

Scriven, Michael

Die Methodologie der Evaluation

Wulf, Christoph [Hrsg.]: *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München : R. Piper & Co. Verlag 1972, S. 60-91. - (*Erziehung in Wissenschaft und Praxis*; 18)



Quellenangabe/ Reference:

Scriven, Michael: Die Methodologie der Evaluation - In: Wulf, Christoph [Hrsg.]: *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München : R. Piper & Co. Verlag 1972, S. 60-91 - URN: urn:nbn:de:0111-opus-14235 - DOI: 10.25656/01:1423

<https://nbn-resolving.org/urn:nbn:de:0111-opus-14235>

<https://doi.org/10.25656/01:1423>

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, aufführen, vertrieben 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

MICHAEL SCRIVEN

Die Methodologie der Evaluation

Einführung

Die bisherigen Konzeptionen der Evaluation sind in Theorie und Praxis noch unzureichend. Mit diesem Beitrag soll versucht werden, einige Unzulänglichkeiten aufzudecken und zu verringern. Geistiger Fortschritt ist nur möglich, weil junge Wissenschaftler auf den Arbeiten von Koryphäen aufbauen können. Dazu gehört jedoch auch, daß beide Seiten die Leistungen der anderen Seite respektieren. Verpflichtet bin ich Lee Cronbachs Aufsatz von 1963, weiterführenden Gesprächen mit den Mitarbeitern des Center for Instructional Research and Curriculum Evaluation (CIRCE) an der Universität von Illinois, Urbana, und anregendem Schriftwechsel mit mehreren Kollegen, insbesondere James Shaver und Ray Barglow.

Überblick

Der Schwerpunkt dieses Beitrags liegt auf der Curriculumevaluation; fast alle Überlegungen können jedoch ohne weiteres auf andere Arten der Evaluation übertragen werden. Die Überschriften der einzelnen Abschnitte sind aus sich selbst heraus verständlich und erscheinen in folgender Reihenfolge:

1. Überblick
2. Ziele und Rollen der Evaluation; die formative und summative Rolle der Evaluation
3. Professionelle und Amateur-Evaluation
4. Evaluationsuntersuchungen und Prozeßuntersuchungen
5. Evaluation und Überprüfung der Zielerreichung
6. Intrinsische Evaluation und Ergebnissevaluation
7. Praktische Vorschläge für eine Mischform der Evaluation (Hybrid Evaluation)
8. Das Für und Wider einer reinen Ergebnissevaluation
9. Vergleichende und nicht-vergleichende Evaluation
10. Praktische Verfahrensweisen für die Evaluation mit Kontrollgruppen.

Ziele der Evaluation und Rollen der Evaluation; die formative und die summative Rolle der Evaluation

Die Funktion der Evaluation kann von zwei Seiten her begriffen werden. Auf der methodischen Ebene können wir von den *Zielen* der Evaluation sprechen; darüber hinaus lassen sich in einem bestimmten soziologischen oder pädagogischen Kontext verschiedene *Rollen* der Evaluation unterscheiden.

In bezug auf die Ziele der Evaluation kann man davon ausgehen, daß Evaluation bestimmte *Fragen* über bestimmte *Einheiten* zu beantworten versucht. Die Einheiten sind die verschiedenen pädagogischen Instrumente (Prozesse, Personal, Verfahrensweisen, Programme usw.) Zu den Fragen gehören solche über die Form: *Wie gut* funktioniert dieses Instrument (in bezug auf diese oder jene Kriterien)? Funktioniert es *besser* als ein anderes Instrument? *Was* leistet dieses Instrument, d. h. welche Variablen der uns interessierenden Gruppe werden von seiner Anwendung signifikant beeinflußt? *Rechtfertigt* der Gebrauch dieses Instruments seine Kosten? Evaluation an sich ist ein methodisches Vorgehen, das im Grunde genommen gleich ist, unabhängig davon, ob man Kaffeemaschinen, Lehrmaschinen, Pläne für ein Haus oder ein Curriculum zu evaluieren versucht. Es besteht einfach im Sammeln und Kombinieren von Verhaltensdaten mit einem gewichteten Satz von Skalen, mit denen entweder vergleichende oder numerische Beurteilungen erlangt werden sollen, und in der Rechtfertigung (a) der Datensammlungsinstrumente, (b) der Gewichtungen, (c) der Kriterienauswahl.

In einem bestimmten pädagogischen Kontext kann die *Rolle* der Evaluation jedoch sehr unterschiedlich sein. Evaluation kann Bestandteil der Lehrerbildung, der Curriculumentwicklung, eines Feldversuchs zur Verbesserung einer Lerntheorie oder einer Untersuchung von Unterrichtsmaterial vor einer Kaufentscheidung sein. Evaluation kann ferner zu einer Datensammlung führen, mit deren Hilfe eine Steuererhöhung oder ein Forschungsvorhaben unterstützt wird; sie kann aber auch z. B. in einem Trainingsprogramm, in einem Gefängnis oder in einer Schulklasse als Mittel zu einer positiven oder negativen Rückmeldung (feedback) dienen. Häufig hat man versäumt, diese wichtige Unterscheidung zwischen Rollen und Zielen der Evaluation zu treffen. Das hat u. a. zu einem starken Substanzverlust in der Evaluation geführt, so daß sie nicht länger zur Beantwortung von Fragen über den Wert von Bildungsprogrammen beiträgt, obwohl gerade darin ihre Aufgabe liegen sollte. Man darf Evaluation nur ablehnen, wenn man begründen kann, daß man sich nicht um eine Antwort auf Fragen nach dem Wert pädagogischer Instrumente bemühen sollte. Dazu wür-

de jedoch auch der kaum zu erbringende Nachweis gehören, daß diese Fragen bei überhaupt *keinen* Aktivitäten gestellt werden dürfen. Aus der Tatsache, daß der Evaluation manchmal eine unangemessene Rolle zugewiesen wird, darf man nicht schließen, daß man Fragen nach dem Ziel der Evaluation niemals beantworten sollte. Die besonders bei Lehrern oder Schülern häufig anzutreffende Furcht vor Evaluation ist eine oft zu Unrecht verallgemeinerte Reaktion auf die berechtigte Ablehnung einer Situation, in der Evaluation eine Rolle zugewiesen wurde, die nicht mehr in ihren Geltungsbereich fällt und der sie auch nicht gerecht werden kann.

Zu Recht verweist man häufig auf die Rolle der Evaluation im Prozeß der *Curriculumentwicklung*, die natürlich die Evaluation des *Endergebnisses* dieses Prozesses nicht ersetzen kann. In der Regel kann und muß Evaluation verschiedenen Rollen gerecht werden. Aus der Behandlung der Evaluation in einer Reihe neuerer Veröffentlichungen und Forschungsprojekte läßt sich jedoch erkennen, daß man die Anforderungen an Evaluation bereits zu erfüllen meint, wenn man an *irgendeiner* Stelle im Laufe des Projekts evaluiert. Evaluation kann jedoch im Rahmen eines pädagogischen Projekts nicht nur verschiedene Rollen übernehmen; sie kann auch innerhalb jeder Rolle mehrere spezifische Ziele haben. So kann Evaluation bei der Verbesserung des Curriculum eine Rolle spielen. Um ihr gerecht zu werden, kann man z. B. folgende Fragen stellen: Gelingt es dem Curriculum wirklich, den Unterschied zwischen Vorurteil und politischer Einstellung zu vermitteln? Benötigt es dafür zuviel Unterrichtszeit? Diese Form der Evaluation kann man als *formative* Evaluation bezeichnen. Ihre Ergebnisse bleiben innerhalb der die Curriculumentwicklung tragenden Institution und dienen zur Verbesserung des Programms.

Der Evaluation kommt eine andere Rolle zu, wenn sie den Beamten der Schulverwaltung bei der Entscheidung helfen soll, ob das fertiggestellte Curriculum den anderen Alternativen so weit überlegen ist, daß sich die Ausgaben für seine Einführung in das Schulsystem rechtfertigen lassen. Man kann diese Form der Evaluation als *summative* Evaluation bezeichnen. Ihre Ergebnisse werden Interessenten außerhalb der die Curriculum tragenden Institution zur Verfügung gestellt und dienen zum besseren *Verständnis* und zur besseren *Verwendung* des Programms.

Die häufige Verwechslung und fehlende Unterscheidung zwischen Rollen und Zielen der Evaluation liegt zum Teil in dem wohlgemeinten Versuch begründet, die Beunruhigung der Lehrer durch Evaluation zu verringern. Indem man jedoch die konstruktive Funktion der Evaluation so stark betont, übergeht man stillschweigend, daß zu den Zielen der Evaluation auch die Beurteilung von Leistung, Bedeutung, Wert usw. gehört, die in einer anderen Rolle der Evaluation zu Entscheidungen über die

Förderung von Personen und Curricula beitragen kann. Man sollte die Furcht vor Evaluation nicht dadurch zu verringern suchen, daß man diese wichtige Funktion der Evaluation nicht genügend berücksichtigt und die Darstellung der Evaluationsergebnisse verfälscht. Die nachteiligen Folgen für das Erziehungswesen sind zu groß. Die Wirtschaft kann es sich auch nicht erlauben, Fabriken zu unterhalten oder leitende Angestellte zu beschäftigen, die ganz offensichtlich keine gute Arbeit leisten. Ebenso sollte die Gesellschaft – solange gute Leistungen erbracht werden können – keine schlechten Bücher und Kurse verwenden und unzulängliche Lehrer und Schulaufsichtsbeamte beschäftigen. Um der Furcht vor Evaluation entsprechend zu begegnen, muß man für die Personen, deren Position oder Ansehen in Gefahr ist, Aufgaben schaffen, für die sie besser geeignet sind. Wenn man die Schülerleistung nicht beurteilt, führt das zu unzulänglichen Leistungen in der Schulklasse. Scheut man sich, die Lehrerleistung zu evaluieren, bewirkt das inkompetenten Unterricht. Daher dürfen wir, wenn wir der gesellschaftlichen Verantwortung für das Erziehungswesen gerecht werden wollen, uns gegenüber dem einzelnen nicht immer so zaghaft verhalten. Vielleicht trifft es zu, daß »der größte Beitrag, den Evaluation leisten kann, darin liegt, die Aspekte des Curriculum herauszuarbeiten, für die eine Neubearbeitung erforderlich ist« (Cronbach 1963, 41, 46). Jedoch gibt es ohne Zweifel auch im Rahmen der meisten Curriculumprojekte und Innovationen gleich wichtige andere Funktionen der Evaluation. Es gibt z. B. viele Situationen, in denen man durch die abschließende Evaluation eines Projekts oder einer Person seiner Verantwortung gegenüber der Person, dem Projekt oder dem Steuerzahler nachkommt. Wenn man mit Hilfe einer umfassenden Ergebnisevaluation deutlich machen kann, daß ein teures Schulbuch nicht besser als ein anderes Schulbuch ist, mit dem es verglichen wurde, oder aber daß das teure Schulbuch bei weitem besser als alle anderen ist, so sollte man doch nicht wie Cronbach diesen Beitrag der Evaluation für unbedeutend erklären. Das heißt: Wenn man z. B. zeigen kann, daß ein bestimmtes Verfahren, Mathematik zu unterrichten, in keiner von Mathematikern für wichtig gehaltenen Dimension signifikant bessere Schülerleistungen erzielt, dann kann dieses Ergebnis Zeit und Geld sparen helfen und somit einen Beitrag zur Entwicklung des Bildungswesens leisten. Das gleiche gilt natürlich auch, wenn mit Hilfe des Verfahrens signifikant bessere Leistungen erzielt werden können. Demnach müssen zunächst einige erhebliche Einschränkungen erfolgen, bevor man Cronbach zustimmen und der formativen Evaluation größere Bedeutung zuschreiben kann als der summativen: »Evaluation, die auf die Verbesserung von noch in der Entwicklung befindlichen Curricula zielt, trägt mehr zur Verbesserung des Unterrichts bei als die Evaluation, die nur dazu dient, Produkte

zu bewerten, die bereits auf dem Markt sind« (Cronbach 1963, 41 46). Das beste Gegenbeispiel bilden die erfolgreichen, jedoch rassistischen Grundschulbücher der späten fünfziger Jahre. Es bedurfte einer summativen Evaluation mit negativem Ergebnis, um sie aus den Schulen zu verdrängen und durch andere bessere Bücher zu ersetzen. Doch zum Glück braucht man sich nicht für eine der beiden Rollen der Evaluation zu entscheiden. Bei pädagogischen, besonders jedoch bei curricularen Projekten muß man versuchen, beide Rollen der Evaluation zu erfüllen.

