

Stake, Robert E.

Verschiedene Aspekte pädagogischer Evaluation

Wulf, Christoph [Hrsg.]: *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München : R. Piper & Co. Verlag 1972, S. 92-112. - (Erziehung in Wissenschaft und Praxis; 18)



Quellenangabe/ Reference:

Stake, Robert E.: Verschiedene Aspekte pädagogischer Evaluation - In: Wulf, Christoph [Hrsg.]: *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München : R. Piper & Co. Verlag 1972, S. 92-112 - URN: urn:nbn:de:0111-opus-14242 - DOI: 10.256656/01:1424

<https://nbn-resolving.org/urn:nbn:de:0111-opus-14242>

<https://doi.org/10.256656/01:1424>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, auführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Evaluation

Beschreibung und Bewertung von Unterricht,
Curricula und Schulversuchen

Texte

herausgegeben von Christoph Wulf



R. Piper & Co. Verlag
München

ISBN 3-492-01985-4
© R. Piper & Co. Verlag, München 1972
Gesamtherstellung Clausen & Bosse, Leck/Schleswig
Umschlagentwurf Gerhard M. Hotop
Printed in Germany

ROBERT E. STAKE

Verschiedene Aspekte pädagogischer Evaluation

Präsident Johnson, Präsident Conant, Mrs. Hull (Saras Lehrer) und Herr Tykoziner (der Mann nebenan) ähneln sich in ihrem Vertrauen auf Erziehung. Aber sie haben recht unterschiedliche Ideen darüber, was Erziehung ist. Der Wert, den sie der Erziehung beimessen, gibt keine Aufschlüsse über die Art, wie sie Erziehung bewerten. Genauso unterschiedliche Auffassungen haben Pädagogen über den Inhalt und Wert eines Bildungsprogramms. Die vielen Möglichkeiten in den Zielen und Methoden der Evaluation erlauben es jedem, seine eigene Perspektive zu behalten. Da viele Pädagogen ein zu begrenztes Verständnis von Evaluation haben, sehen nur wenige ihre eigenen Programme in voller Komplexität. Um den eigenen Unterricht besser zu verstehen und zur Verbesserung der Wissenschaft vom Unterricht beizutragen, sollte jeder Pädagoge sich sämtliche Möglichkeiten der Evaluation vergegenwärtigen. Pädagogische Evaluation hat ihre formalen und informalen Seiten. Informale Evaluation ist durch gelegentliche Beobachtungen, implizite Ziele, intuitive Normen und subjektive Urteile gekennzeichnet. Weil diese vielleicht auch im täglichen Leben üblich sind, kommt informale Evaluation zu Ergebnissen, die selten in Frage gestellt werden. Sorgfältige Untersuchungen zeigen, daß informale pädagogische Evaluation von unterschiedlicher Qualität ist; manchmal ist sie scharfsinnig und einsichtsvoll, manchmal oberflächlich und irreführend.

Formale pädagogische Evaluation ist durch Strichlisten, strukturierte Unterrichtsbeobachtungen, Vergleiche mit Kontrollgruppen und Untersuchungen von Schülern mit standardisierten Tests gekennzeichnet. Einige dieser Verfahren haben sich seit langem bewährt. Bei der Planung von Evaluation denken leider nur wenige Pädagogen an diese vier Verfahren. Es ist viel gebräuchlicher, informal zu evaluieren: den Lehrer nach seiner Meinung zu fragen, über die Logik des Programms nachzudenken oder das Ansehen seiner Befürworter in Betracht zu ziehen. Selten sucht man nach relevanten Forschungsberichten oder Verhaltensdaten, die das Ergebnis curricularer Entscheidungen sind.

Die Unzufriedenheit mit dem formalen Ansatz hat gute Gründe. Es gibt wenige wirklich relevante, lesbare Forschungsberichte. Die pädagogischen Zeitschriften sind nicht bereit, Evaluationsuntersuchungen zu veröffentlichen. Verhaltensdaten zu gewinnen, ist kostspielig; auch sie geben oft nicht die gesuchten Antworten. Zu vielen Pädagogen, die Unterrichtsbesichtigungen machen, fehlt eine entsprechende Schulung oder Erfahrung in Evaluation. Viele Strichlisten sind ungenau; einige betonen die äußeren Bedingungen einer Schule zu sehr. Psychometrische Tests werden eher dazu entwickelt, zwischen Schülern mit etwa gleicher Bildung zu differenzieren als die Auswirkungen des Unterrichts auf den Erwerb von Fertigkeiten und Verständnis zu erfassen. Ein moderner Pädagoge kann sich wenig auf formale Evaluation verlassen, weil ihre Ergebnisse selten die von ihm gestellten Fragen beantworten.

Der mögliche Beitrag formaler Evaluation

Die Skepsis des Pädagogen gegenüber formaler Evaluation resultiert zum Teil auch aus seiner Empfindlichkeit gegen Kritik. Häufig versteckt er sich hinter Begriffen wie »Innovationsphase« und »akademische Freiheit«, um Evaluation zu vermeiden. Die »Politik« der Evaluation ist ein interessantes Problem, das in diesem Zusammenhang jedoch nicht erörtert werden soll. Das Thema unserer Ausführungen ist der *mögliche* Beitrag formaler Evaluation zur Erziehung. Pädagogen sehen heute kaum, welche Hilfe formale Evaluation ihnen leisten könnte. Sie sollten Testkonstrukteure bitten, eine Methodologie zu entwickeln, die den Reichtum, die Komplexität und die Wichtigkeit ihrer Programme berücksichtigt. Das geschieht jedoch bisher noch nicht.

Wenn man die gegenwärtigen Bemühungen um formale Evaluation in der Pädagogik untersucht, findet man geringe Anstrengungen, die Voraussetzungen (antecedent conditions) und die Unterrichtsprozesse (transactions) – einige werden von Beobachter-Teams aufgezeichnet – zu erforschen; auch findet man zu wenig Versuche, sie mit den verschiedenen Ergebnissen – einige werden in konventionellen Testwerten ausgedrückt – in Verbindung zu bringen. Man hat selten versucht, das Verhältnis zwischen dem, was ein Pädagoge zu tun beabsichtigt, und dem, was er wirklich tut, zu erfassen. Das traditionelle Bemühen der Testkonstrukteure um die Reliabilität der Punktwerte individueller Schüler und die Voraussagevalidität (vgl. Lindquist 1951) ist ein zweifelhaftes Mittel. Bei der Evaluation von Curricula sollte man, anstatt die individuellen Unterschiede zwischen den Schülern zu betonen, besser die Kontingenzen zwischen den

Voraussetzungen, den Unterrichtsaktivitäten und den schulischen Ergebnissen beachten.

In diesem Beitrag soll nicht erörtert werden, was oder wie man messen sollte; es soll ein Hintergrund für die Entwicklung eines Evaluationsplans gegeben werden. Was und wie evaluiert wird, muß später entschieden werden. Mir geht es hier eher um Bildungsprogramme als um die Ergebnisse der Bildung. Ich setze voraus, daß der Wert eines Bildungsergebnisses auf dem verwendeten Programm beruht. Die Evaluation eines Programms schließt die Evaluation seiner Materialien ein.

Dieses Verständnis von pädagogischer Evaluation scheint sich zu verändern. Auf den folgenden Seiten will ich zeigen, welches Verständnis von Evaluation sich m. E. empfiehlt. Ich werde eine Konzeptualisierung der Evaluation zu entwickeln versuchen, die sich an dem komplexen und dynamischen Charakter der Erziehung orientiert und die die verschiedenen Zielsetzungen und Urteile des Praktikers angemessen berücksichtigt.

Ein großer Teil des neuerlichen Interesses an Curriculumevaluation liegt in den gegenwärtig umfangreichen Bemühungen um Curriculuminnovation begründet; aber die Ausführungen in diesem Beitrag gelten für herkömmliche und neue Curricula gleichermaßen. Sie betreffen z. B. Titel I- und Titel III-Projekte, die im Rahmen des Elementary and Secondary Education Act von 1965 finanziert worden sind. Die Erörterungen sind für alle Curricula relevant, unabhängig davon, ob sie sich mehr an den fachspezifischen Inhalten oder mehr an den Interessen der Schüler orientieren. Dabei ist es gleichgültig, ob das Curriculum allgemeine Zielsetzungen oder spezielle Förderungsaufgaben hat.

