen Charakters der Erziehung. Das Politische wird erstens der Maßstab, mit dessen Hilfe die Ergebnisse des Unterrichts in solche von öffentlich-politischem Interesse und solche programmatischen Charakters unterteilt werden, zweitens der Maßstab, Wertvorstellungen von großer öffentlich-politischer Bedeutung von solchen geringerer politischer Relevanz zu unterscheiden, drittens der Maßstab für das Recht des Evaluators, Urteile abzugeben. Daß auch der programmatische Bereich nicht ohne politische Implikationen ist, wird bei der Unterscheidung der beiden Bereiche durch Berlak nicht übersehen. Die programmatischen Ergebnisse von Curricula werden durch »Lernziele« angestrebt, die öffentlich-politischen durch »Bildungsziele«.

Der Evaluator muß sich bewußt werden, ob er »Lernziele« im programmatischen Bereich oder »Bildungsziele« im öffentlich-politischen Bereich evaluiert. Danach richtet es sich, ob er beschreibt, Urteilskriterien empfiehlt oder urteilt. Dabei stellt sich das Problem, wie die Grenze zwischen den beiden Bereichen bestimmt wird. Berlak nennt vier Kriterien zur Bestimmung des öffentlich-politischen Bereichs, die er in Form von Fragen formuliert:

1. Ändert des Programm direkt oder indirekt das Verhältnis der Macht zwischen Bürger und Staat?

2. Betrifft das Programm sofort oder später den Status einer Person oder die Macht, die sie in dem sozialen System ausüben kann?
3. Hat das Programm einen Effekt, der danach strebt, politische oder soziale Spannungen zu vermehren oder abzubauen?
4. Bewirkt das Programm einen Wechsel im Selbstverständnis oder Selbstgefühl des Individuums?

Öffentlich-politische und programmatische Ergebnisse können *geplant* oder *ungeplant* und, wenn ungeplant, noch halbwegs *erwartet* oder völlig *unerwartet* auftreten. Je nach der Zugehörigkeit zum öffentlich-politischen oder programmatischen Bereich bedarf es unterschiedlicher Evaluationsstrategien, die von verschiedenen ausgebildeten Personen angewendet werden. Eine sorgfältige Unterscheidung der beiden Bereiche bietet sich aus politischen und finanziellen Gründen an. Soll die Erreichung oder Verfehlung von vorher formulierten Lernzielen evaluiert werden, handelt es sich im allgemeinen um den programmatischen Bereich. Werden dagegen die Wertvorstellungen und Meinungen der Eltern über einschneidende schulorganisatorische Veränderungen evaluiert, befindet man sich im öffentlich-politischen Bereich.

Für den Evaluator schlägt Berlak folgendes Verhalten vor: Im programmatischen Evaluationsbereich ist der Evaluator aufgerufen, Werturteile abzugeben. Im Bereich der öffentlich-politischen Fragen ist es beim jetzigen Wissenschaftsstand angemessen, das Schwergewicht auf die Beschreibung zu legen, da es für diesen besonderen Bereich noch keine Evaluationsexperten gibt. Bei der Evaluation kann es in beiden Bereichen zu einander widersprechenden Ergebnissen kommen, mithin auch zu Konflikten, die ausgeglichen werden müssen. So könnte z. B. die Evaluation von Curricula im programmatischen Bereich in den Gesamtschulen zeigen, daß die kognitiven Ergebnisse geringer sind als die in herkömmlichen Schulen. Dennoch kann die Evaluation im öffentlich-politischen Bereich zeigen, daß diese Schulart erfolgreich ist, z. B. bei der Verwirklichung des Bildungsziels der optimalen Sozialisation des einzelnen.

3. Aspekte der Urteils- und Entscheidungsproblematik der Evaluation

Umstritten ist in der pädagogischen Diskussion, ob Wertkonflikte rational lösbar sind, ob Werturteile bewiesen und widerlegt werden können, »ob ethische Urteile in objektiver Weise gerechtfertigt werden können« (Berlak 1970, 269). Sind Wertkonflikte nicht rational lösbar, so können diejenigen, die Entscheidungsmacht haben, Ergebnisse der Evaluation auf Grund ihrer Wertvorstellungen ablehnen. Geiger (1961) und Scriven (1966) treten entschieden für die Lösbarkeit ethischer Differenzen mit-

tels rationaler Strategien ein, wie sie in den Sozialwissenschaften verwendet werden. Berlak (1970) nimmt im Anschluß an Stevenson (1944) eine Zwischenposition ein. Er tritt für die Lösbarkeit einiger Aspekte von ethisch strittigen Fragen ein, andere hält er nicht für lösbar. Schwierig dürfte die Bildung von Kriterien zur Zuordnung der Aspekte zu einer der Wertkategorien sein. Die Frage nach der Rechtfertigung von Werturteilen in objektiver Weise berührt die komplizierte Problematik von Entscheidungen im curricularen Bereich. Nach einer Unterscheidung von Cronbach und Suppes (1969) zielt eine Gruppe von Evaluationsuntersuchungen auf ein besseres Verständnis des untersuchten Gegenstandes, eine andere auf die wissenschaftliche Vorbereitung von Entscheidungsprozessen.