Wahrscheinlich hat jeder Curriculumentwickler seine Aufgabe übernommen, weil er das gegenwärtige Curriculum aufgrund vorläufiger summativer Evaluation für unzureichend hält. Während er das neue Curriculummaterial entwickelt, evaluiert er es laufend, indem er es besser als das bereits vorhandene Material zu machen versucht. Wenn er sich auch nur ein wenig der Begrenztheit seines Urteils über die eigene Arbeit bewußt ist, wird er das Curriculum noch während seiner Entwicklung in der Schule testen. Dadurch erhält der Curriculumentwickler eine Rückmeldung, auf deren Basis er es revidieren kann. Dieses Vorgehen ist eine formative Evaluation; wenn diese Felduntersuchung gut ausgeführt wird, kann sie sogar zu einer summativen Evaluation der *frühen Formen* des neuen Curriculum werden. Im allgemeinen arbeitet der Curriculumentwickler mit Lehrern oder anderen Kollegen zusammen, die das Material fortwährend kommentieren und beurteilen. Auch das ist eine Form der Evaluation, mit deren Hilfe das Curriculum verbessert werden kann.

Wenn formative Evaluation sinnvoll durchgeführt werden soll, dann sollte möglichst ein *professioneller Evaluator* in dem Curriculumprojekt mitarbeiten. Im allgemeinen werden dabei die Vorteile überwiegen; dennoch deuten einige Erfahrungen in der Praxis darauf hin, daß die Mitarbeit eines professionellen Evaluators auch Nachteile haben kann. Diese Frage berührt natürlich nicht die Frage nach der summativen Evaluation und der Rolle des professionellen Evaluators in ihr. Beide Fragen sollen in einem Teil des nächsten Abschnitts weiter erörtert werden.

Professionelle und Amateur-Evaluation

Der Evaluator ist zwar in seinem Gebiet ein Fachmann, selten aber in dem für das Curriculum inhaltlich relevanten Bereich. Im Unterschied zu den Curriculumentwicklern steht er dem Projekt im allgemeinen distanzierter gegenüber. Diese unterschiedliche Einstellung führt nicht selten zu Spannungen und Zwistigkeiten, die jedem Projektleiter nur zu vertraut sind.

Die unzulängliche Kommunikation zwischen Evaluatoren, Lehrern und

Curriculumentwicklern hat leider zu zwei extremen Reaktionen geführt. Es entstand einmal eine starre Anti-Evaluations-Haltung; sie ist oft nur eine Rationalisierung der Furcht, die durch die Gegenwart eines externen Beurteilers ausgelöst wird, der sich mit den Zielen des Projekts nicht identifiziert hat und der ihnen auch nicht verpflichtet ist. Das andere ebenso unerfreuliche Extrem besteht in dem rigiden Evaluator, der nur Operationalisierungen gelten läßt und der deshalb oft dem Sinn nach sagen dürfte: »Wenn Sie mir nicht in operationalisierter Form sagen, welche Variablen Sie beeinflussen wollen, kann ich keinen entsprechenden Test entwickeln, und solange die Variablen nicht getestet worden sind, dürfen Sie nicht annehmen, daß Sie die Variablen erfolgreich beeinflußt haben.«

Zur Präzisierung dieser beiden Positionen wollen wir den Unterschied zwischen einem großen Curriculumprojekt der Gegenwart und einem in den späten dreißiger Jahren von zwei oder drei Lehrern gemeinsam verfaßten Algebra-Text deutlich herausarbeiten.

Erstens werden die heutigen Projekte häufig mit umfangreichen öffentlichen Geldern finanziert. Um diese Ausgaben zu rechtfertigen, muß ein objektiver Nachweis über den Wert des Produkts erbracht werden. Sodann ist für die *weitere* Finanzierung der Arbeit in diesem Bereich oder anderer Projekte derselben Curriculumentwickler ein objektiver Leistungsnachweis erforderlich. Da die Finanzen nicht ausreichen, alle Antragsteller zu unterstützen, muß man den Wert der Projekte aufgrund eines Vergleichs beurteilen. Objektive Grundlagen dafür sind natürlich den persönlichen Meinungsäußerungen von Kollegen überlegen. Schließlich werden durch die hohen Kosten für die *Einführung* solcher Curriculummaterialein in ein Schulsystem weitere Steuergelder verbraucht; diese Ausgaben sollten nur dann erfolgen, wenn sie sich auf Grund ausreichender Daten rechtfertigen lassen. Daher müssen Projektleiter, finanzierende Institutionen und Schulen auf die Durchführung einer summativen Evaluation drängen. Da formative Evaluation zu jedem rationalen Versuch gehört, gute Ergebnisse in der summativen Evaluation zu erzielen, muß sie auch von Anfang an erfolgen; nach unserem Verständnis ist sie sogar bis zu einem gewissen Grad durch den Prozeß der Curriculumentwicklung selbst bedingt. Davon unabhängig ist die Frage, ob und wie man professionelle Evaluatoren zum Projekt hinzuziehen soll. Die Beantwortung der Frage hängt davon ab, inwieweit formative Evaluation zur Verbesserung des Curriculum beitragen, bzw. seine Entwicklung behindern kann; und es gibt durchaus eine Reihe von Situationen, in denen sie den Prozeß der Curriculumentwicklung eher negativ beeinflußt.

Professionelle Evaluatoren können z. B. so kritisch sein, daß sie die Arbeit einer produktiven Gruppe ernsthaft gefährden. Auch wenn sie der

Gruppe, im ganzen gesehen, in der Regel behilflich sind, stellen sie z. B. oft so hohe Anforderungen an die operationale Formulierung der Ziele, daß zuviel Zeit für eine im Grunde sekundäre Tätigkeit verwendet wird. Daher muß ein Kompromiß geschlossen werden. Der Evaluator muß einen Teil *seiner* Aufgabe darin sehen, einen Satz von testbaren Kriterien für das Curriculum zu entwickeln. Dabei kann ihm die Tatsache helfen, daß sich das Projektteam für bestimmte Ziele ausdrücklich entschieden und andere verworfen hat. Für die Formulierung der Kriterien wird ihm außerdem die Kritik des Teams nützlich sein. Die Kommunikation zwischen Evaluator und Curriculumentwickler muß jedoch in beiden Richtungen erfolgen. Den Curriculumentwicklern unterlaufen häufig Fehler; sie sind oft voreingenommen oder zu sehr von ihrem Projekt begeistert. Evaluatoren wiederum haben, solange sie noch nicht mit den Themen und Zielen der Curriculumentwickler vertraut sind, nur begrenzte Wirkungsmöglichkeiten. Wenn sie sich aber mit diesen Zielen und dem Gesamtprojekt identifizieren, verlieren sie leicht die für eine objektive Evaluation wichtige Unabhängigkeit. Daher sollte man die formativen Evaluatoren nach Möglichkeit auch deutlich von den summativen Evaluatoren unterscheiden, denen sie natürlich bei der Entwicklung eines summativen Evaluationsplans behilflich sein können. Wenn man zwischen formativen und summativen Evaluatoren unterscheidet, kann man die Vorteile objektiver professioneller Evaluation wahrnehmen, ohne eine Störung der Kooperation im Team zu riskieren.

Bei der Evaluation im Bildungswesen und der Verwendung eines Evaluators im Prozeß der Curriculumentwicklung ergeben sich noch viele weitere Probleme. Auf mehrere hat J. Myron Atkin (1963) hingewiesen. Einige von ihnen sollen später in diesem Beitrag behandelt werden; auf zwei Probleme sei jedoch bereits hier aufmerksam gemacht. Eines besteht darin, daß das Testen bestimmter differenzierter Begriffe dadurch eine negative Auswirkung haben kann, daß es dem Schüler die Rolle eines Begriffs zu früh bewußt werden läßt und dadurch die natürliche Entwicklung des Begriffsverständnisses verhindert. Das zweite Problem besteht darin, daß manchmal bei einigen Begriffen die Verständnisfähigkeit eines Kindes im Verlauf der Arbeit mit einem Curriculum oder während eines bestimmten Schuljahrs nur wenig zunimmt. Dennoch kann dieser geringe Zuwachs für die langfristige Entwicklung der Verständnisfähigkeit von großer Bedeutung sein. Der Verständniszuwachs würde sich jedoch in Tests nicht niederschlagen und könnte sogar durch die Verwendung von Tests beeinträchtigt werden; dennoch muß er potentiell im Curriculum enthalten sein, um das gewünschte Endprodukt hervorzubringen. In einem solchen Fall wäre eine Evaluation jedoch unergiebig und vielleicht sogar hinderlich.

Wenn ein Curriculumteam seine Arbeit mit den Lehrern des gegenwärtigen Curriculum diskutiert, bringt das trotz möglicher Vorteile auch Nachteile mit sich. Das gilt auch für eine frühzeitige Hinzuziehung eines Evaluators. Ein einfallsreicher Projektleiter kann eine solche Situation mit verschiedenen Möglichkeiten zum Besten des Projekts nutzen. Er kann z. B. dem Evaluator lediglich die Curriculummateriale geben, ohne daß dieser die Curriculumentwickler selber kennenlernt. Die Kommentare des Evaluators werden dann dem Projektleiter zugeleitet, der vielleicht zunächst nur die grundsätzliche und ernsthafte Kritik an das Team weitergibt und die anderen kritischen Anmerkungen bis zu dem Zeitpunkt zurückhält, an dem eine umfassende Revision erfolgen soll. Das sind jedoch Überlegungen für die Praxis. Es bleiben zwei grundsätzliche Einwände, die kurz erwähnt werden sollen und von denen der erste sich unmittelbar auf Atkins Befürchtungen bezieht.

Jeder, der ein neues Curriculum in der Schule getestet hat, weiß, daß dieses Curriculum sehr unterschiedliche Wirkungen auf die Schüler haben kann, die sich häufig aus ihren vorherigen Leistungen nicht voraussagen lassen. So findet z. B. ein Kind, das sich bereits für die Beobachtung von Vögeln interessiert, einen entsprechenden Zugang zur Biologie vielleicht attraktiver als einen anderen. Bei einigen Kindern hängt ihr Interesse davon ab, wie weit das Unterrichtsmaterial für die ihnen bereits vertrauten Probleme relevant ist; für andere hingegen sind z. B. die Eigenschaften der Möbiusschen Fläche¹ sofort faszinierend. Im allgemeinen kann die Unterrichtsorganisation die Motivation der Schüler in unterschiedlicher Weise beeinflussen. Der nicht-direktive Erziehungsstil, der wegen seines vermuteten Zusammenhangs mit der induktiven Unterrichtsmethode häufig bevorzugt wird, ist für diejenigen Kinder nicht geeignet, die stärker durch eine aggressive, rivalitätsgeladene, kritische Interaktion zur Aktivität herausgefordert werden. Trotz dieser Unterschiede greift man noch immer auf Tests für die ganze Klasse als undifferenzierte Evaluationsinstrumente zurück. Doch selbst wenn man die Testergebnisse in bezug auf individuellen Leistungszuwachs aufschlüsselt, hat man die Möglichkeiten des Materials noch nicht voll ausgenutzt. Sie würden sich erst dann zeigen, wenn man das richtige Material *und* die richtige Unterrichtstechnik für jedes Kind mit seinen entsprechenden Einstellungen, Interessen und Fähigkeiten auswählt. Wenn jemand der Evaluation skeptisch gegenübersteht, wird er vielleicht vorschlagen, man solle seine Hoffnung auf den kreativen und wissenschaftlich kompetenten Curriculumentwickler setzen und die Felduntersuchungen nur als Nachweis dafür ansehen, daß man Schüler unter geeigneten Bedingungen für das Curriculummateriale interessieren und mit ihm unterrichten könne. Das bedeutet, unser Kriterium sollte der deutliche

Leistungszuwachs bei *einigen* oder *mehreren* Schülern, nicht aber der Leistungszuwachs der ganzen Klasse sein. Dem muß der Evaluator mit dem Einwand begegnen, man dürfe nicht übersehen, daß z. B. ein unzulängliches Verständnis vieler Schüler und eine deutliche relative Verschlechterung der Leistungen mehrerer Schüler den Leistungszuwachs bei einigen Schülern aufhebt. Auch dürfe man nicht ausschließen, daß das pädagogische Geschick oder das Engagement des Lehrers und nicht die Materialien für den Erfolg bei der Felduntersuchung verantwortlich sind. Die Curriculummaterialien müssen daher auch anderen Lehrern gegeben werden, damit sie feststellen können, ob sie ihrer Meinung nach brauchbar sind. Um diese Fragen beantworten zu können, benötigt man professionelle Evaluation.

