

# Die Evaluation von Schulversuchen : Vorschlag eines methodisch flexiblen Vorgehens

Autor(en): **Ramseier, Erich**

Objektyp: **Article**

Zeitschrift: **Bildungsforschung und Bildungspraxis : schweizerische Zeitschrift für Erziehungswissenschaft = Éducation et recherche : revue suisse des sciences de l'éducation = Educazione e ricerca : rivista svizzera di scienze dell'educazione**

Band (Jahr): **6 (1984)**

Heft 1

PDF erstellt am: **13.09.2024**

Persistenter Link: <https://doi.org/10.5169/seals-786393>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Die Evaluation von Schulversuchen : Vorschlag eines methodisch flexiblen Vorgehens

*Erich Ramseier*

*Die Evaluation von Schulversuchen muss sich auf die Funktion dieser Versuche für die Schulreform ausrichten und die unverzichtbaren Informationsbedürfnisse formativer und summativer Art erfüllen. Weder Evaluation als Handlungsforschung noch Evaluation als empirisch-analytische Ursachenbestimmung kann diesem Anspruch allein genügen. Es wird daher eine Evaluationsstrategie vorgeschlagen, die die ökologische Situation und die internen Prozesse der Versuchsschulen, beachtet und traditionelle Exaktheitskriterien hinterfragt. Im Zentrum stehen Einzelfallanalysen zu den einzelnen Versuchsschulen, die auf einem gemeinsamen Raster aufbauen. Dieser Raster soll den Systemcharakter und die Komplexität der Schule einfangen. Zu einzelnen Themen sind empirisch-analytische Vergleiche mit dem Regelschulsystem notwendig, die dessen Variabilität systematisch erfassen. Damit wird gleichzeitig ein Beitrag zur Problemanalyse des Regelschulsystems geleistet.*

## **1. Einleitung**

Die folgenden Ausführungen zielen darauf ab, eine geeignete Form der Evaluation von Schulversuchen zu finden. Die Überlegungen beziehen sich auf ein mögliches, regionales Versuchsprogramm im Rahmen einer Reform der Sekundarstufe I, die in Richtung einer Integration der Schultypen und einer Individualisierung des Unterrichts zielt. Sie lassen sich in dem Masse auf andere Schulversuche übertragen, als die zu Beginn von Abschnitt 4 formulierten Bedingungen auf diese zutreffen.

Die Evaluation muss entsprechend den Bedürfnissen, die sie abzudecken hat, konzipiert werden. In Abschnitt 2 wird deshalb diskutiert, welche Funktionen Schulversuche und deren Evaluation für die Reform des Bildungswesens haben. In Abschnitt 3 werden zwei bekannte Evaluationsstrategien kritisch diskutiert. Dies dient als Grundlage zum Vorschlag einer «idiographisch-systematischen» Evaluation (Abschnitt 4).

## **2. Schulversuche und ihre Funktion für die Schulreform**

Wer von der Dringlichkeit bestimmter Verbesserungen im Schulsystem unserer Zeit überzeugt ist, ist immer wieder vom hartnäckigen Widerstand gegen Neuerungen und vom Beharrungsvermögen unserer Schulen überwältigt. Entgegen diesem Eindruck befindet sich das Bildungssystem in vielleicht langsamer, jedoch anhaltender Wandlung. Um sich davon zu überzeugen, genügt es, einen längeren Zeitabschnitt zu betrachten und sich etwa die Beschreibung der öffent-

lichen Schule zu Gemüte zu führen, die Gotthelf (1838) in seinem « Schulmeister » gegeben hat. Die meisten dieser Wandlungen vollziehen sich, indem die Schule allmählich auf Verschiebungen in den gesamtgesellschaftlichen Wertvorstellungen und den materiellen Voraussetzungen reagiert oder für ein bestimmtes, neu aufkommendes Bedürfnis ein zusätzliches Angebot einführt. Im Gegensatz dazu soll hier von Reformen die Rede sein, bei denen versucht wird, geplant eine Neustrukturierung und Verbesserung des vorhandenen Schulsystems zu erreichen. Es geht somit um Innovationen als Änderungen in einem sozialen System, die von der Bewertung des Bestehenden ausgehen, auf einen höher bewerteten Zielzustand hinführen wollen und als absichtliche und planvolle Prozesse in einem bestimmten zeitlichen Rahmen angelegt sind (nach Teschner 1979, S. 300).