Ziele und Verfahren pädagogischer Evaluation sind von Fall zu Fall verschieden. Was für eine Schule angemessen ist, mag für eine andere weniger zweckmäßig sein. Manchmal empfehlen sich standardisierte Leistungstests und manchmal nicht. In einem Fall stehen geringe, im anderen Fall umfangreiche finanzielle Mittel zur Verfügung. Wie unterscheiden sich Ziele und Verfahren der Evaluation? Was sind die grundlegenden Charakteristika der Evaluation? In den folgenden Ausführungen werden sie als Evaluationshandlungen, Datenquellen, Kongruenz und Kontingenzen, Normen und Verwendungsarten der Evaluation identifiziert. Doch zuerst soll zwischen Beschreibung und Beurteilung in der Evaluation unterschieden werden.

Die Erwartung, die der Pädagoge an die Evaluation stellt, ist nicht gleich der Auffassung des Evaluators. Dieser erblickt seine Aufgabe in der Beschreibung von Einstellungen, Umwelt und Leistungen. Der Lehrer und der Beamte der Schulverwaltung jedoch erwarten vom Evaluator, daß er etwas oder jemanden nach seiner Leistung bewertet. Sodann erwarten

sie, daß er Dinge nach äußeren Normen beurteilt, vielleicht mit Kriterien, die nur eine geringe Beziehung zu den Mitteln und Zielen der örtlichen Schule haben.

Keiner begreift Evaluation umfassend genug. Beschreibung und Beurteilung sind erforderlich; sie sind in der Tat die beiden grundlegenden Evaluationshandlungen. Ein Evaluator kann versuchen, sich des Urteils oder der Sammlung von Urteilen anderer Personen zu enthalten. Ein anderer Evaluator kann ausschließlich darauf bedacht sein, den Wert des Programms deutlich zu machen. Aber die Evaluation beider ist unvollständig. Um vollständig verstanden zu werden, muß das Erziehungsprogramm vollständig beschrieben und beurteilt werden.

Auf dem Weg zu einer vollständigen Beschreibung

Der Evaluator scheint in zunehmendem Maße die Wichtigkeit einer vollständigen Beschreibung zu betonen. Seit vielen Jahren evaluiert er vor allem dadurch, daß er feststellt, inwieweit Lernziele von Schülern erreicht worden sind. Diese Lernziele wurden gewöhnlich mit den traditionellen Disziplinen, z. B. Mathematik, Englisch und Politischer Bildung (social studies) gleichgesetzt. Standardisierte oder vom Lehrer hergestellte Leistungstests hielt man für nützlich, um das Ausmaß zu beschreiben, in dem einige curriculare Lernziele von einzelnen Schülern in einem speziellen Kurs erreicht wurden. In diesem frühen Stadium war Evaluation für viele Evaluatoren und Pädagogen nichts anderes als der Einsatz und die normative Interpretation von Leistungstests.

In den letzten Jahren haben darüber hinaus einige Evaluatoren abzuschätzen versucht, inwieweit einzelne Schüler bestimmte interdisziplinäre und extracurriculare Ziele erreicht haben. Dabei bestand ihr Ziel vor allem darin, die Integration von Verhaltensweisen in Individuen festzustellen, das Verständnis der Beziehungen zwischen den wissenschaftlichen Disziplinen zu erfassen und die Entwicklung von Haltungen, Fertigkeiten und Einstellungen zu untersuchen, die ein Individuum dazu befähigen, ein Handwerker oder Wissenschaftler zu sein. Für die beschreibende Evaluation solcher Ergebnisse hat die Eight-Year-Study (Smith/Tyler 1942) als Modell gedient. Das National Assessment Program wird vielleicht – wie aus den in einem Zwischenbericht erschienenen folgenden Ausführungen hervorgeht – zu einem anderen Modell werden: »... Alle Kommissionen arbeiteten innerhalb der folgenden allgemeinen Definition des National Assessment.

1. Um in etwa die Ziele der Erziehung in den USA zu reflektieren, sollte

das National Assessment traditionelle und moderne Curricula ins Auge fassen, alle Zielsetzungen berücksichtigen, die die Schulen für die Entwicklung von Einstellungen, Motivation, Kenntnissen und Fertigkeiten haben« (Educational Testing Service 1965).

In seinem Beitrag *Evaluation zur Verbesserung von Curricula* forderte Lee Cronbach (1963) eine möglichst umfangreiche Einbeziehung verhaltenswissenschaftlicher Variablen, um die Ursachen und Auswirkungen von gutem Unterricht zu untersuchen. Nach seinem Vorschlag liegt das Hauptziel der Evaluation darin, solche dauerhaften Beziehungen zu entdecken, die für die Entwicklung zukünftiger Bildungsprogramme relevant sind. Die traditionelle Beschreibung der Schülerleistung ergänzen wir durch die Beschreibung des Unterrichts und die Beschreibung der Beziehungen zwischen ihnen. Wie der Bildungsforscher versucht nach unserer Auffassung auch der Evaluator, Generalisationen über pädagogische Praktiken zu entwickeln. Viele Evaluatoren von Curriculumprojekten haben sich diese Definition von Evaluation zu eigen gemacht.

Die Rolle des Urteils

Beschreibung ist eine Sache, Beurteilung eine andere. Die meisten Evaluatoren haben sich entschlossen, keine Urteile abzugeben. In seiner kürzlich erschienenen *Methodologie der Evaluation* hat Michael Scriven (1967) jedoch den Evaluatoren die Aufgabe zugeschrieben, Urteile über den Wert einer pädagogischen Handlung zu fällen. Er hat vom Evaluator verlangt, die Erwartungen zu erfüllen, die Pädagogen an ihn stellen. Nach Scrivens Ansicht findet Evaluation nicht statt, bevor nicht Beurteilung erfolgt. Wenn der Evaluator sich dessen bewußt ist, ist er am besten zur Abgabe von Urteilen qualifiziert.

Aufgrund seiner zahlreichen Erfahrungen und Kenntnisse in diesem Bereich der Forschung und pädagogischen Praxis ist der Evaluator wenigstens teilweise dazu befähigt, Urteile abzugeben. Aber soll er wirklich diese Aufgabe übernehmen? Selbst zur Zeit, da wenige Evaluatoren bereit sind, Urteile zu fällen, wehren Pädagogen sich gegen formale Evaluation. Wenn man die Funktion der Evaluatoren häufiger mit der Aufgabe identifiziert, Urteile über den Unterschied zwischen schlechteren und besseren Programmen, über die Gewährung von Unterstützung und über die Formulierung von Kritik abzugeben, würde sich der Zugang der Evaluatoren zu Daten wahrscheinlich erschweren. Evaluatoren arbeiten mit anderen Sozialwissenschaftlern und Verhaltensforschern zusammen. Alle Forscher, die keine Urteile fällen wollen, bedauern die Übernahme dieser Aufgabe durch ihre

Kollegen. Sie sind der Überzeugung, daß viele Praktiker noch mehr Einwände als bisher gegen die Sozialwissenschaften und die verhaltenswissenschaftliche Forschung erheben würden.

Viele Evaluatoren glauben, sie seien nicht in der Lage – wozu ihrer Meinung nach ein Beurteiler fähig sein sollte –, den eindimensionalen Wert alternativer Programme wahrzunehmen. Sie erwarten z. B. folgendes Dilemma: Curriculum I hat als Ergebnis drei Fertigkeiten und zehn Erkenntnisse, Curriculum II vier Fertigkeiten und acht Erkenntnisse. Sie scheuen sich, zu beurteilen, ob der Gewinn einer Fertigkeit den Verlust von zwei Erkenntnissen wert ist. So bestärkt der Evaluator, sei es aus Ängstlichkeit, Desinteresse oder aufgrund rationaler Entscheidung, häufig die Entscheidungen der Gemeinden, ihr Recht, eigene Normen aufzustellen und den Wert ihres Bildungssystems selbst zu beurteilen. Er setzt voraus, daß das, was für eine Gemeinde gut ist, auch für eine andere gut sein muß; er traut sich nicht zu, eine Entscheidung darüber zu fällen, was für eine ihm erst seit geraumer Zeit bekannte Gemeinde am besten ist.

Scriven macht darauf aufmerksam, daß gegenwärtig wenige und in Zukunft noch weniger Evaluatoren komplexe Curricula beurteilen können. Verschiedene Entscheidungen müssen getroffen werden, z. B. ob das Physical Science Study Committee Program oder das Harvard Physical Program unterrichtet werden soll. Sie sollen jedoch nicht aufgrund trivialer Kriterien – beispielsweise Erwähnung in der Presse, Persönlichkeit des Vertreters des Projekts, administrative Bequemlichkeit oder pädagogischer Mythos – getroffen werden. Wer soll Urteile fällen? Scriven findet die Antwort z. T. so leicht, weil er zwischen Schüler und Curriculum wenige Interaktionen erwartet. D. h.: er geht davon aus, daß das, was für einen Schüler – wenigstens in groben Umrissen – am besten ist, auch für andere am besten sein muß. Er setzt ferner voraus, daß, wenn die Interessen einer Gemeinde sich nicht mit denen der Gesamtgesellschaft decken, erstere denen der Gesamtgesellschaft abträglich sind, so daß daher das freie Entscheidungsrecht eingeschränkt werden muß. – Nach Scriven muß der Evaluator Urteile fällen.