Aus dieser Kritik läßt sich jedoch eine wichtige Anregung ableiten. Auf jeden Fall muß man nämlich die Ansichten und Urteile der Fachwissenschaftler über die Qualität von Curriculuminhalten gewissenhaft berücksichtigen. Manchmal wird man zwar kaum Informationen für ihre Weiterentwicklung erhalten können; in einigen Fällen werden sie jedoch für bestimmte Entscheidungen durchaus genügen. Auf jeden Fall sollten diese Urteile sorgfältig bedacht und ausdrücklich in der Evaluation berücksichtigt werden; denn eine *fehlende* Unterstützung durch das Urteil von Fachwissenschaftlern ist oft bereits ein ausreichender Grund für eine vollständige Ablehnung des Curriculummaterials.

Schließlich trifft man in vielen Diskussionen auf die Ansicht, daß Evaluation Werturteile erfordert und daß diese Werturteile im Grunde genommen subjektiv und nicht wissenschaftlich sind. Dies ist genau so abwegig wie die Ansicht, daß Aussagen einer Person über sich selbst im Grunde genommen subjektiv sind und damit nicht rational vorgebracht werden können. Einige Werturteile sind im wesentlichen Äußerungen von wichtigen persönlichen Präferenzen (Geschmacksfragen) und als solche faktische Aussagen, die durch die üblichen Verfahren der psychologischen Forschung nachgewiesen werden können. Der Nachweis solcher Urteile gibt keine Auskunft darüber, ob es für jemanden richtig oder falsch ist, solche Wertvorstellungen zu haben. Man erfährt lediglich, ob jemand diese Wertvorstellungen hat oder nicht. Eine andere Form eines Werturteils ist die Einschätzung (assessment) der Leistung oder relativen Leistung einer curricularen Einheit in einem klar definierten Kontext, die schließlich zu dem Urteil führt, die Leistung dieser Einheit sei in bezug auf deutlich identifizierbare und deutlich gewichtete Kriteriumsvariablen so gut wie oder gar besser als die einer anderen curricularen Einheit. Bei Werturteilen dieser Art läßt sich allerdings nicht nur feststellen, ob die Personen, die sie abgeben, zu ihnen stehen oder nicht, sondern man kann auch feststellen, ob es richtig oder falsch ist, diese Werturteile zu haben. Sie sind lediglich komplexe

Gesamturteile aufgrund der Einstufung und Gewichtung verschiedener Leistungen. In diesem Sinne können wir also genau feststellen, daß die Palek Quartz zur Zeit die beste Armbanduhr ist oder daß ein bestimmtes Wörterbuch für Benutzer mit umfangreichen wissenschaftlichen Interessen am besten geeignet ist. Schließlich gibt es noch Werturteile, bei denen die Kriterien selbst umstritten sind; diese Werturteile sind, philosophisch gesehen, die wichtigsten; ihre Strittigkeit verweist darauf, daß wichtige Probleme meist nur schwer lösbar sind. Ein Beispiel für ein solches Urteil ist die Behauptung, daß die wichtigste Rolle der Evaluation im Prozeß der Curriculumentwicklung liegt, daß der Intelligenztest ein überholtes Untersuchungsinstrument ist oder daß die Kopenhagener Interpretation der Quantenphysik allen bekannten Alternativen überlegen ist.

In allen diesen Fällen streitet man sich darüber, was als gut gelten soll, und argumentiert dabei kaum mit den »wirklichen Fakten« der Situation. Dennoch darf man solche Urteile nicht außer acht lassen. Vielmehr sollte man auf jeden Fall die Gründe untersuchen, die für solche Urteile angeführt werden, und dann erst entscheiden, ob und wie diese Fragen rational diskutiert werden können. Nach einer häufig vertretenen Auffassung müssen wir im Umgang mit Menschen, also etwa bei der Erziehung, *ethische* Werturteile fällen, die im Grunde genommen subjektiv sind. Aber erstens sind Werturteile über Menschen keineswegs notwendigerweise ethisch, weil sie sich auch auf ihre Gesundheit, ihre Intelligenz oder ihre Leistungen beziehen können. Zweitens, selbst wenn sie ethisch sind, sind wir alle wohl einem ethischen Prinzip, der Rechtsgleichheit für alle Menschen, verpflichtet. Auf dieser Voraussetzung und einem entsprechenden Bezugsrahmen beruht der größte Teil der öffentlichen Auseinandersetzungen über ethische Fragen. Wenn man nicht dieses Axiom in Frage stellen und rationale Argumente für eine Alternative beibringen will, sind selbst ethische Werturteile rationaler Diskussion zugänglich. Was auch immer das Ergebnis einer solchen Diskussion ist, die Tatsache, daß Evaluation manchmal eine ethische Evaluation ist und daß ethische Evaluation zum Teil kontrovers ist, läßt bei weitem nicht den Schluß zu, daß Curriculumevaluation nicht ein wichtiger Bereich der angewandten Wissenschaft ist, zu der man sonst auch die Ingenieurwissenschaften und die Medizin nicht zählen dürfte.

Evaluationsuntersuchungen und Prozeßuntersuchungen

Wenn man den Begriff Evaluation zu erklären versucht, sollte man sich vor Simplifizierung in acht nehmen. Obwohl Evaluation im allgemeinen auf

Urteile über Leistung und Wert zielt, ist eine analytische Beschreibung und Interpretation des Prozesses notwendig, in dem jemand eine *Situation* oder die *Auswirkung* bestimmter Materialien evaluiert. In diesem Sinne kann man einige Formen von Prozeßforschung als Evaluation begreifen. Prozeßforschung überschneidet sich jedoch mit Evaluation nur teilweise und sollte nicht unter Evaluation subsumiert werden. Nach Cronbach lassen sich drei Arten der Prozeßforschung unterscheiden.

(1) Die erste Art der Prozeßforschung besteht in der deskriptiven Untersuchung der wirklichen Unterrichtsgeschehnisse. Vielleicht kann man sie am ehesten als eine Untersuchung des Lehr- und Lernprozesses kennzeichnen. Man kann z. B. erforschen, wie lange der Lehrer in einer Unterrichtsstunde spricht, wieviel Zeit die Schüler pro Unterrichtsstunde für Hausarbeiten aufwenden oder wieviel Gesprächszeit z. B. für Erklärungen, Definitionen und ähnliches benötigt wird (Meux/Smith 1961). Bei einigen dieser Untersuchungen wird man nur schwer ihren Wert einsichtig machen können. Denn nicht selten sind sie nur Forschungen um ihrer selbst willen. Deshalb muß auch die Arbeit von Smith und Meux besonders erwähnt werden, da sie wirklich originell und sehr erfolgsversprechend ist. Dennoch darf man im ganzen davon ausgehen, daß der größte Teil dieser Art der Prozeßforschung in der Erziehung und der Psychotherapie weder für die Theorie noch für die Praxis fruchtbar ist.

(2) Die zweite Art von Prozeßforschung zielt auf die Erforschung der kausalen Beziehungen zwischen den Prozeßelementen («dynamische Hypothesen»). Hier will man z. B. erforschen, ob ein größerer Zeitaufwand für eine an den curricularen Zielen orientierte Diskussion, die auf Kosten der Zeit für Übungsaufgaben geführt wird, zu einem besseren Verständnis für Algebra oder Geographie führt.

In einer anderen Variante dieser Art der Prozeßforschung versucht man Fragen zu beantworten wie: Wird durch die Betonung des Lehrer-Schüler-Dialogs die Bildung von Untergruppen und die Identifikation mit dem Lehrer gefördert? Diese Untergruppe von Prozeßhypothesen unterscheidet sich von Evaluationshypothesen dadurch, daß die unabhängigen Variablen entweder gar nicht unter den Kriterien einer summativen Evaluationsuntersuchung auftauchen oder daß sie nur eine Untergruppe der summativen Kriterien bilden würden. In beiden Fällen versucht man jedoch nicht, aufgrund von Korrelationsuntersuchungen die Vorzüge zu bestimmen.

Prozeßhypothesen dieser zweiten Art sind im allgemeinen genauso schwierig zu konkretisieren wie Hypothesen über Ergebnisse. Tatsächlich lassen sie sich manchmal sogar noch schwerer konkretisieren, weil sie vielleicht nur die Messung einer einzigen unter mehreren unabhängigen Va-

riablen erfordern und die gebräuchlichen Verfahren der Parallelisierung sich nur schwer dazu verwenden lassen, die anderen Variablen zu kontrollieren. Einige summative Evaluationsuntersuchungen haben den Vorteil, daß sie sich nur mit der Evaluation der Gesamtauswirkungen eines von Lehrern unterrichteten Curriculum befassen und daher nicht die spezifischen Elemente aufdecken müssen, die für die Verbesserung oder Verschlechterung der Ergebnisse verantwortlich sind. Dieser Vorteil wird jedoch hinfällig, wenn wir herausfinden wollen, welche nur Auswirkungen auf das Curriculum und nicht auf den Lehrer zurückgehen.

(3) Formative Evaluation. Diese Art der Forschung wird oft Prozeßforschung genannt, aber sie ist natürlich nur eine Ergebnisevaluation in einem Zwischenstadium der Curriculumentwicklung. Zwischen formativer Evaluation und der oben beschriebenen zweiten Art der Prozeßforschung gibt es zwei Unterschiede. Der eine Unterschied liegt in den Rollen. Die Rolle der formativen Evaluation besteht darin, die Schwächen und Stärken in der vorläufigen Fassung eines neuen Curriculum zu entdecken. Die Rolle der Forschung bei dieser zweiten Art der Prozeßforschung besteht aus spezifisch eigenen Aufgaben. Sie soll wichtige Fragen über Unterrichtsmechanismen zu beantworten versuchen. Der zweite Unterschied zur formativen Evaluation besteht in der unterschiedlichen Bedeutung, die der Frage zukommt, inwieweit die angewandten Kriterien den Zielen des Curriculum entsprechen. Im Unterschied zur formativen Evaluation braucht die Prozeßforschung der zweiten Art, die sich auf die Erforschung der kausalen Beziehungen zwischen Prozeßelementen richtet, die curricularen Ziele nicht zu berücksichtigen. Diese zwei Arten der Prozeßforschung lassen sich jedoch nicht immer scharf voneinander trennen; sie sind beide für die Curriculumforschung von großer Bedeutung.

Natürlich sollte man, wenn man einen Schulversuch durchführt, seine *Ergebnisse* evaluieren. Im allgemeinen ist ein Versuch sogar so angelegt, daß die Verfahren für die Evaluation der Ergebnisse mitgeplant werden. Das bedeutet jedoch nicht, daß der größte Teil der Forschung Evaluationsforschung ist. Sogar Prozeßforschung ist nicht immer Evaluationsforschung. Daß die Interpretation von Daten als Evaluation von Ergebnissen beschrieben werden kann, bedeutet noch nicht, daß die Interpretation und die Erklärungen sich auf die *Leistung* eines Curriculum beziehen. Sie können sich z. B. auch auf die zeitliche Dauer seiner verschiedenen Elemente richten. Darin liegt ein deutlicher Unterschied; allerdings verrät ein erheblicher Teil der Diskussion pro und contra Evaluationsforschung beträchtliche Unkenntnis über die Grenzen der Evaluation.