Es besteht Übereinstimmung darüber, daß das letztere eine zentrale Aufgabe der Evaluation ist. Um so erstaunlicher ist es, daß erst seit kurzem dieses Problem ausdrücklich thematisiert wird und daß bisher kaum Untersuchungen über Entscheidungsabläufe in diesem Bereich vorliegen. Es werden deskriptive Studien der Entscheidungsabläufe im pädagogischen Bereich gebraucht, die dazu beitragen können, den Entscheidungsprozeß in curricularen Fragen durchsichtig zu machen. Entscheidungsträger müssen befragt und beobachtet werden. Einige von ihnen sind fähig, den Prozeß, der zu den Entscheidungen führt, zu verbalisieren, bei anderen wird man nur versuchen, die Prinzipien und Kriterien der Urteile von ihrem Verhalten abzulesen. Beide Arten der Untersuchung können zur Entwicklung von Modellen und Kriterien der Entscheidungsabläufe führen. Darüber hinaus dürfte auch die Untersuchung von Entscheidungsabläufen unter experimentellen Bedingungen wertvolle Hilfe zur Entwicklung derartiger Modelle und Kriterien bieten. Ihrer Kenntnis bedarf der Evaluator, damit er die Ergebnisse der Evaluation so formulieren kann, daß sie den individuellen Bedingungen der Entscheidungssituation Rechnung tragen und die Informationen enthalten, die für die betreffenden Entscheidungen wichtig sind.

Cooler (o. J.) hat ein Verfahren entwickelt, das es einerseits dem Entscheidungsträger erleichtert, diejenigen Informationen zu finden, die er für die Entscheidung eines komplexen Problems benötigt, das andererseits dem Evaluator eine Möglichkeit bietet, die Datensammlung so anzulegen, daß sie die für den Entscheidungsprozeß relevanten Daten enthält. Es umfaßt in fünf Schritten die Identifikation

- der Gruppe von Hauptinteressenten,
- von Schlüsselfragen, die diese Gruppe hat,
- derjenigen Daten, die benötigt werden, um die Schlüsselfragen zu beantworten,

- von Quellen für diese Daten,
- von Methoden, diese Daten zu erhalten.

Je präziser die Vorstellungen des Entscheidungsträgers in einzelnen Punkten sind, desto größer ist die Hilfe für die Evaluatoren, die für die Entscheidung wichtigen Daten zu sammeln. Dabei empfiehlt sich eine Kooperation von Entscheidungsträgern und Evaluatoren zu einem möglichst frühen Zeitpunkt. Die hier entwickelte Methode bietet wenig mehr als eine Formalisierung des Datensammelungsprozesses, dessen Ergebnisse eine begründete Entscheidung ermöglichen. Weitere Untersuchungen zum Beitrag der Evaluation zu Entscheidungsprozessen werden dringend benötigt (vgl. Wulf 1972c). Dabei muß u. a. die von Stufflebeam und Gooler nicht problematisierte Kompetenz und Legitimität der Entscheidungsträger und die Effektivität ihrer Entscheidungen im Hinblick auf ihre Folgen thematisiert werden. Auch gilt es Prinzipien und Regeln zu entwickeln, die den Entscheidungsprozeß leiten, bzw. Kriterien, auf Grund derer er beurteilt werden kann. Darüber hinaus gilt es zu untersuchen, inwieweit in der Evaluation gewonnene Daten überhaupt Einfluß auf Entscheidungen haben.

V. Probleme der Curriculumevaluation

Curriculumevaluation steht in den USA und in der BRD vor gleichen Aufgaben und Problemen. Im Augenblick erscheint es notwendig, Begriffe und Modelle, Methoden und Instrumente der Evaluation zu entwickeln, um so die immer neu und präzise zu formulierenden Zielsetzungen der Evaluation einzelner Curricula zu verwirklichen. Die Ausweitung der zu berücksichtigenden Gesichtspunkte in der Curriculumevaluation in den letzten Jahren hat noch nicht die anstehenden Probleme gelöst, vielleicht jedoch ihre Lösung ermöglicht. Manche der dargestellten Ansätze sind es wert, aufgegriffen und weiterentwickelt zu werden.

Die Entwicklung von Theorie und Methodologie der Evaluation kann nur in engem Zusammenhang mit der tatsächlichen Evaluation von Curricula erfolgen. Das bedeutet: es muß wenigstens bei der Planung einiger Curricula das Schwergewicht auf ihre Evaluation gelegt werden mit dem Ziel, die Probleme der Evaluation zu erforschen. Dabei gilt es zugleich, die Aufmerksamkeit auf die Ausbildung von Wissenschaftlern zu richten, die sich der vielschichtigen Probleme der Evaluation bewußt sind und ihre zahlreichen Methoden beherrschen¹⁸. Im Rahmen einer zu entwickelnden Theorie und zu erarbeitender Methoden der Curriculumevaluation sind vor allem folgende Probleme zu lösen:

1. Welches ist die gesellschaftliche Funktion der Evaluation?
2. Wie wird die Beziehung zwischen Curriculum und Evaluation bestimmt?
3. Wie werden Kriterien für die Evaluation von Curricula gewonnen?
4. Wie wird das Verhältnis von Beschreibung und Werturteilen in der Evaluation bestimmt?
5. Inwieweit ist es notwendig, das Vorfeld, das Aktionsfeld und das Wirkungsfeld in den verschiedenen Phasen der Curriculumentwicklung zu evaluieren, und welche Methoden sind angemessen?
6. Welche Funktion hat Evaluation für die verschiedenen Ebenen der Entscheidung in den unterschiedlichen Phasen der Curriculumentwicklung?
7. Welchen Beitrag kann Evaluation zur Auswahl und Revision von Lernzielen, zur Entwicklung von Curricula, zur Entwicklung spezifischer Lernsituationen und -methoden, zur Optimierung des Lernens von spezifischen Schülergruppen und Individuen leisten?
8. Wie kann eine Technologie der Evaluation entwickelt werden, und welche Instrumente und Methoden muß sie ausarbeiten?