Ob die Evaluatoren Scrivens Aufforderung berücksichtigen oder nicht, bleibt abzuwarten. Wahrscheinlich werden jedoch Beurteilungen einen zunehmend größeren Teil des Evaluationsberichts ausmachen. Die Evaluatoren werden sich darum bemühen, die Ansichten von qualifizierten Personen aufzuzeichnen. Obwohl diese Auffassungen subjektiv sind, können sie sehr nützlich sein und objektiv gesammelt werden, d. h. unabhängig von denen, die diese Ansicht vertreten. Der Evaluator kann sich eher der Aufgabe unterziehen, Urteile bei der Evaluation zu verwerten als sie selber abzugeben.

Taylor und Maguire (1966) haben fünf Gruppen genannt, deren Ansichten über Erziehung wichtig sind: Sprecher der Gesamtgesellschaft, Fachwissenschaftler, Lehrer, Eltern und Schüler. Die Urteile der Vertreter dieser und anderer Gruppen sollten gehört werden. Oberflächliche Umfragen, Briefe an den Herausgeber und andere beiläufig geäußerte Urteile sind unzureichend. Die Evaluation sollte den Wert und die Unzulänglichkeiten eines Schulprogramms nach dem Urteil gut informierter Gruppen mit Hilfe systematisch gesammelter und verarbeiteter Daten deutlich machen. D. h. also: Urteilsdaten und Beschreibungsdaten sind gleichermaßen für die Evaluation von Bildungsprogrammen erforderlich.

Datenmatrizen

Um evaluieren zu können, muß ein Pädagoge bestimmte Daten sammeln. Sie werden wahrscheinlich in zahlreichen heterogenen Bereichen mit mehreren verschiedenartigen Mitteln gewonnen. Unabhängig davon, ob das unmittelbare Ziel Beschreibung oder Beurteilung ist, sollten Informationen in drei Bereichen gesammelt werden. Im Evaluationsbericht empfiehlt es sich, zwischen *Voraussetzungsdaten*, *Prozeßdaten* und *Ergebnisdaten* zu unterscheiden.

Eine Voraussetzung ist jede Bedingung, die vor dem Unterrichten und Lernen besteht und die Einfluß auf die Ergebnisse haben kann. Die Ausgangssituation eines Schülers vor dem Unterricht, z. B. seine Fähigkeit, vorherige Erfahrung, Interesse und Bereitschaft, ist eine komplexe Voraussetzung. Beim Programmierten Unterricht nennt man einige Voraussetzungen »Eingangsverhalten«. Die »akkreditierende Institution des Staates« wiederum richtet ihre besondere Aufmerksamkeit auf die Investition der Ressourcen der Gemeinde. Dies sind Beispiele für die Voraussetzungen, die ein Evaluator beschreiben kann.

Prozesse sind die zahllosen Begegnungen zwischen Schülern und Lehrern, Schülern und Schülern, Autoren und Lesern, Eltern und Schulpsychologen – das Aufeinanderfolgen von pädagogischen Handlungen, das den Prozeß der Erziehung ausmacht. Beispiele sind die Vorführung eines Films, eine Unterrichtsdiskussion, die Lösung einer Hausaufgabe, eine Erklärung am Rand einer Prüfungsarbeit und der Einsatz eines Tests. Smith und Meux (1962) haben solche Prozesse genau untersucht und dazu ein 18 Kategorien umfassendes Klassifikationssystem aufgestellt. Auf eine bestimmte Art von Prozessen wurde durch die Förderung der Entwicklung audiovisueller Medien im Rahmen des National Defense Education Act besonderer Wert gelegt.

Während Voraussetzungen und Ergebnisse relativ statisch sind, sind Prozesse dynamisch. Die Grenzen zwischen den Bereichen sind nicht deutlich. Wir können z. B. während eines Prozesses bestimmte Ergebnisse identifizieren, die als Rückmeldung Voraussetzung für nachfolgendes Lernen sind. Die Abgrenzung zwischen den Bereichen braucht nicht exakt zu erfolgen. Die Kategorien sollen eher dazu dienen, eine umfangreiche Sammlung von Daten anzuregen, als sie in Gruppen zu unterteilen.

In der Vergangenheit konzentrierte man sich bei formaler Evaluation vorwiegend auf Ergebnisse wie Fähigkeiten, Leistungen, Einstellungen und Erwartungen der Schüler, die sie aufgrund einer pädagogischen Erfahrung gewonnen hatten. Versteht man unter Ergebnissen aber den gesamten Bereich der dazu gehörenden Informationen, müßte man auch die Auswirkungen des Unterrichts auf Lehrer, Verwaltungsbeamte, Schulpsychologen und andere untersuchen. Hierzu gehören auch Daten über die Abnutzung der Ausstattung, den Einfluß der Lernbedingungen und die Kosten. Bei der Evaluation müssen außer den nachweisbaren oder sogar deutlich greifbaren Ergebnissen auch die Anwendung des Gelernten, der Transfer und die Auswirkungen wiederholenden Lernens, die sich vielleicht erst viel später messen lassen, berücksichtigt werden. Die Beschreibung der Ergebnisse von Fahrschulunterricht z. B. könnte sinnvollerweise Berichte darüber enthalten, inwieweit jemand im Laufe seines Lebens

rationale

Begründung

dung

Intentionen

Beobachtungen

Normen

Urteile

			Voraussetzungen		
			Prozesse		
			Ergebnisse		

Beschreibungsmatrix

Urteilmatrix

Abb. 1: Eine Matrix für Daten, die vom Evaluator eines Bildungsprogramms gesammelt werden sollen

Unfälle vermieden hat. Ergebnisse sind also, kurz gesagt, die unmittelbaren und langfristigen kognitiven und affektiven, persönlichen und gesellschaftlichen Folgen der Erziehung.

Voraussetzungen, Prozesse und Ergebnisse sind Elemente der Evaluationsmatrix und müssen – wie Abbildung 1 zeigt – bei der Beschreibung und Beurteilung berücksichtigt werden. Um diese Matrix auszufüllen, sammelt der Evaluator Urteile und Beschreibungen, z. B. über Vorurteile in der Gemeinde, über Stile des Problemlösens und die Persönlichkeit des Lehrers. Aus Abbildung 1 geht auch hervor, daß Urteile entweder als allgemeine Qualitätsnormen oder als spezifische Urteile über ein gegebenes Programm klassifiziert werden. Beschreibende Daten werden als Intentionen und Beobachtungen klassifiziert. Der Evaluator kann die Sammlung seiner Daten entsprechend den Kategorien der Abbildung 1 organisieren.

Der Evaluator kann aufzeichnen, was Pädagogen beabsichtigen und Beobachter wahrnehmen, was die verantwortlichen Geldgeber im allgemeinen erwarten und wie Beurteiler das gegenwärtige Programm bewerten. Die Aufzeichnung kann versuchen, Voraussetzungen, Prozesse und Ergebnisse getrennt innerhalb der vier Gruppen als *Intentionen, Beobachtungen, Normen und Urteile* zu identifizieren (vgl. Abb. 1). Die folgenden Ausführungen liefern ein Beispiel für zwölf Daten, die in die zwölf Felder eingetragen werden können. Sie beginnen mit einer intendierten Voraussetzung und gehen jede Spalte hinunter, bis ein Urteil über die Ergebnisse gefällt worden ist.

Im Wissen, daß

(1) Kapitel 11 als Aufgabe aufgegeben worden ist, und daß er beabsichtigt (2), am Mittwoch über das Thema eine Vorlesung zu halten, gibt ein Professor an (3), was die Studenten bis zum Freitag können sollen – z. T. dadurch, daß er einen Fragebogen über das Thema bearbeiten läßt. Er beobachtet, (4) daß einige Studenten am Mittwoch abwesend waren, (5) daß er wegen der langen Diskussion nicht die Vorlesung beenden konnte und (6) daß einen wichtigen Begriff im Fragebogen nur zwei Drittel der Hörer zu verstehen schienen. Im allgemeinen erwartet er (7), daß einige abwesend sind, aber daß das Versäumnis durch die für den Fragebogen aufgewandte Zeit aufgeholt wird; er erwartet (8), daß seine Vorlesungen für etwa 90 % der Zuhörer so klar sind, daß sie ohne Schwierigkeiten verstehen können; und er weiß (9), daß seine Kollegen erwarten, daß nur einer von zehn Studenten alle wichtigen Begriffe in solchen Vorlesungen versteht. Nach seinem Urteil bot (10) die aufgegebenene Lektüre keine ausreichenden Hintergrundinformationen für seine Vorlesung; Studenten äußerten (11), daß die Vorlesung provokativ war; der Hilfsassistent, der die Fragebogen las, sagte (12), daß eine entmutigend große Zahl der Studenten einen wichtigen Begriff mit einem anderen zu verwechseln schien.