Evaluation und Überprüfung der Zielerreichung

Eine Reaktion auf die Verunsicherung durch Evaluation und vielleicht auch auf die Verwendung zu wenig sensibler Evaluationsverfahren besteht in der extremen Relativierung der Evaluationsforschung. In ihrem Verlauf wird die Frage, wie gut ein Curriculum seine Ziele erreicht, anstelle der Frage, wie gut ein Curriculum ist, die zentrale Frage der Evaluation. Es ist jedoch recht unwichtig, wie gut man Ziele erreicht, wenn sie überhaupt nicht wert sind, erreicht zu werden. Dieser Relativismus im Bereich der Evaluation konnte nur dadurch entstehen, daß man davon ausging, Urteile über Ziele seien subjektive, nicht auf rationaler Begründung beruhende Werturteile. Das verhält sich zweifellos oft so; jedoch bedeutet es nicht, daß es in diesem Bereich keine Objektivität geben kann. So könnte z. B. von einem Curriculum über amerikanische Geschichte, das nur auf das Auswendiglernen von Namen und Daten zielt, auf keinen Fall behauptet werden, es sei ein gutes Curriculum, auch dann nicht, wenn es seine Ziele gut erreicht. Genau so unzulänglich wäre jedoch auch ein Curriculum, bei dem überhaupt keine Namen und historische Daten gelernt werden. So wäre auch ein Curriculum in moderner Mathematik, in dem die Mehrzahl der Sekundarschulabgänger nicht gelernt hat, zuverlässig zu addieren und zu multiplizieren, völlig unzulänglich, unabhängig davon, was es sonst noch vermittelt. Solche Werturteile über Ziele werden durchaus abgegeben. Dafür, daß sie unterbleiben sollten, hat noch keiner *gute* Argumente angeführt. Denn dies sind gut begründete Werturteile, die auch spezifisch genug sind.

So gehören zu einem angemessenen Verständnis von Evaluation neben der Leistungsmessung in bezug auf die Ziele auch Verfahren zur Evaluation dieser Ziele. In den nächsten beiden Abschnitten werden wir Evaluationsverfahren erörtern, die sich auf Ziele beziehen und Verfahren einschließen, die solche Beziehungen zu umgehen versuchen. Zuerst soll dargelegt werden, daß Urteile über curriculare Ziele Teil der Evaluation sind, d. h. z. B., daß man nicht einfach beliebige Ziele akzeptieren darf. Das wiederum bedeutet jedoch nicht, daß diese Ziele für jede Schule, jeden Schulbezirk, jeden Lehrer, jede Altersstufe gleich sind. Eine Schule, in der die Mehrzahl der Schulabgänger direkt in den Beruf geht, sollte andere Ziele als eine Schule haben, die 95 Prozent der Schulabgänger zur Hochschule entläßt. Das heißt natürlich nicht, daß die Lehrer, Schulleiter oder Curriculumentwickler bei der Auswahl der Ziele nicht kritisiert werden dürfen. Ein großer Teil der Energie in den gegenwärtigen Bemühungen um Curriculumreform geht unmittelbar auf die Überzeugung zurück, daß die bisherigen Ziele grundsätzlich falsch waren, daß z. B. Lebensanpassung als

Erziehungsziel viel zu stark betont wurde. Nun in die entgegengesetzte Richtung einzuschwenken ist nur allzu leicht und in keiner Weise besser.

Der Prozeß der Relativierung hat jedoch nicht nur zu übertriebener Toleranz für zu restriktive Ziele geführt, sondern hat auch zu inkompetenter Evaluation des Ausmaßes, in dem diese Ziele erreicht wurden, beigetragen. Wie man sich auch zur Evaluation stellt, es läßt sich leicht nachweisen, daß es gegenwärtig in den USA nur sehr wenige professionell kompetente Evaluatoren gibt. Der Gedanke, daß jedes Schulsystem oder jeder Lehrer seine Leistung sinnvoll evaluieren kann, ist genauso abwegig wie die Ansicht, daß jeder Psychotherapeut dazu fähig ist, seine Arbeit mit den Patienten zu evaluieren. Gewiß können sie durch die sorgfältige Untersuchung ihrer eigenen Arbeit viel lernen; sie können ohne Zweifel darin einige gute und schlechte Aspekte identifizieren. Aber wenn man die wichtigen Fragen über den Prozeß oder das Ergebnis beantwortet haben will, braucht man Fertigkeiten und Mittel, die nur sehr schwer zu finden sind.

Intrinsische Evaluation und Ergebnisevaluation

Für die Evaluation eines Unterrichtsinstruments lassen sich zwei Ansätze unterscheiden, die beide auch in der Literatur häufig einander gegenübergestellt werden. Wenn man ein Werkzeug, etwa eine Axt evaluieren will, kann man einmal die Konstruktion der Schneide, die Gewichtsverteilung, die Stahllegierung, die Güte des Ahorngriffs untersuchen; man kann aber auch die Art und Geschwindigkeit der Hiebe untersuchen, die sie in der Hand eines guten Holzfällers macht. In beiden Fällen kann die Evaluation summativ oder formativ sein; denn diese beiden Begriffe kennzeichnen Rollen, nicht unterschiedliche Verfahrensweisen der Evaluation.

Der erste Ansatz zielt auf eine Bewertung des Instruments selbst; in unserem Zusammenhang würde er z. B. der Evaluation der Inhalte, Ziele, Zensierungsverfahren, Lehrereinstellungen entsprechen. Wir nennen diesen Ansatz *intrinsische Evaluation*. Seine Kriterien sind im allgemeinen nicht operational formuliert; sie beziehen sich auf das Instrument selbst und nur indirekt auf seine pädagogischen Auswirkungen. Der zweite Ansatz zielt ausschließlich auf die Untersuchung der Wirkungen des Unterrichtsinstruments auf den Schüler und spezifiziert diese gewöhnlich operational, wobei die Auswirkungen auf Lehrer, Eltern usw. auch berücksichtigt werden können. Zu diesem Ansatz gehört z. B. eine Bewertung der Unterschiede zwischen Vor- und Nachtests, zwischen den Tests der Versuchsgruppe und der Kontrollgruppe in bezug auf eine Anzahl von Kriterien. Wir nennen diese Form der Evaluation *Ergebnisevaluation*. Ihre Befür-

worter würden behaupten, daß nur von Bedeutung ist, welche Wirkungen das Curriculum auf die Schüler hat, und daß die Evaluation der Ziele und Inhalte nur insofern sinnvoll ist, als ihre Ergebnisse mit der Ergebnisevaluation korrelieren. Im Gegensatz dazu könnte der intrinsische Evaluator darauf hinweisen, daß viele wichtige Werte und Charakteristika des Curriculum sich in der reinen Ergebnisevaluation nicht niederschlagen. Um seine Auffassungen zu veranschaulichen, braucht der intrinsische Evaluator nur auf Qualitäten wie Eleganz, Modernität, Struktur eines Curriculum zu verweisen, die sich am besten durch eine unmittelbare Analyse des gesamten Materials erfassen und durch Berücksichtigung von Aspekten des Unterrichts wie Schülerbeteiligung und Unterrichtsklima beurteilen lassen.

Im vorherigen Abschnitt wurde die Behauptung vertreten, daß die bloße Überprüfung des Erreichens von Zielen ein schlechter Ersatz für summative Evaluation sei, da man damit der Grundproblematik der Evaluation ausweiche. Wenn man unter Berücksichtigung der Ziele evaluieren will, muß man die Ziele selbst einer Evaluation unterziehen. Dabei liegt eine Schwierigkeit darin, daß die intrinsische Evaluation sekundäre Ziele und Kriterien benötigt und deshalb sofort auch die Frage nach dem Wert dieser Kriterien unter Bezug auf die primären Ergebniskriterien stellen muß. Im Rahmen der Evaluation ist ein sinnvoller Kompromiß möglich, wenn man einige intrinsische Kriterien und einige Ergebniskriterien berücksichtigt. Zweifellos läßt sie sich in zahlreichen Evaluationssituationen anwenden. Doch bevor wir weiter zu beurteilen versuchen, welches Verhältnis zwischen beiden Arten der Kriterien im Rahmen der Evaluation angemessen ist, wollen wir die entsprechenden praktischen Verfahren etwas näher untersuchen.

Praktische Vorschläge für eine Mischform bei Evaluationsuntersuchungen (Hybrid Evaluation)

Zu Beginn liegen allen Curriculumprojekten allgemeine Zielvorstellungen zugrunde. Selbst wenn sie nur einen interessanteren oder moderneren Unterricht bewirken sollten, wurden sie begonnen, weil man mit dem gegenwärtigen Curriculum in bestimmten Punkten nicht zufrieden war und mit Hilfe des Projekts eine Verbesserung der Situation herbeiführen wollte. Im allgemeinen werden die Ziele und Vorstellungen während der Planungsdiskussion stärker spezifiziert. Wenn drei gleich vertretbare Ziele formuliert werden können, die zu unvereinbaren Anforderungen an das Curriculum führen, kann man sich nach einer gründlichen Diskussion der

Projektziele z. B. dazu entschließen, ein dreigliedriges, auf die unterschiedlichen Lehrer- und Schülerinteressen zielendes Curriculum zu entwickeln. Oder man entschließt sich dazu, dasselbe Ziel – in sehr allgemeinem Sinn – mit drei verschiedenen, gleichwertigen Curriculumvarianten zu erreichen. Diese Varianten werden dann zu den sekundären Kriterien für das Curriculum. Daß diese Varianten oft als miteinander unvereinbar angesehen werden, macht deutlich, daß sie ziemlich wichtigen Inhalts sind.

Ein anderes wichtiges sekundäres Kriterium bezieht sich auf den Geltungsbereich; von Anfang an weiß man, daß wenigstens bestimmte Themen behandelt werden sollten. Wenn das nicht möglich ist, muß eine Abdeckung durch andere Themen erfolgen. Im allgemeinen enthält ein Projekt wenigstens einige abstrakt formulierte Kriterien für primäre und sekundäre Eigenschaften, z. B. für die Verhaltensziele und die intrinsischen Qualitäten eines Instruments. In diesem Fall sollte man eine Mischform der Evaluation wählen, in der beide Kriterienarten berücksichtigt werden, um den Erfolg eines Curriculum festzustellen. In diesem frühen Stadium curricularer Planung sollten einige Projektmitglieder die Aufgabe übernehmen, die Ziele zu formulieren. Die häufig dagegen geäußerten Bedenken sind eine Reaktion auf die rigide Forderung nach präziser Formulierung der Ziele in dieser Phase der curricularen Planung. Alle vom Projektteam akzeptierten Ziele – wie abstrakt oder spezifisch sie auch formuliert sein mögen – einschließlich der Ziele, die nur als vorläufige oder mögliche Ziele angesehen werden, sollten schon in diesem Entwicklungsstadium in einer Liste zusammengestellt werden. Keines der Ziele sollte als absolut verbindlich angesehen werden, da sie lediglich eine Hilfsfunktion haben. Dabei sollte man durchaus auch an das negative Beispiel von Projekten denken, bei denen das kreative Engagement der Mitarbeiter dazu geführt hat, die Bedingungen und Erfordernisse der Realität zu vernachlässigen. Daher sollte man von Anfang an beachten, daß bei zu großer Abweichung vom traditionellen Curriculum, die Implementation des neuen Curriculum in der Schule sehr schwierig wird. Wenn eine umfassende Implementation eine zentrale Zielsetzung des Curriculum ist, sollte sie zusammen mit den kognitiven und affektiven Zielen genannt werden. Eine solche Zielsetzung ist durchaus angemessen, da man das Bildungswesen ja nicht mit Curricula reformieren kann, die niemals in die Schule gelangen. Bereits in einem frühen Entwicklungsstadium empfiehlt es sich, unter Bezug auf diese Ziele die Inhalte auszuwählen.