In einer solchen Reform, die zu recht oder unrecht als rational gesteuerter Prozess verstanden wird, haben Schulversuche und deren Evaluation hauptsächlich drei Funktionen: die Entwicklung des Neuen in einem kleineren Rahmen, also mit beschränktem Risiko, die Überprüfung der Neuerung und das Darstellen und Bekanntmachen des Neuen.

Die Überprüfung ist wohl die bekannteste Funktion. Ähnlich einem Experiment soll sich zeigen, ob eine Änderung die erwarteten Folgen hat. Die Ähnlichkeit mit Experimenten ist jedoch kleiner, als oft angenommen und in der Öffentlichkeit befürchtet wird. Jeder Schulversuch ist einmalig, in eine bestimmte historische Situation mit besonderen gesellschaftlichen und politischen Randbedingungen eingebettet. Ein ganzes Massnahmenpaket, dessen Tauglichkeit schon im voraus möglichst weitgehend abgesichert wird, wird eingeführt und den konkreten Gegebenheiten laufend angepasst. Die Auswirkungen in den verschiedensten Bereichen interessieren (z.B. fachliche Schulleistungen, soziale Kompetenz der Schüler, Schulklima, Kosten). Schulversuche werden nur so angesetzt und gestaltet, dass ein Scheitern wenig wahrscheinlich ist. Als Analogie bieten sich demnach eher technische Entwicklungsarbeiten an, z.B. die Erprobung neuer Automobile oder Medikamente, als Experimente: Die Entwicklung der Neuerung ist mindestens gleichwertig zu deren Überprüfung. Um diese Abgrenzung zu reinen Versuchen auch sprachlich zu dokumentieren, sollte statt von Schulversuchen und Versuchsschulen besser von Schulentwicklung und Pilot- oder Modellschulen die Rede sein.

Die Parallelsetzung von Schulversuchen und technischen Entwicklungen geht allerdings von einem technologischen Innovationsmodell aus, wonach Änderungen im Bildungswesen zentral entwickelt und dann über politische Entscheide und Verwaltungsmassnahmen «per Dekret» durchgesetzt werden können. Sofern sich aber Bildungsreform nicht nur auf organisatorische und strukturelle Änderungen beschränken, sondern auch im Unterricht wirksam werden soll, genügt eine solche Änderungsstrategie nicht (vgl. Gstettner und Seidl, 1977, S. 96). Wenn Änderungen im täglichen Umgang von Lehrern mit Schülern erreicht werden sollen, müssen die Lehrer die neuen Verhaltensweisen selbst zur Richtschnur ihres Verhaltens machen. Dies kann nur erreicht werden, indem sie

die betreffenden Einsichten in ihrer eigenen Praxis nachvollziehen und erarbeiten, sich also aktiv an der Entwicklung der Neuerung beteiligen. Im Falle der Reform der Sekundarstufe I zeigt sich diese Grenze eines technologischen Innovationsvorgehens bei der Frage der inneren Differenzierung: Eine Verschiebung des Zeitpunkts der äusseren Differenzierung kann zwar zentral entschieden werden. Die innere Differenzierung des Unterrichts, die notwendig ist, wenn ein solcher Aufschub nicht zu einer qualitativen Verschlechterung des Unterrichts und zu einer Benachteiligung von Schülern mit sehr günstigen bzw. ungünstigen Lernvoraussetzungen führen soll, kann nur über die intensive Partizipation der Lehrer erreicht werden.

Schulversuche sollen eine schulische Neuerung bekannt machen (dritte Funktion). Sie sollen demonstrieren, dass und wie etwas Neues möglich ist. Gerade wenn Entscheide breit gestützt durch Volk und Parlament zu fällen sind, hat weniger eine abschliessende Beurteilung des Versuchs Einfluss auf bildungspolitische Entscheide als vielmehr die laufende Berichterstattung. Regelmässige Informationen und besonders direkte Kontakte wichtiger Meinungsbildner machen den Versuch allmählich bekannt und vertraut, sie nehmen ihm den Charakter des Ungewissen und zum Teil Bedrohlichen.