Nicht einmal für die ferne Zukunft erwarten Pädagogen und Evaluator,den,

daß Daten so genau aufgezeichnet werden. Meine Absicht war es, hier zwölf Beispiele für Daten zu geben, die den zwölf verschiedenen Feldern in der Matrix zugeordnet werden können. Im folgenden möchte ich die Matrix für die Beschreibungsdaten erläutern.

Ziele und Intentionen

Seit vielen Jahren sind Unterrichtstechnologen und Testkonstrukteure für eine stärker explizite Formulierung der pädagogischen Ziele eingetreten. Nach meiner Auffassung sind Ziele, Lernziele und Intentionen synonym. Ich benutze als Bezeichnung der Kategorie *Intentionen*, weil heute viele Pädagogen »Ziele« und »Lernziele« mit »intendiertem Schülerverhalten« gleichsetzen. In diesem Beitrag umfassen Intentionen die intendierten Bedingungen der Umwelt, die geplanten Demonstrationen, die beabsichtigte Behandlung von bestimmten fachspezifischen Inhalten und das angestrebte Schülerverhalten. Zu den drei Feldern dieser Reihe gehören erwünschte, erhoffte, erwartete und sogar befürchtete Auswirkungen. Diese Datengruppe enthält die Ziele und Pläne anderer Personen, vor allem jedoch die der Schüler. (Man sollte bedenken, daß Pädagogen nicht das Recht haben, die Untersuchung einer Variablen dadurch auszuschließen, daß sie nicht als ein Lernziel angesehen wird. Der Evaluator sollte die Variable und ihre Ablehnung erfassen.) Die daraus sich ergebende Sammlung der *Intentionen* ist eine – nach Priorität geordnete – Aufzeichnung aller möglicherweise eintretenden Ereignisse.

Die Tatsache, daß viele Pädagogen heute die »Ziele« mit den »intendierten Schülerverhaltensweisen« gleichsetzen, geht auf die Behavioristen, vor allem jedoch auf die Vertreter des Programmierten Unterrichts zurück. Indem sie das Schwergewicht auf die spezifischen Unterrichtshandlungen und -übungen gelegt haben, die zur Verbesserung der Schülerantworten beitragen, haben sie eine gewisse Reform des Unterrichts bewirkt. Das American Association for the Advancement of Science Elementary Project (A. A. S. S.) hat z. B. sein Curriculum erfolgreich mit Hilfe von Verhaltenszielen entwickelt. Einige innovative Curriculumprojekte haben jedoch festgestellt, daß die Betonung behavioristischer Ergebnisse für kreativen Unterricht hinderlich ist (vgl. Atkin 1963). Der pädagogische Evaluator sollte Ziele nicht nur als erwartetes Schülerverhalten formulieren. Um ein Bildungsprogramm zu *evaluieren*, muß man untersuchen, was gelehrt und gelernt werden soll. (Viele Voraussetzungen und viele Unterrichtsprozesse können auf Wunsch behavioristisch formuliert werden.) Wie Intentionen formuliert werden, ist kein Kriterium für ihre Berücksichtigung bei der

Evaluation. Intentionen können die allgemeinen Ziele der »Educational Policies Commission« oder die detaillierten Ziele der Hersteller von Programmen sein (vgl. Mager 1962). Taxonomische, mechanistische, humanistische, biblische – alle noch so verschiedenartigen Zielformulierungen müssen bei der Evaluation berücksichtigt werden.

Bei dem Versuch, die Ziele des Pädagogen aufzuzeichnen, stößt ein Evaluator gegenwärtig auf Schwierigkeiten. Zu Beginn seiner Arbeit fordert er den Pädagogen auf, seine Lernziele so darzulegen, daß Verfahren für das Testen der Ergebnisse entwickelt werden können. Dabei erfährt er, daß der Pädagoge sich entweder sträubt oder unfähig ist, seine Ziele zu verbalisieren. Obwohl der Evaluator das Formulieren von Verhaltenszielen für die Aufgabe des Pädagogen hält, hilft er ihm sorgfältig und gern dabei. Nach unserer Auffassung ist es jedoch nicht Aufgabe des Pädagogen, Verhaltensziele zu formulieren. In Übereinstimmung mit Scrivens Ausführungen liegt u. E. die Beschreibung curricularer Lernziele beim Evaluator. Er ist mit der Terminologie des Verhaltens und seiner Ausdrucksformen vertraut. So wie es seine Aufgabe ist, die Verhaltensweisen eines Lehrers und die Antworten eines Schülers in Daten umzuformen, muß er auch die Intentionen und Erwartungen eines Pädagogen in Daten transformieren. Wiederholt muß der Evaluator den Pädagogen bitten, seine Intentionen zu äußern. Er sollte versuchen, die Zahl der Antworten durch Fragen zu erhöhen, wie: »Kann man es auch so sagen? Ist das ein Beispiel für das, was Sie meinen?« Natürlich kann der Evaluator den interessierten Pädagogen über Verhaltensziele unterrichten. Das kann seine Arbeit erleichtern. Darauf zu bestehen, daß jeder Pädagoge Verhaltensziele verwendet, ist jedoch falsch.

Authentische Formulierungen der Intentionen zu erhalten, ist für den Evaluator eine schwierige Aufgabe. Die benötigte Methodologie muß noch entwickelt werden. Im weiteren soll nun die zweite Reihe der Datenfelder behandelt werden.

Die Auswahl von Methoden der Beschreibung

Die meisten deskriptiven Daten, die am Anfang des vorherigen Abschnitts erwähnt wurden, werden als *Beobachtungen* klassifiziert. Wenn der Evaluator¹ die Voraussetzungen, Prozesse und die daraus sich ergebenden Folgen beschreibt, gibt er (nach Abb. 1) seine Beobachtungen wieder. Manchmal macht er die Beobachtungen direkt und persönlich, manchmal benutzt er Instrumente. Zu seinen Instrumenten gehören Inventurverzeichnisse, Listen mit bibliographischen Daten, Routine-Interviews, Strichlisten, Fragebogen zur Erforschung von Meinungen und alle Arten psy-

chometrischer Tests. Der erfahrene Evaluator konzentriert seine Aufmerksamkeit auf das Messen der Schülerleistungen; aber er beobachtet auch die anderen Ergebnisse, Voraussetzungen und unterrichtlichen Prozesse.

Viele Pädagogen fürchten, daß der von außen kommende Evaluator nicht die Merkmale berücksichtigt, die nach dem Urteil des Lehrerkollegiums die wichtigsten sind. Dies trifft manchmal zu; oft aber richten Evaluatoren *zuviel* Aufmerksamkeit auf das, was sie beobachten sollen, und zuwenig Aufmerksamkeit auf andere Dinge. Bei der Auswahl der Variablen für die Evaluation muß der Evaluator eine subjektive Entscheidung treffen. Selbstverständlich muß er für eine Untersuchung die Zahl der Elemente begrenzen. Alle Elemente, die nicht berücksichtigt werden, tragen nach seinem Urteil nicht zum Verständnis des pädagogischen Geschehens bei. Der Evaluator sollte die Variablen besonders beachten, die durch die Lernziele des Pädagogen angegeben werden. Darüber hinaus muß er aber auch zusätzliche Variablen beobachten und die ungewollten Nebenwirkungen und zufälligen Ergebnisse untersuchen. Der Evaluator hat die Beobachtungsgegenstände und Meßverfahren auszuwählen.

Ohne die rationale Begründung (rationale) des Programms darzulegen, ist eine Evaluation nicht vollständig. Sie muß gesondert berücksichtigt werden (vgl. Abb. 1). Jedes Programm enthält eine allerdings oft nur implizite Begründung. Sie macht den philosophischen Hintergrund und die grundlegenden Ziele des Programms deutlich. Berlack (1966) hat dargelegt, wie wichtig die rationale Begründung für die Evaluation ist. Sie soll eine Grundlage für die Evaluation der Intentionen bieten. Der Evaluator muß sich oder anderen Beurteilern die Frage stellen, ob der von den Pädagogen entwickelte Plan einen logischen Schritt zur Implementation der grundlegenden Ziele darstellt. Die Begründung ist auch für die Wahl der Personen wichtig, die das Programm verwenden sollen, z. B. für die Mathematiker und Mathematiklehrer, die später verschiedene Aspekte des Programms beurteilen sollen.