Im Verlauf der Projektentwicklung sollten drei mit der Zielformulierung zusammenhängende Dinge gesehen werden. Erstens sollten alle formulierten Ziele regelmäßig überprüft und unter Berücksichtigung der im Verlauf der Curriculumentwicklung entstandenen Divergenzen an den

Stellen modifiziert werden, an denen diese Veränderungen zu anderen, wertvolleren Zielen geführt haben. Doch selbst wenn keine Modifikation der Ziele erfolgt, dient die Überprüfung der Ziele dazu, die Curriculumentwickler an die übergreifenden Ziele des Projekts zu erinnern.

Zweitens sollte man möglichst rechtzeitig mit der Konstruktion einer Sammlung von Testaufgaben beginnen. Aus den Leistungstests können Testaufgaben in diese Sammlung aufgenommen werden. Mit ihrer Entwicklung soll eine operationale Fassung der Ziele erstellt werden. Deshalb ist ihre Überprüfung gleichzeitig eine Überprüfung der allgemeiner formulierten Ziele. Schon wenn das Projekt bei der Entwicklung der ersten Einheit eines auf zehn Einheiten angelegten Curriculum ist, empfiehlt es sich, die Testaufgaben so zu formulieren, daß sie in der Schlußprüfung bei der letzten Einheit oder in einem ein Jahr später eingesetzten Test verwendet werden können. Bekanntlich verändert sich die Konzeption der Curriculumziele bei der Formulierung solcher Aufgaben. Man sollte nicht zuviel Zeit dafür aufwenden, und doch sollte man im Zusammenhang mit den Zielen darüber nachdenken, welche Testaufgaben eine bestimmte Lernleistung oder eine Veränderung der Motivation in der abschließenden Prüfung oder in einer späteren Untersuchung erfassen. Manchmal werden sich keine Testaufgaben entwickeln lassen, da nicht alle Werte eines Curriculum in der abschließenden oder in einer später erfolgenden Prüfung direkt in Erscheinung treten. Wo sie sich nicht zeigen, sollte wenigstens angegeben werden, wann und wie sie sich etwa in der Berufswahl, in den Einstellungen von Erwachsenen oder im Unterrichtsverhalten ausdrücken.

Drittens sollte man in einem mittleren Stadium der Curriculumentwicklung versuchen, einige externe Beurteilungen über den Zusammenhang zwischen den angegebenen Zielen, den wirklichen curricularen Inhalten und den gesammelten Testaufgaben zu erhalten. Denn ohne solche Beurteilungen dürfte die Validität der Tests und der praktische Nutzen des Curriculum wohl eingeschränkt sein. Um diese Aufgaben zu erfüllen, muß der einzelne Beurteiler nicht unbedingt ein professioneller Evaluator sein. Professionelle Evaluatoren sind sogar oft wenig geeignet. Ein Fachwissenschaftler, ein pädagogischer Psychologe oder ein Curriculumfachmann kann diese Aufgaben besser erfüllen. Die dazu benötigten Qualitäten sind nicht mit professionellen Fähigkeiten identisch. Sie bestehen in der Fähigkeit zur »Konsistenzanalyse«. Dies ist ein Gebiet, für das man Mitarbeiter nicht ohne Probezeit einstellen sollte. Vielleicht sollte der Wissenschaftler, der die Konsistenzanalyse macht, wenigstens in der Probezeit nicht persönlich mit dem Projektteam zusammenarbeiten. Zu diesem Zeitpunkt genügt vielleicht ein kurzer schriftlicher Bericht, in dem die vorhandenen In-

formationen zur Verfügung gestellt werden. In einem späteren Stadium jedoch, wenn große Divergenzen zwischen (a) verbalisierten, (b) impliziten und (c) getesteten Zielen vermieden werden sollen, ist die Konsistenzanalyse sehr wichtig. Ein Wissenschaftler kann mit einer guten Konsistenzanalyse nicht nur verhindern, daß das Projekt durch den Übereifer seiner Mitarbeiter oder durch Fehleinschätzungen seiner tatsächlichen Auswirkungen in Sackgassen gerät, sondern er kann auch wertvolle Anregungen zur Entwicklung des Projekts in eine neue Richtung liefern. Er muß auf fehlende und überflüssige Testaufgaben in der entsprechenden Sammlung und auf fehlende und irrelevante Ziele in der entsprechenden Liste achten. Schließlich läßt sich auch die Psychotherapie nicht dadurch rechtfertigen, daß der Psychotherapeut *meint*, er würde dem Patienten helfen, sondern nur dadurch, daß er es tatsächlich tut; entsprechendes gilt für die Curriculumforschung.

Daher braucht man ein besser entwickeltes, wenn auch ähnliches Verfahren, um die Diskrepanz zwischen den curricularen Materialien, Zielen, Tests und den Normen eines solchen Curriculum zu identifizieren. Die Curriculumhersteller neigen leicht zu der Annahme, daß diese Größen kongruent sind. Daher bedarf es eines externen Evaluators. Wenn er gut ist und über hinreichende Erfahrungen verfügt, wird er Nebenwirkungen und Diskrepanzen entdecken, die für die finanzierenden Ministerien oder Stiftungen und die Adressaten aufschlußreich sind. Der Beweis dafür, daß der Evaluator nicht nur seine eigenen Vorurteile zum Maßstab seiner Evaluation macht, muß in seinen Argumenten liegen, die in vielen Fällen die Curriculumhersteller durchaus überzeugen können.

Wenn man während der ganzen Curriculumentwicklung in der beschriebenen Weise vorgeht, wird man am Ende eine große Testaufgabensammlung haben. Die Antworten auf diese Testaufgaben können dazu dienen, jedes angestrebte Ergebnis des Curriculum zu überprüfen; das Ergebnis dürfte dann im allgemeinen wirklich nur auf das Curriculum zurückzuführen sein. Diese Sammlung hat mehrere erhebliche Vorteile. Sie ist eine operationale Fassung der Curriculumziele, die erstens den Schülern eine Vorstellung von den an sie gestellten Erwartungen vermitteln kann, die zweitens ein wertvolles Hilfsmittel für die Konstruktion der abschließenden Prüfung ist und die drittens dem Curriculumentwickler ein detailliertes Bild seines eigenen Erfolgs bietet. Um sich darüber Gewißheit zu verschaffen, kann er jedem Schüler eine jeweils unterschiedliche Zufallsauswahl von Testaufgaben im Rahmen einer formativen Evaluation vorlegen, anstatt jedem Schüler eine bestimmte Zufallsauswahl zur Beantwortung zu geben, wie man es vielleicht gerechterweise in einer abschließenden Prüfung machen müßte.

Bisher wurden die Grundzüge einer Evaluation beschrieben, die unter Zuhilfenahme sekundärer Kriterien erfolgt. Dabei haben wir vor allem auf inhaltliche Charakteristika als eine Form der Zielangabe hingewiesen, da Curriculumteams oft behaupten, daß ein Vorteil ihres Curriculum in der Modernität und Aktualität seiner Inhalte liegt. Um diese Behauptung zu verifizieren, braucht man das Curriculum nur von einigen qualifizierten Fachwissenschaftlern analysieren zu lassen. Dabei ergeben sich jedoch besondere Schwierigkeiten. Bestenfalls können wir in Erfahrung bringen, ob das Curriculum starke Verzerrungen oder Mängel hinsichtlich der wichtigsten derzeitigen Kenntnisse und Anschauungen enthält. Offen bleibt bei diesem Verfahren die Frage – worauf vor allem der Befürworter der Ergebnissevaluation hinweisen würde –, inwieweit die Ziele und Inhalte des Curriculummaterials den Schülern wirklich vermittelt werden. Selbst wenn ein Curriculum im Hinblick auf seine fachliche Qualität für wissenschaftliche Experten nicht ganz zufriedenstellend ist, kann es von den fachlichen Inhalten manchmal durchaus eine bessere Vorstellung vermitteln als ein nur nach fachwissenschaftlichen Kriterien entwickeltes anderes Curriculumprojekt. Der Vorteil der geschilderten Methode besteht darin, ein Verfahren bereitzustellen, die Lücke zwischen intrinsischer Evaluation und Ergebnisevaluation, zwischen bloßem Messen, ob die Lernziele erreicht worden sind, und vollständiger Evaluation auszufüllen.

Weitere Verbesserungen der obigen Ausführungen sind erforderlich und in jeder guten Untersuchung unerlässlich. Sie hängen mit der Rolle der Konsistenzanalyse zusammen und sind für formative Evaluationsuntersuchungen noch wichtiger als für summative, da sie die Gründe für schlechte Ergebnisse zu entdecken helfen. Man muß in Erfahrung bringen, ob es gelungen ist, eine Entsprechung zwischen drei zusammenhängenden Problemen herzustellen:

1. die Entsprechung von Zielen und Curriculuminhalten,
2. die Entsprechung von Zielen und Prüfungsinhalten,
3. die Entsprechung von Curriculuminhalten und Testinhalten.

Im Grunde genommen, brauchte man nur zwei der Probleme zu lösen, um auch das dritte evaluieren zu können. Aber in der Praxis empfiehlt es sich, um Irrtümer möglichst auszuschließen, jedes Problem unabhängig vom anderen zu behandeln. Aufgrund dieser Überlegungen könnte man den Eindruck gewinnen, als könnte eine Person oder Gruppe alle Entsprechungen abschätzen. Es empfiehlt sich jedoch, alle Einschätzungen unabhängig voneinander durchführen zu lassen und sogar von nicht an dem Projekt beteiligten Personen wiederholen zu lassen. Nur so kann man wahrscheinlich die wirklichen Gründe für enttäuschende Ergebnisse herausfinden. Sogar das Curriculumprojekt des Physical Science Study Committee, das so sorg-

fällig wie die meisten derzeitigen Curriculumprojekte getestet wurde, hat an keiner Stelle die hier als notwendig bezeichneten Verfahren der Analyse angewandt.

Das schwierigste Problem der Testtheorie und der Herstellung von Tests liegt in der Konstruktvalidität, die durch das angesprochene Problem vor allem berührt wird. Man darf die mit ihr verbundenen Probleme nur vernachlässigen, wenn man dafür in Kauf zu nehmen bereit ist, (1) daß die intendierten Ziele nicht im Curriculum realisiert werden oder (2) daß die Prüfungen nicht testen, was das Curriculum lehrt, oder (3) daß die Prüfungen nicht die Werte und Materialien testen, die das Curriculum vermitteln soll. Es gibt in der Praxis viele Möglichkeiten, mit denen man die hier beschriebenen Vergleiche durchführen kann: die Anwendung von Q- und R-Techniken (Q-sorts, R-sorts), von parallelisierten Tests und projektiven Tests für die Analyse usw. Die Aufgabe muß jedoch im Rahmen der Evaluation in irgendeiner Form gelöst werden.

Das Für und Wider einer reinen Ergebnisevaluation

Der »reine« Ergebnisevaluator betrachtet die sich bei der beschriebenen experimentellen Planung ergebenden Schwierigkeiten mit Skepsis. Nach seiner Auffassung ist die Berücksichtigung von Ziel- und Inhaltsbewertung oder einer anderen sekundären Bewertung im Rahmen der Curriculum-evaluation nicht nur irrelevant, sondern auch unzuverlässig. Seiner Ansicht nach braucht man weder zu untersuchen, was ein Lehrer zu tun behauptet oder nach Aussage seiner Schüler tut, noch was er im Unterricht sagt und was die Schüler in ihren Schulbüchern lesen. Wichtig ist lediglich, was der Schüler nach Beendigung seiner Arbeit mit dem Curriculum sagt und was er nicht gesagt hätte, wenn er nicht mit diesem Curriculum gearbeitet hätte. Nach seiner Auffassung kommt es also nur darauf an, die Auswirkungen eines Curriculum festzustellen, und nicht, ob ihm gute Intentionen zugrunde liegen.

Für den »reinen« Ergebnisevaluator gibt es jedoch auch erhebliche Schwierigkeiten. Er kann das Problem der Konstruktvalidität nicht ganz umgehen, d. h. er kann den Schwierigkeiten nicht ausweichen, die in dem Versuch liegen, die Lernleistung der Schüler *mit einem sinnvollen Grad an Allgemeinheit* zu beschreiben. Es ist zwar einfach, die Testergebnisse so darzustellen, daß ersichtlich wird, wieviel Schüler (in Prozenten) die einzelnen Testaufgaben gelöst haben; man muß jedoch wissen, ob man aufgrund dieser Lösung sagen kann, daß sie bestimmte Elemente der Astronomie oder den ökologischen Ansatz in der Biologie *besser verstehen*. Doch von

Daten über die Lösung spezifischer Testaufgaben zu derartigen Schlußfolgerungen ist es ein weiter Weg. Der Ergebnissevaluator hat zwar recht mit der Behauptung, daß man für solche Schlußfolgerungen nicht unbedingt eine Diskussion der Ziele benötigt. *Ohne* eine solche Zieldiskussion verfügt man jedoch nicht über die benötigten Daten, um eine Entscheidung über die Angemessenheit unterschiedlicher Erklärungen des Erfolgs oder Versagens bestimmter Aspekte des Curriculum zu fällen. Wenn man z. B. den Ansatz der reinen Ergebnissevaluation wählt und dann entdeckt, daß die von den Schülern erinnerten und reproduzierten Inhalte von Fachwissenschaftlern als inadäquat bezeichnet werden, dann weiß man noch nicht, ob dies auf die Unzulänglichkeit der Intentionen, der curricularen Realisierung der Ziele oder der curriculumspezifischen Prüfungen zurückzuführen ist.

Das soll weiter erläutert werden: Wenn man eine reine Ergebnissevaluation durchführen will und die Schülerleistung nur am Ende des Curriculum von einem externen Beurteiler beurteilen lassen will, muß man jemanden für die Beurteilung auswählen. Dabei zeigt es sich, daß die Auswahl von bestimmten Interessen und bestimmten Zielen abhängt, die man genauso gut explizit machen könnte. Genauso muß der Evaluator die Schülerleistung auf ein *bestimmtes* Kriterium beziehen. Das kann seine Auffassung von einem angemessenen Fachverständnis des ganzen Bereichs oder seine Ansicht über die angemessene Verständnisfähigkeit eines Schülers der zehnten Klasse sein. Der Ergebnissevaluator behauptet zu Recht, daß man von jeder Zieldiskussion absehen und doch genau feststellen kann, was Schüler gelernt haben. Mit dem gleichen Recht geht er davon aus, daß letzteres die wichtigste Variable überhaupt ist. Aber er irrt sich in der Annahme, daß man ohne weiteres die Ergebnisse des Lernens so beschreiben kann, daß sie für unsere Zwecke nützlich sind, oder daß man das Curriculum ohne Bezug auf allgemeine Ziele rechtfertigen kann. Deshalb ist eine reine Ergebnissevaluation etwas oberflächlich, so daß beim augenblicklichen Diskussionsstand eine Mischform der Evaluation vorzuziehen ist.

Der Ergebnissevaluator weist konsequenterweise zu Recht darauf hin, daß es unverantwortlich ist, »elegante«, »moderne«, »präzise« Curricula zu entwickeln, wenn ihre Qualitäten nicht bis zu den Schülern gelangen. Solange es sich dabei nur um sekundäre Qualitäten handelt, reicht es aus, ihre Existenz lediglich anzunehmen; sobald man sie aber für wesentlich hält, darf man sich damit nicht begnügen. Deshalb muß man einen wissenschaftlichen Evaluator berufen, dessen Aufgabe darin besteht, nicht nur die Curriculummaterialien oder die Sammlung der Testaufgaben, sondern auch die exakte Leistung der Klasse bei jeder Testaufgabe zu untersuchen. Mit Hilfe dieser Ergebnisse soll er abschätzen, inwieweit das Cur-

riculum die Inhalte angemessen vermittelt. Trotzdem fehlt uns dann immer noch die Diagnose der Ursache für die Mängel. Deshalb ist dies ein schlechtes Verfahren für formative Evaluation; doch wir können summative Evaluation durch dieses Verfahren vereinfachen. Deshalb müssen wir unseren umfassenden Plan durch eine genaue Analyse der *Ergebnisse* der Schülertests und nicht nur des Curriculum und der Testinhalte ergänzen. Es lohnt sich nicht, viel Mühe auf die Aufstellung und wechselseitige Analyse der Ziele, Tests und Inhalte eines Curriculum zu verwenden, wenn man lediglich versucht, eine Prozentangabe in bezug auf die maximal möglichen Punkte als Index für das Ausmaß zu benutzen, in dem die Ziele erreicht worden sind – es sei denn, diese Angabe liegt zufällig ziemlich dicht bei 100 oder 0 Prozent. Die Leistungen der Schüler in den Tests der mittleren Stadien müssen analysiert werden, um exakt festzustellen, wo z. B. ein ausreichendes Verständnis grundlegender Fakten und die Übung wichtiger Fertigkeiten fehlen. Prozentangaben sind dabei nicht so wichtig. Es ist vielmehr die *Art* der Fehler, die für die Evaluation und für die Neufassung des Curriculum wichtig ist. Daher braucht man für die formative und die summative Evaluation eine klare Beschreibung der Stärken und Schwächen des Curriculum. Die umfangreiche Sammlung von Testaufgaben ist ein bewährtes Verfahren, um die Unzulänglichkeiten im Curriculum zu lokalisieren. Aber es kann nur dann voll ausgenutzt werden, wenn die Ergebnisse adäquat evaluiert werden. Dazu muß man in gleicher Weise unabhängige Beurteiler, Hypothesenentwicklung, Testen der Art der Fehler, Längsschnittanalysen von Leistungsunterschieden bei gleichen Schülern verwenden. Daher ist eine angemessene Evaluation von Curriculummaterialien sehr schwierig. Zudem ist die Verwendung von Aufsätzen, die Entwicklung und Anwendung von neuen Instrumenten, die Verwendung der Berichte der Versuchsleiter, die Übertragung dieses gesamten Materials in ein speziell entwickeltes Beurteilungsschema kostspielig und zeitraubend. Doch ist diese Art auch nicht zeitraubender als gute Forschungs- und Entwicklungsarbeit in der Technik. In diesem Zusammenhang soll eine weitere Unterscheidung zwischen zwei Methoden getroffen werden.

Vergleichende und nicht-vergleichende Evaluation

Die Ergebnisse der Evaluation neuerer Curricula sind oft erstaunlich ähnlich. Bei dem Vergleich zwischen Schülern, die nach dem alten Curriculum unterrichtet werden, und Schülern, die nach dem neuen unterrichtet werden, schneiden letztere gewöhnlich besser bei den für ihr Curriculum kon-

struierten Tests ab als erstere und schlechter bei den für das alte Curriculum konstruierten Tests; für die Schüler, die nach dem alten Curriculum unterrichtet werden, sind die Ergebnisse entsprechend umgekehrt. Es fehlt im allgemeinen ein größerer Leistungszuwachs für dieselben Kriterien. Leicht hat man den Eindruck, daß ein solches Ergebnis kaum wirklich relevant ist; denn daß das Ergebnis positiv und nicht negativ ist, hängt ausschließlich von den der Evaluation zugrunde gelegten Kriterien, d. h. von den benutzten Tests ab. Aufgrund dieses Sachverhalts erhebt sich die berechtigte Frage, ob man nicht den von den Fachwissenschaftlern den Inhalten und Zielen zugeteilten Wert viel schwerer gewichten sollte als die geringen Unterschiede im Leistungsniveau in bezug auf ungewichtete Kriterien. Wenn man sich dazu entschließt, werden relativ unbedeutende Leistungsverbesserungen hinsichtlich der richtigen Ziele sehr wertvoll, und, so gesehen, schneidet das neue Curriculum bei dem Vergleich wesentlich besser ab. Ob diese Veränderung der Gewichtung sich wirklich rechtfertigen läßt, muß gründlich untersucht werden. Dazu muß z. B. die wirkliche Bedeutung der zum Verständnis moderner Physik im alten Curriculum fehlenden Elemente analysiert werden. Denn nur zu leicht ist man in der Versuchung, die neue Gewichtung für richtig zu halten, da man ja von der Überlegenheit des neuen Curriculum fest überzeugt ist.

Ferner muß man sich fragen, ob die Tests bei den nach dem neuen Curriculum unterrichteten Schülern wirklich das Verständnis erfassen. In diesem Zusammenhang empfiehlt es sich, eine umfangreiche Sammlung von Testaufgaben zu verwenden. Cronbach schlägt eine Sammlung von 700 Testaufgaben vor. Bei einer gründlichen Evaluation eines ein- oder zweijährigen Curriculum ist diese Größenordnung durchaus sinnvoll. In dieser Sammlung sollte man keine Testaufgaben aufnehmen, die nur auf die terminologischen Unterschiede des neuen Curriculum zielen. Wenn die Sammlung hauptsächlich solche Aufgaben enthält, werden die Schüler des neuen Curriculum natürlich viel besser abschneiden, obwohl eine echte Überlegenheit nicht besteht. Cronbach weist daher mit Recht auf die Unzulänglichkeit curriculumabhängiger Terminologie hin, obwohl er mit der Unterscheidung und Trennung zwischen Verständnis und Terminologie zu weit geht. Deshalb sollte man auch hier externe Evaluatoren zur Konstruktion und Beurteilung der Aufgabensammlung hinzuziehen.

Die Reaktionen auf diesen Sachverhalt reichen von der etwas naiven Vermutung, daß solche Resultate nur die Schwächen von Evaluationsverfahren zeigen, bis zur folgenden interessanten Überlegung Cronbachs: »Da die Ergebnisse von Gruppenvergleichen sehr fragwürdig sind, sollte man meiner Meinung nach eine gute Untersuchung in erster Linie so planen, daß sie es erlaubt, Leistungen einer genau umschriebenen Gruppe am

Ende eines Curriculum im Hinblick auf die wesentlichen Ziele und Nebenwirkungen zu bestimmen« (1963, 43, 48). Cronbach schlägt offensichtlich ein Verfahren vor, bei dem wir zwar den Vergleich mit Lernzielen nicht vermeiden können, wohl aber den mit einer anderen Gruppe, die der Testgruppe in bezug auf relevante Variablen entspricht. Wie sieht nun eine nicht-vergleichende alternative Verfahrensweise für Evaluation aus? Cronbach führt dazu aus: »Unser Problem ist mit dem eines Ingenieurs, der ein neues Auto überprüft, vergleichbar. Er kann sich die Aufgabe stellen, die Leistungsfähigkeit und Zuverlässigkeit des Autos genau zu bestimmen. Es würde aber an dem Problem vorbeiführen, wenn er sich die Frage stellen würde: Ist dieses Auto besser oder schlechter als die konkurrierende Automarke?« (a. a. O.). Es ist richtig, daß der Ingenieur vielleicht nur an der Frage der Leistung und Verlässlichkeit des neuen Autos interessiert ist. Aber kein Ingenieur hat jemals nur dieses Interesse gehabt, und keiner wird es jemals haben. Ziele sind nur im Kontext einer praktischen Entscheidung wichtig. Unrealistische Lernziele z. B. sind nicht wichtig. Das Maß der Leistung und der Verlässlichkeit eines Autos und unser Interesse daran hat seinen Ursprung *ausschließlich* in dem Wissen darüber, was sich bisher innerhalb einer bestimmten Preisklasse mit bestimmtem Raum und bestimmtem Gesamtgewicht als möglich erwiesen hat. Die Anwendung von geeichten Instrumenten ist keine Alternative, sondern eine indirekte Form der vergleichenden Untersuchung. Was wir messen, hat eine absolute Qualität. Der Grund dafür, daß wir sie messen, liegt darin, daß wir sie für eine wichtige Variable im Rahmen eines Vergleichs halten. Wenn man genau wüßte, daß alle Autos die Eigenschaft P haben, dann brauchte man sie nicht zu messen. Aber im allgemeinen ist P eine stärker oder geringer bewertete Variable, der unterschiedliche Bedeutung zugemessen wird und die wir messen, weil sie eine Grundlage für den Vergleich bildet.