Eine Formulierung der Begründung zu erhalten ist oft schwer. Häufig ist ein effektiver Lehrer beim Formulieren der Begründung für sein pädagogisches Handeln recht uneffektiv. Wenn er gedrängt wird, kann er schließlich vielleicht das sagen, was man von ihm erwartet. Die Begründung sollte jedoch in der Sprache des Pädagogen formuliert werden. Die Vorschläge des Evaluators können leicht hinderlich werden, da sie vielleicht übernommen werden, weil sie attraktiv sind, ohne jedoch die wirklichen Gründe für die Handlungen des Pädagogen anzugeben.

Die Urteilmatrix bedarf weiterer Erläuterung. Aber ich verschiebe das bis nach der Behandlung der Grundlagen für die Verarbeitung deskriptiver Daten.

Kontingenz und Kongruenz

Um deskriptive Evaluationsdaten zu verarbeiten, gibt es für jedes Bildungsprogramm zwei wichtige Verfahren. Man muß die Kontingenzen zwischen den Voraussetzungen, Prozessen und Ergebnissen und die Kongruenz zwischen den Intentionen und Beobachtungen finden. Die Verarbeitung der Urteile folgt einem anderen Modell. Die ersten beiden Spalten der Datenmatrix in Abbildung 1 enthalten die deskriptiven Daten. Das Schema für die Verarbeitung dieser Daten ist in Abbildung 2 dargestellt. Wenn das, was intendiert ist, wirklich geschieht, sind die Daten für ein Curriculum *kongruent*. Um vollständig kongruent zu sein, müssen die intendierten Voraussetzungen, Prozesse und Ergebnisse eintreten. (Das geschieht selten – und oft sollte es nicht geschehen.) Innerhalb einer Reihe der Datenmatrix sollte der Evaluator die Felder, die Intentionen und Beobachtungen enthalten, vergleichen, um Diskrepanzen festzustellen und den Grad der Kongruenz in dieser Reihe zu beschreiben (die Wichtigkeit der Kongruenz der Ergebnisse wurde in dem von Taylor/Maguire (1966) erarbeiteten Evaluationsmodell hervorgehoben). Die Kongruenz gibt keinen Hinweis darauf, ob die Ergebnisse reliabel oder valide sind, sondern lediglich darauf, daß das Intendierte eintrat.

Ähnlich dem Gestaltpsychologen, der in dem Ganzen mehr findet als die Summe seiner Teile, findet der Evaluator bei der Untersuchung der Variablen von beliebigen zwei der drei Bereiche in einer Spalte der Datenmatrix mehr zu beschreiben als die Variablen selbst. Die Beziehungen oder *Kontingenzen* zwischen den Variablen verdienen zusätzliche Aufmerksamkeit. Insofern als Evaluation die Suche nach Beziehungen ist, die die Verbesserung der Erziehung ermöglichen, ist es die Aufgabe des Evaluators, die Ergebnisse zu identifizieren, die mit bestimmten Voraussetzungen und Unterrichtsprozessen kontingent sind.

Unterrichtsplanung und Curriculumrevision haben in den letzten Jahren auf dem Vertrauen in bestimmte Kontingenzen beruht. Täglich organisiert der gute Lehrer seinen Unterricht und wählt seine Curriculummateriale entsprechend seinen unterrichtlichen Zielen aus. Für ihn sind die Kontingenzen in der Hauptsache logisch intuitiv, die durch zahlreiche befriedigende und bestätigende Erfahrungen unterstützt werden. Sogar der erfahrene und zweifellos der weniger erfahrene Lehrer müssen ihre intuitiv erwarteten Kontingenzen der Überprüfung durch Evaluatoren unterziehen.

Als erster Schritt in der Evaluation ist es wichtig, die Kontingenzen aufzuzeichnen. Ein Film über eine Überschwemmung kann dazu dienen (intendierter Prozeß), Schülern einen Hintergrund für eine entsprechende Schutzgesetzgebung (intendiertes Ergebnis) zu geben. Denen, die die

Beschreibende Daten

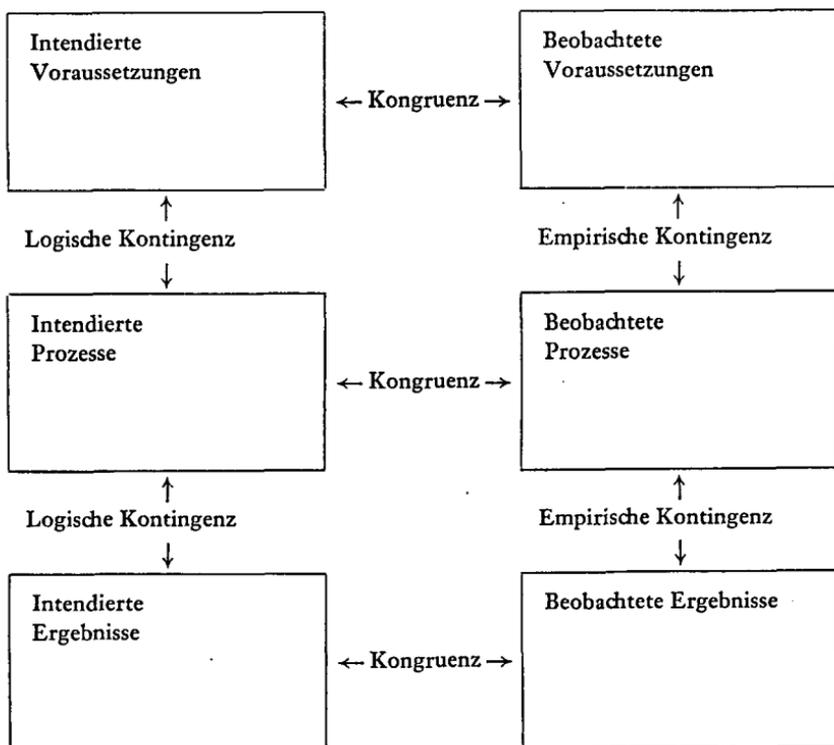


Abb. 2: Eine Darstellung des Prozesses der Verarbeitung von beschreibenden Daten

Sach- und die pädagogischen Probleme kennen, kann man folgende Frage stellen: »Gibt es eine logische Verbindung zwischen diesem Prozeß und dieser Zielsetzung?« Wenn es sie gibt, existiert eine logische Kontingenzenz zwischen diesen beiden Intentionen. Die Aufzeichnung sollte sie deutlich machen. Bei der Evaluation von Intentionen ist das Kontingenzenzkriterium immer ein Kriterium der Logik. Um die Logik einer pädagogischen Kontingenzenz zu testen, greifen die Evaluatoren auf ihre vorherigen Erfahrungen und vielleicht auch auf Forschungserfahrungen mit ähnlichen Erscheinungen zurück. Keine unmittelbare Beobachtung dieser Variablen ist jedoch erforderlich, um die Stärke der Kontingenzenzen zwischen Intentionen zu testen.

Die Evaluation von Beobachtungskontingenzenzen hängt von der empirischen Evidenz ab. Um sagen zu können: Diese Klasse macht im Rechnen

schnelle Fortschritte, weil der Lehrer gute, aber nicht zu differenzierte Kenntnisse in Mathematik hat, braucht man empirische Daten aus der Evaluation oder aus der Forschungsliteratur (vgl. Bassam 1962). Aus der Evaluation eines einzigen Programms allein kann man nicht die Daten erhalten, die für die Formulierung einer Kontingenz notwendig sind. Dazu bilden auch frühere Erfahrungen mit ähnlichen Erscheinungen eine grundlegende Qualifikation für den Evaluator.

Die Kontingenzen und Kongruenzen, die die Evaluatoren identifiziert haben, müssen genauso wie einheitliche Beobachtungsdaten der Beurteilung von Experten und Teilnehmern unterzogen werden. Das Auftreten einer *Nicht-Kongruenz* wird entsprechend den unterschiedlichen Standpunkten verschieden bewertet werden. So sind vielleicht ein Schulrat und ein Schulpsychologe verschiedener Ansicht darüber, welche Bedeutung der Streichung von Stunden für die im Stundenplan vorgesehene Sexualhygiene zukommt. Für die Beurteilung von Kontingenzen bietet sich als Beispiel das Ausmaß an, in dem die Lehrfähigkeit eines Lehrers während eines ganzen Schultages kontingent ist; darin kann ein Beurteiler einen ausreichenden Grund dafür sehen, auf eine Unterrichtsstunde am frühen Morgen zu verzichten, ein anderer jedoch nicht. Die Vorstellungen, die über die Bedeutung von Kongruenz und Kontingenz bestehen, müssen vom Evaluator sorgfältig untersucht werden.

Normen und Urteile

Nach allgemeinem Konsens besteht das Ziel der Erziehung in der optimalen Bildung der Schüler. Wie und unter welchen Umständen die Schüler sie jedoch erreichen, wird immer umstritten sein. Unabhängig davon, ob die Ziele von den örtlichen Gemeinden aufgestellt werden und nur für sie gelten oder ob sie für das ganze Land gelten sollen, erfordert die Evaluation des Bildungsniveaus eher explizite als implizite Normen (standards). Die gegenwärtigen Bildungsprogramme werden keiner normorientierten Evaluation unterzogen. Das bedeutet nicht, daß sich die Schulen nicht anstrengen oder keine Erfolge erzielen; es bedeutet lediglich, daß Normen – allgemeingültige Verhaltensformen – nicht überall gebräuchlich sind. Selbst wenn Schulen in allen Teilen des Landes die gleichen Evaluationsbogen verwenden², wird die Interpretation der gewonnenen Daten mit unklaren, individuell gebrauchten Begriffen erfolgen. Sogar im Rahmen informaler Evaluation kann keine Schule die Auswirkungen ihres Curriculum evaluieren, ohne zu wissen, wie andere Schulen ähnliche Lernziele zu erreichen versuchen. Leider wehren sich viele Pädagogen gegen die sy-

stematische Sammlung solcher Kenntnisse (vgl. Hand 1965 u. Tyler 1965). Über die Qualität der Erziehung eines Schülers weiß man gegenwärtig wenig. Die Schulzensuren beruhen auf den privaten Kriterien und Normen eines einzelnen Lehrers. Die meisten Werte in standardisierten Tests geben eher Auskunft darüber, wo ein Schüler bei der Lösung psychometrisch brauchbarer Aufgaben im Verhältnis zu seiner Bezugsgruppe steht; als über das Ausmaß an Kompetenz, mit der er wesentliche schulische Aufgaben erfüllt. Obwohl die meisten Lehrer in der Lage sind, ihre Fächer zu unterrichten und Lernschwierigkeiten zu erkennen, haben nur wenige die Fähigkeit, zu *beschreiben*, wie ein Schüler sich mit seiner geistigen Umwelt auseinandersetzt. Weder Schulzensuren noch Punktwerte in standardisierten Tests, noch die Ansichten der Lehrer enthalten genügend Informationen über das Bildungsniveau der Schüler.

Selbst wenn die Meßwerte erfolgreich interpretiert werden, ist Evaluation aufgrund der zahlreichen Normen schwierig. Die Normen unterscheiden sich von Schüler zu Schüler, Lehrer zu Lehrer und Bezugsgruppe zu Bezugsgruppe, und das ist auch richtig so. In einer pluralistischen Gesellschaft haben verschiedene Gruppen unterschiedliche Normen. Die Aufgabe der Evaluation besteht zum Teil darin, deutlich zu machen, wer welche Normen hat.

Es wurde bereits dargelegt, daß im Verlauf eines längeren Zeitraums die *Intentionen* eines Pädagogen sich ändern. Das bedeutet, daß sich während des Unterrichts die Kriterien und die Normen des Pädagogen wandeln. Während der Entwicklung und Dissemination eines Curriculum ändern sich sogar die Hauptgruppen der Kriterien. In einer umfassenden Analyse des Prozesses, in dem neue Curricula an die speziellen Bedingungen der einzelnen Schulen adaptiert werden, identifizierten Clark und Guba (1965) acht Stadien der Veränderung. Für jedes Stadium erarbeiteten sie spezifische Kriterien (jedes mit seinen eigenen Normen), aufgrund derer das Curriculum evaluiert werden soll, bevor man zum nächsten Stadium fortschreitet. Alle ihre Kriterien bedürfen weiterer Ausführung; hier soll nur angedeutet werden, daß es in jedem aufeinanderfolgenden Stadium der Curriculumentwicklung recht unterschiedliche Kriterien gibt. In der informalen Evaluation werden die Kriterien oft unspezifiziert gelassen. Formale Evaluation ist spezifischer. Je sorgfältiger die Evaluation ist, desto weniger Kriterien scheint es zu geben; je sorgfältiger die Kriterien spezifiziert sind, desto weniger wird die Angemessenheit der ihnen zugrunde liegenden Normen beachtet. Leider haben die am besten ausgebildeten Evaluatoren die Erziehung mit einem Mikroskop anstatt mit einem Weitwinkelsucher untersucht.

Es gibt keine genauen Kenntnisse darüber, was Schulen und Curriculum-

projekte gegenwärtig leisten; z. T. liegt es daran, daß die Methoden für die Verarbeitung von Urteilsdaten unzulänglich sind. Bei dem gegenwärtig geringen Ausmaß an formaler Evaluation berücksichtigt man zu wenig Kriterien, ist zu tolerant für implizite Normen und kümmert sich nicht um die Vorteile relativer Vergleiche. Es bedarf weiterer Ausführungen über relative und absolute Normen.

Vergleichen und Urteilen

Die Beurteilung der Charakteristika eines Bildungsprogramms kann erfolgen (1) in bezug auf absolute Normen, wie sie sich in persönlichen Urteilen äußern, und (2) in bezug auf relative Normen, wie sie in den Charakteristika alternativer Curricula zum Ausdruck kommen. Man kann das School Mathematics Study Group Project in bezug auf persönliche Ansichten darüber, was ein Mathematikcurriculum sein soll, oder in bezug auf andere Mathematikcurricula evaluieren. Die Vergleiche und Beurteilungen des Evaluators sind in Abbildung 3 dargestellt. Der obere linke Teil der Abbildung entspricht der Datenmatrix in Abbildung 2. Auf der oberen rechten Seite sind Normengruppen dargestellt, mit denen ein Curriculum absolut beurteilt werden kann. Da es zahlreiche Bezugsgruppen oder Gesichtspunkte geben kann, gibt es viele unterschiedliche Normengruppen. Die verschiedenen Matrizen auf der unteren linken Seite stellen verschiedene alternative Curricula dar, mit denen das Curriculum, das evaluiert wird, verglichen werden kann.

Wenn alle absoluten Normengruppen formalisiert werden, würden sie angemessene und wertvolle Bezugs Ebenen für die Voraussetzungen, Prozesse und Ergebnisse bilden. Bislang ging es nur um das Aufstellen von Normen, nicht um ihre Beurteilung. Bevor der Evaluator ein Urteil fällt, muß er bestimmen, ob alle Normen getroffen werden. Wenn Normen nicht vorhanden sind, müssen sie gesetzt werden. Der Urteilsakt selbst entscheidet, welche Normengruppe berücksichtigt wird. Genauer gesagt, Urteile fällen heißt: jeder Normengruppe eine bestimmte Bedeutung zuordnen. Rationales Urteilen in der pädagogischen Evaluation ist eine Entscheidung darüber, wieviel Beachtung den Normen jeder Bezugsgruppe bei der Entscheidung darüber zukommt, ob eine administrative Handlung erfolgen oder nicht erfolgen soll.