Dasselbe gilt für den Bereich der Curriculumentwicklung. Es gibt bereits Curricula für fast jedes Thema, und es besteht wahrhaftig kein Interesse daran, Curricula um ihrer selbst willen zu produzieren. Man ist an neuen Curricula interessiert, weil sie vielleicht in wichtigen Aspekten besser als die vorhandenen sind. Man kann jemanden beauftragen, ein Curriculum in bezug auf bestimmte Variablen zu bewerten, ohne gleichzeitig zu fordern, daß er die Leistung anderer Curricula bezüglich dieser Variablen feststellt. Aber wenn man das Curriculum – im Unterschied zur Beschreibung seiner Leistung – *evaluiert*, dann ist man unweigerlich mit der Frage seiner Über- oder Unterlegenheit in bezug auf andere Curricula konfrontiert. Die Behauptung, ein Curriculum sei ein »wertvoller Beitrag«, ein »wünschenswertes«, »nützliches« oder »gutes« Curriculum, bedeutet bereits, ihm einen relativen Wert zuzuordnen. Tatsächlich sind die Skalen,

die wir zur Leistungsmessung von Curricula anwenden, oft Prozentskalen oder andere Skalen mit implizitem Vergleich.

Es gibt sogar wichtige Gründe, die Frage sofort in der Form eines Vergleichs zu formulieren. Vergleichende Evaluation ist oft viel einfacher als nicht-vergleichende zu handhaben, weil man häufig Tests benutzen kann, die Unterschiede erbringen, anstatt eine absolute Skala konstruieren zu müssen, um dann schließlich die absoluten Testwerte (scores) zu vergleichen. Handelt es sich z. B. um Curricula für das Lernen von Schach, kann man zwei Gruppen in bezug auf Hintergrundvariablen parallelisieren, unterrichtet sie nach verschiedenen Curricula und läßt sie dann in einem Turnier gegeneinander spielen. Ein absolutes Maß für eine Fertigkeit aufzustellen wäre außerordentlich schwierig; durch die Art der vergleichenden Evaluation jedoch können wir leicht konsistente und signifikante Unterschiede erhalten. Cronbach macht nicht den Fehler der reinen Ergebnisevaluation, den Bezug auf allgemeine Ziele zu leugnen; aber er schlägt ein Verfahren vor, das das implizit vergleichende Element in jedem Bereich des »Social Engineering« einschließlich der Curriculumevaluation unterschätzt, so wie Ergebnisevaluation die implizite Aussagekraft abstrakter Kriterien unterschätzt.

Sodann entwickelt Cronbach in diesem Abschnitt einen Gedankengang, über den es keine Meinungsverschiedenheiten gibt. Er weist darauf hin, daß in allen Vergleichen zwischen sehr unterschiedlichen Unterrichtsinstrumenten kein wirkliches Verständnis der Gründe für eine Leistungsdifferenz dadurch gewonnen wird, daß man die Überlegenheit eines Unterrichtsinstruments gegenüber den anderen entdeckt. Denn niemand kennt einzelne Elemente, die für die Überlegenheit verantwortlich sind. Aber ein Verständnis der Leistungsdifferenz zwischen Curricula ist nicht das *einzig* Ziel der Evaluation. Sie richtet sich ebenso auf die Fragen der Unterstützung, Annahme, Anerkennung, Verbesserung von Curricula usw. Diese äußerst wichtigen Fragen können schon – wenn auch nicht immer vollständig – durch die Feststellung der Überlegenheit des Curriculum beantwortet werden. Wie wir in einem früheren Abschnitt ausgeführt haben, ist die reine Ergebnisevaluation der zielbezogenen Evaluation darin unterlegen, daß zu ihren Ergebnissen nicht die Daten gehören, die uns helfen, die Ursachen der Schwierigkeiten usw. aufzudecken. Nach Cronbachs Auffassung kann hier eine nicht-vergleichende Methode mit größerer Wahrscheinlichkeit die Daten liefern, die für spätere Verbesserungen benötigt werden. Das ist jedoch nicht ein Vorteil der nicht-vergleichenden Methode als solcher. Es ist lediglich der Vorteil von Methoden, bei denen eine größere Zahl von Variablen genauer untersucht wird. Wenn man aber die Gründe für die Unterschiede zwischen den Curricula feststellen will, kann man »gut

kontrollierte Untersuchungen kleinerer Größenordnung mit Gewinn dazu benutzen, alternative Fassungen desselben Curriculum zu vergleichen«, während der umfassende Vergleich großen Ausmaßes weniger wertvoll ist. Das bedeutet aber nicht, daß keine vergleichenden Untersuchungen im Lauf des Evaluationsverfahrens benötigt werden. Cronbachs Argument besagt lediglich, daß man, um *Erklärungen* zu erhalten, mehr Kontrollgruppen und kurzfristige Untersuchungen braucht, als für summative *Evaluation* notwendig sind. Das ist unbestreitbar, aber beweist nicht, daß man für die umfassende Evaluation einen umfassenden Vergleich vermeiden sollte.

Man kann den entscheidenden Punkt in Form folgender Analogie fassen: In der Geschichte der Konstruktion von Automotoren ist es häufiger vorgekommen, daß ein Konstrukteur einen Motor entwarf, der der Konkurrenz ganz überraschend überlegen war. Vielleicht hatte man etwa 30 Variablen bei Konstruktion des neuen Motors verändert; nachdem der Motor in die Produktion gegangen war, wußte man noch lange nicht, der Konstrukteur eingeschlossen, welche von diesen Variablen vor allem für die Verbesserung verantwortlich war. Aber die Entscheidung, den Motor in die Produktion zu geben, die Entscheidung, weitere Forschung in den Motor zu investieren, machte es möglich, den Grund des Erfolgs herauszufinden. Tatsächlich war für den Beginn einer neuen Ära der Konstruktion von Motoren wirklich vor allem die *vergleichende* Evaluation notwendig. Man setzt ein großes Team ein und hofft, daß es Gold entdecken wird, muß jedoch auf das Metall stoßen, bevor das Kapital investiert wird, das man benötigt, um die Lage von Flözen genau festzustellen. So müssen wir in jedem Bereich arbeiten, wo wir zu viele Variablen und zu wenig Zeit haben.

Praktische Verfahrensweisen für die Evaluation mit Kontrollgruppen

In einer seiner Hauptthesen behauptet Cronbach, daß Vergleiche mit Kontrollgruppen für die Curriculumentwicklung nicht sehr nützlich sind. Nach unseren Ausführungen bietet sein Versuch, eine brauchbare Alternative bereitzustellen, im Kontext einer typischen Evaluation keine heuristische Möglichkeit. Man muß deshalb einige seiner Einwände gegen Kontrollgruppen zu entkräften versuchen, die sich unserer Meinung nach für die Evaluation durchaus eignen.

Der These, daß grobe Vergleiche nur geringe Unterschiede erbringen, kann man zunächst dadurch begegnen, daß man die Präzision der Untersuchungsmethoden verbessert. Das bedeutet die Entwicklung einer grö-

berer Zahl verschiedenartiger Testaufgaben, die Vergrößerung der Gruppe, um differenziertere Unterschiede zu gewinnen, und die Entwicklung neuerer und besserer Tests. Wenn wir einen besonders relevanten Faktor entdecken, versuchen wir, diesen bei der Neukonstruktion des Curriculum stärker zu berücksichtigen, um seinen Erfolg zu vergrößern. Doch bleibt die Tatsache bestehen, daß man die Verbesserung des Curriculum wahrscheinlich nur in geringen Testunterschieden messen kann und daß das Streben nach größeren Unterschieden im allgemeinen eine Methode mit mehreren Angriffspunkten erfordert; daher muß sie nicht nur auf das Curriculum zielen, sondern muß auch die Verfahren der Gruppeneinteilung, die Stoffdarbietung des Lehrers, die Verteilung der Unterrichtszeit berücksichtigen. Darüber hinaus muß sie versuchen, Langzeiteffekte, z. B. ein allgemeines Anwachsen des Interesses, zu erforschen, die von Verbesserungen in jedem Bereich des Schulcurriculum bewirkt werden. Darin liegt eine wichtige Aufgabe der Evaluation. Eine genaue Parallele finden wir im Bereich der Psychotherapie, in der wir menschliches Verhalten dadurch zu ändern versuchen, daß wir Menschen mehrere Jahre lang wöchentlich für einige Stunden behandeln. Vielleicht sind wir allzusehr an die Erfindung von Wunderdrogen oder an technologische Durchbrüche im Bereich der Raumfahrt gewöhnt, daß wir diesen plötzlichen Fortschritt nicht mehr als ungewöhnliche Ausnahme begreifen. Selbst im Bereich der Konstruktion von Autos, um bei Cronbachs Beispiel zu bleiben, deuten zahlreiche Erfahrungen darauf hin, daß die Weiterentwicklung eines bereits erprobten Modells bessere Resultate erbringt als die Einführung eines vielversprechenden, aber radikal neuen Modells. Realistisch gesehen, kann man keine großen Sprünge, sondern nur eine langsame und stetige Verbesserung als Ergebnis erwarten, wobei sich natürlich manchmal Sackgassen nicht vermeiden lassen. Das systematisch geplante Experiment, das ein Curriculum einem anderen gegenüberstellt, sollte genügend eindeutige Ergebnisse haben, um die zu ihrer Gewinnung erforderlichen Kosten zu rechtfertigen. Cronbach berücksichtigt jedoch dabei nicht genügend, daß das Ausbleiben klarer Unterschiede oft gerade das Ergebnis ist, das man benötigt. Wenn man sich wirklich davon überzeugt hat, daß man mit guten Tests die wichtigsten Kriteriumsvariablen erfaßt, dann *ist* es äußerst informativ, zu sehen, daß die Ergebnisse gleich sind, denn »kein Unterschied« bedeutet keineswegs »kein Wissen«.

Natürlich können wir aus einem Null-Resultat nicht schließen, daß alle in einem neuen Curriculum enthaltenen Verfahren wertlos sind. Wir müssen mit Mikro-Untersuchungen fortfahren, aus denen wir entnehmen können, ob eine dieser Techniken etwas wert ist. Die Durchführung einer groben vergleichenden Untersuchung kostet stets dasselbe, unabhängig davon,

welche Resultate sie erbringt; hinzu kommt, daß man sie früher oder später doch durchführen muß. Es ist also falsch aufzuhören, wenn man nicht-signifikante Unterschiede entdeckt hat; vielmehr muß man weitere analytische Forschung der von Cronbach empfohlenen Art betreiben. Wenn auch Cronbach in seinem Beitrag den Untersuchungen mit Kontrollgruppen eine geringe Bedeutung zumißt, kann man diese jedoch lediglich dann als unangemessen bezeichnen, wenn man sie als *einzig*e Evaluationsmethode für den *gesamten* curricularen Bereich ansieht. Wir wollen hier versuchen, einige praktische Vorschläge für die Anlage von Versuchen zu machen, die mehr als eine grob vergleichende Evaluation ergeben.