Der relative Vergleich wird ähnlich durchgeführt, wobei allerdings die Normen aus der Beschreibung anderer Bildungsprogramme stammen. Es ist nicht sehr schwierig, ein Urteil darüber zu fällen, ob ein Curriculum in einem Charakteristikum besser ist als ein anderes, aber es gibt viele Cha-

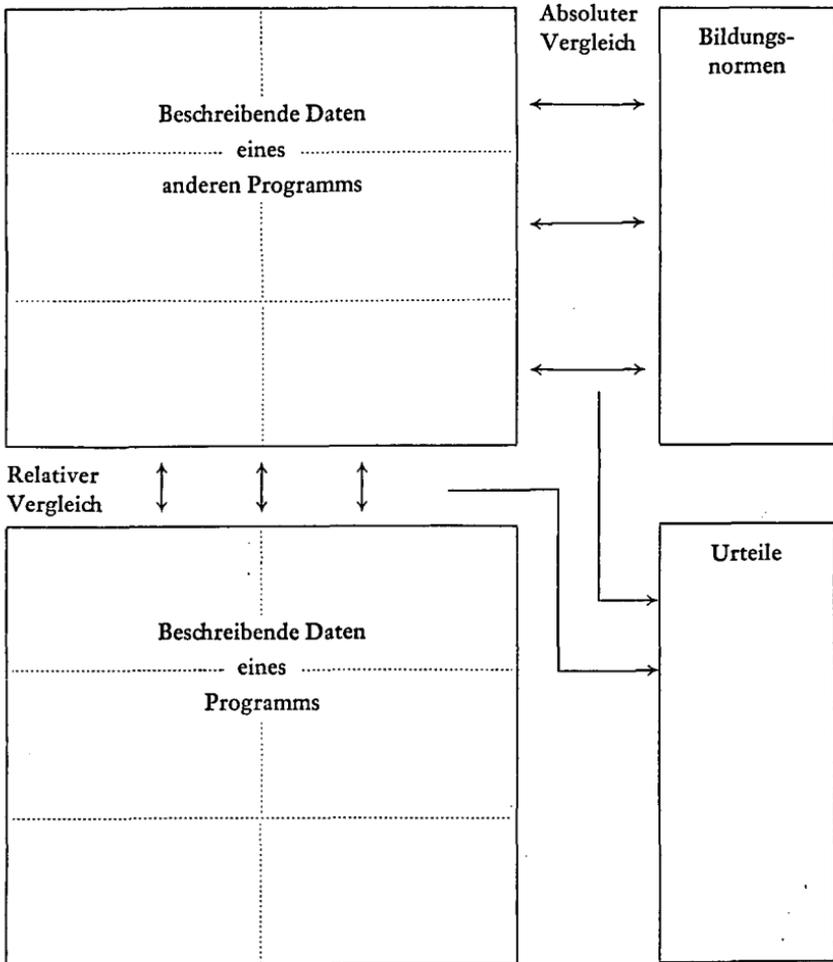


Abb. 3: Eine Darstellung des Prozesses der Beurteilung des Wertes eines Bildungsprogramms

rakteristika, die verschieden wichtig sind. Der Evaluator wählt aus, welche Charakteristika zu berücksichtigen und mit welchen Bildungsprogrammen sie zu vergleichen sind. Mit Hilfe der relativen und der absoluten Beurteilung eines Bildungsprogramms kann man ein Gesamturteil über seine Qualität (vielleicht mit einigen modifizierenden Aussagen) erhalten, das die Grundlage für eine Bildungsentscheidung sein kann. Aufgrund dieser abschließenden Beurteilung kann dann eine Empfehlung ausgesprochen werden.

Absolute und relative Evaluation

Ob absoluter oder relativer Evaluation der Vorzug zu geben ist, ist zwischen Scriven und Cronbach umstritten geblieben. Cronbach (1963) vertritt die Auffassung, es sei sehr gewagt, aus den Ergebnissen curricularer Vergleichsuntersuchungen Generalisationen zu machen, die auf eine örtliche Schulsituation zutreffen sollen – selbst wenn diese Untersuchungen umfangreich, gut angelegt und richtig kontrolliert worden sind –, so daß in Vergleichsuntersuchungen eine unzulängliche Forschungsinvestition besteht. Darüber hinaus ist wahrscheinlich der Unterschied in den Zielsetzungen der verglichenen Curricula so groß, daß die Ergebnisse nicht interpretierbar sind, es sei denn, ein Curriculum ist einem anderen weit überlegen. Da Cronbach diesen Fall selten erwartet, tritt er dafür ein, weniger Vergleichsuntersuchungen, dafür aber mehr intensive Prozeß- und Einzeluntersuchungen von Curricula mit umfangreichen Messungen und sorgfältiger Beschreibung zu machen.

Scriven (1967) andererseits vertritt die Auffassung, daß der Pädagoge vor allem wissen will, ob ein Curriculum besser ist als ein anderes, und daß das beste Verfahren zur Beantwortung dieser Frage der direkte Vergleich ist. Er weist darauf hin, wie schwierig die begrifflich klare Beschreibung der Ergebnisse komplexer Lernprozesse in bezug auf absolute Normen ist, wenn man sie mit der Beobachtung der relativen Ergebnisse zweier Bildungsprogramme vergleicht. Inwieweit Scrivens Ausführungen überzeugend sind, hängt wahrscheinlich vom Adressaten ab. Ein Pädagoge, der über die Adaptation eines Programms entscheiden muß, wird seine Überlegungen eher plausibel finden als ein Curriculuminnovator und Unterrichtstechnologe.

Die von Scriven getroffene Unterscheidung zwischen *formativer* und *summativer* Evaluation ist eine sehr wertvolle Differenzierung im Rahmen der Evaluation. Seine Verwendung der Begriffe bezieht sich vor allem auf das Stadium der Entwicklung von curricularem Material. Solange das Material noch nicht so fertiggestellt ist, daß es an die Lehrer verteilt werden kann, ist Evaluation formativ; nach Abschluß seiner Entwicklung ist Evaluation summativ. Wahrscheinlich empfiehlt es sich eher, zu unterscheiden zwischen einer Evaluation, die sich an den Kriterien und Normen der Curriculumentwickler, Autoren und Verleger orientiert, und einer Evaluation, die sich an den Kriterien und Normen der Schüler, Beamten der Schulverwaltung und Lehrer orientiert. Die Unterscheidung zwischen formativer und summativer Evaluation könnte so definiert werden, und ich will die Begriffe so verwenden. Die Kommission eines Lehrerkollegiums, die ein Curriculum für die Schule auswählen soll, stellt Fragen nach seiner Qua-

lität und Geeignetheit. Der Curriculumentwickler, der Cronbachs Rat befolgt, fragt danach, wie das Curriculum verbessert werden kann. (Keiner befaßt sich mit den individuellen Unterschieden zwischen den Schülern.) Um diese Fragen zu beantworten, muß der Evaluator verschiedene Daten untersuchen und sich auf verschiedene Normen beziehen.

Der Evaluator, der seine Aufgabe eher in der summativen als in der formativen Evaluation sieht, muß die Adressaten über die Qualität des Curriculum informieren. Sein Ziel besteht darin (vgl. Abb. 3), zu Urteilen zu gelangen. Wahrscheinlich wird er die Schulsituationen zu beschreiben versuchen, in denen die Verfahren bzw. Materialien benutzt werden können. Seine Aufgabe kann er darin sehen, herauszufinden, wie gut ein Curriculum in ein bereits bestehendes Schulprogramm paßt. Er muß in Erfahrung bringen, ob die für das Curriculum intendierten Voraussetzungen, Prozesse und Ergebnisse sich mit den finanziellen Mitteln, Normen und Zielen der Schule vereinbaren lassen. Dazu kann es erforderlich sein, der Schule ebensoviel Aufmerksamkeit wie dem Curriculum zuzuwenden.

Der formative Evaluator andererseits interessiert sich stärker für Kontingenzen, wie sie in Abbildung 2 dargestellt worden sind. Er wird in der Evaluationsuntersuchung und in Querschnittsuntersuchungen (across studies) nach gemeinsamen Veränderungen suchen, um auf ihrer Grundlage die Entwicklung gegenwärtiger oder zukünftiger Curricula zu steuern.

Für größere Evaluationsuntersuchungen verfügt ein Evaluator allein natürlich nicht über die vielen benötigten Fähigkeiten und Kenntnisse. Ein Team von Sozialwissenschaftlern ist für viele Aufgaben erforderlich. Solche Teams werden sicherlich aus Experten in der Unterrichtstechnologie, der Psychometrie, in den Skalierungsmethoden, in der Forschungsplanung (research design) und der Dissemination von Informationen bestehen. Curriculare Innovationen haben ohne Zweifel tiefe und umfassende Wirkungen auf unsere Gesellschaft; daher empfiehlt es sich auch, zu manchen Evaluationsteams einen Kulturanthropologen hinzuzuziehen. Auch Wirtschaftswissenschaftler und Philosophen können einen wesentlichen Beitrag zur Evaluation leisten. Experten für die Untersuchung von Wertvorstellungen, für Stichprobenerhebungen und statistische Methoden werden benötigt.

Der Pädagoge, der sich vor der Teilnahme an einer Evaluation scheut, wird erst recht davon unangenehm berührt, daß ein Evaluationsteam in seiner Schule arbeitet. Darüber hinaus erhebt sich für ihn die Frage, wie Evaluatoren die wirkliche Beschaffenheit der Erziehung, die durch ihre Gegenwart beeinflußt wird, beobachten und beschreiben können. Die Bedenken des Pädagogen sind berechtigt; die Evaluation – ja die bloße Gegenwart der Evaluatoren – hat manchmal einen positiven und manchmal einen negativen Einfluß auf die Erziehung. In beiden Fällen trägt sie je-

doch zum atypischen Charakter des Unterrichts bei. Einige Wissenschaftler nehmen an (Webb/Campbell/Schwartz/Sechritz 1966), daß die Verfahren der Evaluation eines Tages soweit entwickelt sein werden, daß sie die Evaluation nicht mehr beeinträchtigen.