Ein wichtiger Grund für Cronbachs Ablehnung von vergleichenden Untersuchungen liegt in der Überzeugung, daß man keine Doppelblindversuche durchführen könne. »In einem pädagogischen Versuch ist es schwer, die Schüler über ihre Rolle als Versuchsgruppe im unklaren zu lassen. Die Fehlerquellen, die durch die Person des Lehrers bedingt sind, können kaum so gut kontrolliert werden wie die durch den Arzt im Doppelblindversuch bedingten. Infolgedessen kann man nicht mit Sicherheit sagen, ob ein beobachteter Gewinn der pädagogischen Innovation an sich zuzuschreiben ist oder dem größeren Engagement von Lehrer und Schülern bei einem Versuch mit einer neuen Methode« (Cronbach 1963, 42, 43). Cronbachs Schlußfolgerungen sind jedoch übereilt. Im medizinischen Bereich bilden nicht die Untersuchungen über Medikamente die Analogie, bei denen wir Doppelblindbedingungen ohne weiteres herstellen können, sondern psychotherapeutische Untersuchungen, bei denen der Therapeut die Behandlung engagiert durchführt und der Patient nicht darüber in Ungewißheit gelassen werden kann, daß er behandelt wird. Wenn Cronbachs Argumentation richtig ist, wäre es nicht möglich, eine adäquate Untersuchung der Ergebnisse einer psychotherapeutischen Behandlung zu planen. Das *ist* jedoch möglich, und die entsprechende Methode besteht darin, mehrere Vergleichsgruppen zu benutzen (vgl. Scriven 1959). Wenn wir nur eine Kontrollgruppe benutzen, können wir keine Aussage darüber machen, ob das Engagement oder die Versuchstechnik den Unterschied erklärt. Wenn wir aber mehrere Versuchsgruppen haben, können wir die Auswirkungen des Engagements einschätzen. Wir vergleichen mehrere Therapiegruppen, in denen der Therapeut jeweils engagiert ist, in denen aber die Therapiemethode jeweils verschieden ist. Nach Möglichkeit sollte man Therapiemethoden mit sehr verschiedenen Verfahrensweisen anwenden. Die Patienten, die nach *einer* Methode behandelt werden, sollten jedoch soweit wie möglich vergleichbar sein. Es gibt eine Anzahl von Therapien, die die erste Bedingung in mehreren Dimensionen erfüllen, und es ist leicht, Pseudo-Therapien zu entwickeln, die vielversprechend genug sind, um bei eini-

gen praktizierenden Ärzten Engagement zu wecken. Die Feststellung der Unterschiede in Verbindung mit der Kovarianzanalyse ermöglicht zu entscheiden, ob Engagement der einzige oder ein Hauptfaktor beim therapeutischen Erfolg ist, wenn Doppelblindbedingungen nicht erreicht werden können. Dies ist auch nicht der einzige Forschungsplan, mit dem das erreicht werden kann; andere Methoden sind verfügbar, und fähige Wissenschaftler werden zweifellos noch weitere Methoden entwickeln können, die es uns ermöglichen, dieses Forschungsproblem zu bewältigen. Eine Doppelblinduntersuchung ist also nicht unbedingt notwendig.

Im curricularen Bereich sind die Fragen noch etwas schwieriger als im Bereich der Psychotherapie, weil es sehr schwer ist, gemeinsame Elemente aus den verschiedenen Vergleichsgruppen auszusondern. Zwar wird der durchschnittlich intelligente Patient aufgrund der vielen unzulänglichen Ärzte und aufgrund des Wunsches, geheilt zu werden, fast alles als Form von Therapie akzeptieren, doch ist es längst nicht so einfach, Schüler und Lehrer davon zu überzeugen, daß sie einen bestimmten Unterricht in Geometrie bekommen oder geben sollen, es sei denn, die Art der Geometrie erscheint ihnen sinnvoll. Und wenn es sich so verhält, dann ist die Interpretation jedes der möglichen Ergebnisse mehrdeutig, d. h., wenn mehrere Gruppen ungefähr gleich gut abschneiden, kann es *entweder* sein, daß das Engagement sich so auswirkt, *oder*, daß die gemeinsamen Inhalte effizient sind. Trotzdem ist vergleichende Evaluation lohnend; denn wenn wir einen *deutlichen* Unterschied zwischen den Gruppen feststellen und Engagement bei Lehrern und Schülern in beiden Fällen vorhanden ist, können wir einigermaßen sicher sein, daß der Unterschied auf die Curriculuminhalte zurückzuführen ist. Die Sequenz der Darbietung, die Methoden, die Schwierigkeiten, die Beispiele usw. können sicherlich genügend variiert werden, so daß nicht unterscheidbare Resultate unwahrscheinlich sind.

Es ist nicht sehr schwierig, entsprechendes Engagement in den Gruppen zu erreichen. In Analogie zu den scharf kalkulierten Vergleichsgruppen der »neuen Therapie«, bei denen die Therapieverfahren in ein oder zwei Tagen freier Assoziation durch ein »Brainstorming« der Wissenschaftler entstehen, konstruieren wir auf folgende Weise einige »neue Curricula«:

Zuerst holen wir uns zwei intelligente Studenten höherer Semester, z. B. aus den Wirtschaftswissenschaften, geben ihnen eine Liste mit wirtschaftswissenschaftlichen Fachausdrücken für die zehnte Klasse und zahlen ihnen 500 Dollar für die Übersetzung eines Kapitels aus Samuelsons Wirtschaftslehre in die Sprache der zehnten Klasse. Wir ermutigen sie, ihre Originalität zu zeigen und neue Ideen zu entwickeln. Sie können den Text wahrscheinlich in einem Sommer bearbeiten; so haben wir für einige tausend Dollar, einschließlich der Kosten für die Reproduktion des Versuchs-

materials, ein Curriculum, das wir einem der teuren wirtschaftswissenschaftlichen Curricula entgegensetzen können, die mit großer finanzieller Unterstützung auf kostspieligen Felduntersuchungen aufbauen. Dann suchen wir einige intelligente jüngere Studenten aus verschiedenen Hochschulen, die Wirtschaftswissenschaften studieren. Sie haben zu diesem Zeitpunkt Erfahrungen beim Absolvieren von Einführungskursen in Wirtschaftswissenschaften gesammelt und ein Problembewußtsein in bezug auf das Begriffsverständnis in diesem Bereich erworben. Wir geben ihnen einen Sommerzeit für die Entwicklung eines Curriculum zur Einführung in die Wirtschaftswissenschaften für die zehnte Klasse, das nicht einen bestimmten Text in den Mittelpunkt rücken soll.

Für eine dritte Vergleichsgruppe suchen wir einige Lehrer aus, die eine hohe Meinung von einem in den Sekundarschulen verwendeten Text für »Wirtschaftswissenschaften« haben. Dann lassen wir sie zusammen mit den Autoren eine Revision erarbeiten und dabei einige Beispiele der Reaktionen ihrer Kollegen auf den im Unterricht benutzten Text berücksichtigen. Da wir vor allem Curriculumentwickler unterrichten lassen, setzen wir sie in nur grob parallelisierten Vergleichsgruppen in Schulsystemen ein, die geographisch weit von denen entfernt sind, in denen wir die teuren Curricula testen. Wir setzen, um ihre Initiative zu wecken, eine vorher angekündigte Geldprämie für diese Gruppe aus, wenn sie nicht von dem großen Curriculum signifikant übertroffen wird. Wenn wir *trotzdem* einen erheblichen Unterschied zugunsten des großen Curriculum bekommen, können wir zu Recht annehmen, daß wir die Engagement-Variable beachtet haben. Darüber hinaus brauchen wir dies nicht für jedes Fach durchzuführen, da Engagement unabhängig vom Fachbereich in seinen Auswirkungen ziemlich konstant ist. Auf jeden Fall sollte eine kleinere Stichprobe genügen, um dies zu überprüfen.

Diese Art der vergleichenden Untersuchung hat einen besonderen Vorzug. Selbst wenn wir nur unbedeutende Unterschiede und damit ein mehrdeutiges Resultat erhalten, das uns darüber in Zweifel geraten läßt, ob ein allgemeines Engagement dafür verantwortlich ist oder ob alle Curricula der Wirtschaftswissenschaften ungefähr gleichen Unterricht bewirken, ersparen wir uns große Unkosten. Wenn wir mit wenig Kapital neue Curricula entwickeln können, die gute Ergebnisse erbringen, so ist das um so besser. Wir können das häufiger praktizieren und uns dadurch die Unterstützung engagierter Projektleiter erhalten und die Chancen vergrößern, einen Newton der Curriculumreform zu finden, der einen grundlegend neuen Ansatz entdeckt.

Weiterhin können wir, ohne daß neue Kosten entstehen, selbst im Fall einer Verbindung zwischen den verschiedenen Curricula die Engagement-

Frage recht schnell lösen, indem wir die Curricula einigen *negativ* und einigen *neutral* eingestellten Lehrern für den Unterricht während des nächsten Jahres oder der nächsten zwei Jahre geben. Andererseits bilden die von Anfang an beteiligten Curriculumentwickler eine Gruppe gut ausgesuchter, innovationswilliger Lehrer sorgfältig für die gleiche Arbeit aus. Vergleiche zwischen der Leistung dieser drei neuen Gruppen und der Leistung der alten sollten es uns ermöglichen, die Rolle des Engagements und zusätzlich die Unabhängigkeit der verschiedenen Curricula gegenüber fehlendem Engagement, die unzweifelhaft eine Variable darstellt, recht genau zu erfassen.

Offensichtlich müssen einige der oben dargestellten Verfahrensweisen in einer tatsächlich durchgeführten Untersuchung erweitert werden, z. B. die Möglichkeit für die neuen Curriculumentwickler, einige Nachmittage auf die Felduntersuchung der ersten Abschnitte ihres neuen Curriculum zu verwenden, um ihnen ein »Gefühl« für die Schnelligkeit zu vermitteln, mit der Schüler dieser Altersstufe die neuen Begriffe aufnehmen können, und um die Lehrer in bezug auf ihre konservative Einstellung, ihre Abneigung oder Lethargie mit Hilfe von Selbsteinschätzung und Einschätzung von Kollegen in Verbindung mit Einstellungsskalen sorgfältig auszusuchen.

Die »Schwierigkeit« mit der Engagement-Variablen ist ein Beispiel für die Auswirkungen der Versuchssituation. Andere Beispiele sind der Placebo-Effekt in der Medizin und der Hawthorne-Effekt in der Betriebs- und Sozialpsychologie. In jedem dieser Fälle sind wir daran interessiert, die Wirkungen eines bestimmten Faktors festzustellen, aber wir können den Faktor nicht in die experimentelle Situation einführen, ohne eine Störung hervorzurufen, die ihrerseits für die beobachteten Veränderungen verantwortlich sein kann. Im medizinischen Bereich besteht die Störung darin, dem Patienten etwas zu geben, was er für ein Medikament hält. Weil das für ihn kein gewöhnlicher Vorgang ist, kann dieser, ganz abgesehen von den intrinsischen Wirkungen des Medikaments, eigene Wirkungen hervorrufen. Beim Hawthorne-Effekt besteht die Störung beispielsweise in der Änderung von Arbeitsbedingungen, die den Arbeiter möglicherweise vermuten läßt, daß er Gegenstand einer speziellen Untersuchung und eines speziellen Interesses ist, und *dies* mag mehr zu einem verbesserten Arbeitsergebnis führen als die physikalischen Änderungen in der Umgebung, die die zu untersuchenden Kontrollvariablen darstellen. Die bisher erwähnten Fälle sind solche, bei denen die Überzeugung der Versuchspersonen der intervenierende Faktor zwischen Störung und mehrdeutiger Wirkung ist. Dies ist im Bereich der Psychologie charakteristisch, aber die Situation ist nicht grundlegend verschieden von der, die in der naturwissenschaftlichen Forschung auftaucht. Dort treffen wir auf Probleme wie die Absorp-

tion von Hitze durch ein Thermometer, wodurch die Temperatur, die gemessen werden soll, geändert wird. Das heißt, einige der beobachteten Wirkungen sind auf die Tatsache zurückzuführen, daß man das, was man messen will, ändern muß, um überhaupt eine Messung zu erhalten. Der Meßprozeß bringt ein anderes physikalisches Objekt in die Nähe des gemessenen Objektes. Das Instrument selbst hat eine bestimmte Wärmekapazität – ein Faktor, an dessen Einfluß man nicht interessiert ist. Dennoch muß man seine Größe abschätzen, um das Gewünschte herausfinden zu können. Das optimale Doppelblindmodell ist nur bei bestimmten Gegebenheiten angemessen, und es ist nur eine von vielen Methoden, mit denen wir diese Wirkungen umgehen können. Cronbachs Annahme, daß die Unmöglichkeit einer Doppelblinduntersuchung in der Curriculumentwicklung vergleichende Evaluation nicht zuläßt, erscheint deshalb zu pessimistisch. Tatsächlich stimmt er der Wichtigkeit vergleichender Arbeit zu, soweit er Längsschnittuntersuchungen erörtert.

Die Schlußfolgerung erscheint zwingend, daß vergleichende Evaluation die richtige Methode für die Behandlung der Probleme der Evaluation ist.

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).