Abschließend möchte ich den Leser darauf aufmerksam machen, daß zur Zeit eine der größten Investitionen im Bildungswesen in der Entwicklung neuer Bildungsprogramme besteht. Schulaufsichtsbeamte können ein Curriculum noch nicht mit Hilfe rationaler Begründungen revidieren, da es die dazu notwendige Evaluation nicht gibt. Wie können die Erfahrungen aus den Anstrengungen der Innovatoren der sechziger Jahre genutzt werden, wenn in den sechziger Jahren keine Evaluationsaufzeichnungen vorhanden sind? Für die Innovatoren und Lehrer der siebziger Jahre sind solche Informationen notwendig, denn die bisherigen Kenntnisse können nicht die Sammlung ausreichender Kenntnisse ersetzen. In unseren Datensammlungen sollten wir die Auswirkungen und ihre Gründe, die Kongruenz zwischen Intentionen und Ergebnissen und die verschiedenen Urteile der Adressaten festhalten. Solche Aufzeichnungen sollten gesammelt werden, um das pädagogische Handeln zu fördern, nicht um es zu hemmen. Evaluation sollte ihre Aufgabe darin sehen, Daten für Entscheidungen zu sammeln, und nicht darin, Unruhe in der Schule zu stiften.

Pädagogen sollten ihre eigenen Evaluationen sorgfältiger machen und stärker formalisieren. Alle, die in ihren Klassen oder in überregionalen Kommissionen daran interessiert sind, können sich vielleicht ihre Aufgabe durch die Beantwortung folgender Fragen verdeutlichen:

- (1) Ist diese Evaluation vorwiegend deskriptiv oder vorwiegend beurteilend oder deskriptiv und beurteilend zugleich?
- (2) Soll diese Evaluation die Voraussetzungen, die Prozesse oder die Ergebnisse allein oder eine Kombination dieser oder aber ihre funktionalen Kontingenzen betonen?
- (3) Soll diese Evaluation die Kongruenz zwischen den Intentionen und den Ergebnissen angeben?
- (4) Richtet sich diese Evaluation auf ein Programm allein, oder findet sie als Vergleich zwischen zwei oder mehreren Programmen statt?
- (5) Soll diese Evaluation eher zur Weiterentwicklung von Curricula dienen, oder soll sie zwischen vorhandenen Curricula auszuwählen helfen?

Mit der Beantwortung dieser Fragen werden restriktive Auswirkungen unvollständiger und unangemessener Auffassungen von Evaluation leichter vermieden.

LEE J. CRONBACH: Evaluation zur Verbesserung von Curricula

Übersetzung von Ines Graudenz (Dipl.-Psych.).

Originaltitel: Evaluation for course improvement; zugrunde gelegte Fassung aus: R. W. Heath, *New curricula*, Harper & Row 1964, benutzt und zitiert nach Abdruck in: N. E. Gronlund (Ed.): *Readings in measurement and evaluation*, London: The Macmillan Company 1970. Erste Fassung in: *Teachers College Record* 64, 1963, 672-683.

1 Meine Ausführungen zu diesen Fragen konnten durch die Reaktionen, die ich auf die erste Fassung dieses Beitrags von einigen Leitern von Curriculumprojekten und von Kollegen erhielt, präzisiert werden.

MICHAEL SCRIVEN: Die Methodologie der Evaluation

Übersetzung von Gisela Spöhring und dem Herausgeber

Originaltitel: *The methodology of evaluation*

Erste Fassung: Publication 110 des Social Science Education Consortium, University of Colorado, Boulder 1966.

Zweite Fassung: American Educational Research Association, Monograph Series on Curriculum Evaluation, No. 1, Chicago: Rand McNally 1967.

Dritte Fassung: Sie wurde dem Herausgeber Anfang 1972 zugesandt und liegt der Übersetzung zugrunde. Zitiert wird nach der zweiten Fassung und nur bei Abweichungen in der dritten Fassung nach letzterer. Der Beitrag ist gekürzt.

1* Möbiussche Fläche (nach F. A. Möbius, 1790-1868): einseitige Fläche, veranschaulicht durch z. B. ein Papierband, das um 180° verdreht und zu einem Ring zusammengefügt wird (nach: dtv-Lexikon Bd. 12, München 1966).

ROBERT STAKE: Verschiedene Aspekte pädagogischer Evaluation

Übersetzung vom Herausgeber.

Originaltitel: *The countenance of educational evaluation*, *Teachers College Record* 68, 1967, 7, 523-40.

(Der Übersetzung wurde ein Nachdruck mit den Seiten 1-11 zugrunde gelegt, der auch den Zitaten zugrunde liegt).

1 Hier und an anderen Stellen dieses Beitrags beziehen wir uns, um die Darstellung zu vereinfachen, auf den Evaluator und den Pädagogen als zwei verschiedene Personen. Der Pädagoge ist oft sein eigener Evaluator oder ein Mitglied des Evaluationsteams.

2 Eine solche Liste bilden die *Evaluative Criteria*, die von der National Study of Secondary School Evaluation (1960) veröffentlicht worden sind. Es ist eine sehr sorgfältig erstellte Liste von Voraussetzungen und möglichen Prozessen, die im allgemeinen nach Unterrichtsfächern und -inhalten gegliedert ist. Sie ist als Liste wertvoll, da sie auf vernachlässigte Bereiche aufmerksam macht. Ihr Wert liegt auch darin, zur Verbesserung eines Curriculum, das im Entwicklungsstadium ist, beitragen zu können. Sie kann jedoch nur von begrenztem Wert für die

Evaluation sein. Denn sie hilft weder bei der Datenerhebung noch der Interpretation der Daten. Aufgrund ihrer Zielsetzung enthält sie Kriterien (zur Auswahl der Variablen) und überläßt die Beantwortung der Frage nach Normen (Welche »ratings« sind angemessen?) dem Beobachter.

* Die vierte Auflage der *Evaluative Criteria* erschien 1969; National Study of Secondary School Evaluation, 1785 Massachusetts Avenue, N. W. Washington D. C. 200 366.

DANIEL L. STUFFLEBEAM: Evaluation als Entscheidungshilfe

Übersetzung von Gudrun Eggert und dem Herausgeber.

Originaltitel: *Evaluation as enlightenment for decision making*, in: *Improving educational assessment and an inventory of measures of affective behavior*, hg. W. H. Beatty, Association for Supervision and Curriculum Development, National Education Association, Washington D. C. 1969, 41-73.

1 Public Law 89-10: The Elementary and Secondary Education Act of 1965, Titel I.

2 Diese Kriterien sind auf den Seiten 70-71 der Titel - III - Richtlinien aufgeführt: *A manual for project applicants and grantees*, U.S. Office of Education, Washington D. C. 1967.

3 Egon G. Guba: *Evaluation and the process of change*, Notizen und Arbeitspapiere bezüglich der Administration von Programmen des Titel III des Public Law 89-10, The Elementary and Secondary Education Act of 1965, erweitert durch Public Law 89-750, April 1967, 312.

4 a. a. O.

5 Citizen Committee for Children of New York, Inc., Newsletter, Bericht von Mrs. Nathan W. Levin, Vorsitzende der Educational Services Section, vor dem Subcommittee on the Elementary and Secondary Education Act of the Education and Labor Committee of the House of Representatives, März 1967.

6 Egon G. Guba: *Methodological Strategies for educational change*, ein Referat, das auf der Tagung über Strategien zur pädagogischen Reform gehalten wurde, die vom 8.-10. November 1965 in Washington D. C. stattfand.

7 Das EPIE-Forum. Monatliche Veröffentlichung des Instituts für Informationsaustausch über Erziehungsprodukte, das von und für Pädagogen geschaffen wurde. New York: Educational Products Information Exchange Institute.

MARVIN C. ALKIN:

Die Aufwands-Effektivitäts-Evaluation von Unterrichtsprogrammen

Übersetzung von Manfred Weiß (Dipl.-Kfm.) und Hans v. Schaper (cand. rer. pol.).

Originaltitel: *Evaluating the cost-effectiveness of instructional programs* (Referat, gehalten auf dem »Symposium on Problems in the Evaluation of Instruction, University of California, Los Angeles, Dezember 1967), Report No. 25, Los Angeles: University of California, Center for the Study of Evaluation of Instructional Programs 1969.