Neben diesen quasi offiziellen Funktionen von Schulversuchen sind auch mögliche heimliche Funktionen zu bedenken. Zum Beispiel kann der Aufschub einer bereits genügend abgesicherten Neuerung im ganzen Schulsystem durch Versuche legitimiert oder politische Konflikte können umgangen werden.

Die Evaluation dient den gleichen Zielen wie der Schulversuch als Ganzes. Sie ist jener Teil der Innovation, in dem Informationen gesammelt und verfügbar gemacht werden. Eingeschränkter und anspruchsvoller formuliert ist es das Ziel der Evaluation, «einzelne Reformmassnahmen in den grösseren Kontext der erzieherischen Wirkung von Umweltvariablen zu stellen und auf verallgemeinerungsfähiger Grundlage Aussagen zu machen, wie sinnvoll bestimmte Reformmassnahmen sind» (Fend, 1977, S. 57). Der Schulversuch soll also nicht nur beschrieben, sondern in seiner Wirkung hinterfragt werden. Umstritten ist der Grad der Verallgemeinerungsfähigkeit. Im einen Extrem werden nomothetische Gesetzesaussagen angestrebt, im andern werden die Aussagen auf den Einzelfall beschränkt (vgl. Abschnitt 3). Wo Schulversuche nicht nur der Entwicklung einer einzelnen Schule, sondern auch dem ganzen Bildungssystem dienen sollen, wird man nicht darum herumkommen, in irgend einer Form verallgemeinernde Schlüsse zu ziehen. Für die Konzeption der Evaluation ist es wichtig, zu sehen, für welche unterschiedlichen Zwecke Informationen bereitzustellen sind. Im Vordergrund stehen fünf Aufgabenkreise.

1. Systematische Informationen sind für die laufende Entwicklung und Verbesserung der Pilotschulen selbst notwendig. Damit leistet die Evaluation ihren Beitrag dazu, dass in den Pilotschulen die Möglichkeiten der zu erprobenden Innovation überhaupt ausgeschöpft werden können (formative Evaluation).



2. Informationen über die Auswirkungen der Versuchsmassnahmen müssen den bildungspolitischen Entscheidungsinstanzen zur Verfügung gestellt werden. Damit leistet die Evaluation ihren Beitrag zur kritischen Prüfung der Neuerung auf dem Hintergrund formulierter Ziele (summative Evaluation).
3. Informationen über die Pilotschulen müssen laufend einer weiteren Öffentlichkeit zur Verfügung gestellt werden. Damit leistet die Evaluation ihren Beitrag zur Förderung eines kritischen Reformbewusstseins und schafft Grundlagen für eine mögliche Erneuerung des Regelschulsystems.
4. Die bisher erwähnten Informationsaufgaben sind nur erfüllbar, wenn versucht wird, bei Beobachtungen ihr Zustandekommen aus dem Zusammenspiel vielfältiger Ursachen zu klären. Auf diese Art lassen sich aber Einzelergebnisse (zum Beispiel zum Thema Schülerbeurteilung) gewinnen, die von allgemeiner Bedeutung ausserhalb der Pilotschulen sind. Diese Bedeutung kann auf der praktischen Ebene liegen; die Ergebnisse sollten dann im Regelschulsystem möglichst rasch und unabhängig von der umfassenden Reform, auf die die Versuche zielen, angewandt werden, z.B. vermittelt über die Lehrerfortbildung.
5. Wo Ergebnisse von theoretischem Interesse vorliegen, sollten sie Eingang in die wissenschaftliche Diskussion finden. Eine Teilnahme an der wissenschaftlichen Diskussion konfrontiert immer wieder mit qualitativen Anforderungen, deren Einhaltung für eine zuverlässige Evaluation notwendig ist.

### **3. Zwei wichtige Strategien der Evaluation**

#### *3.1 Empirisch-analytische Evaluation*

Evaluationen von Schulversuchen sehen je nach dem wissenschaftstheoretischen und methodologischen Standpunkt der Beteiligten (und vorgelagert natürlich je nach Auftrag und Randbedingungen) ganz unterschiedlich aus. Wohl die meisten und bekanntesten Evaluationen verstehen sich als empirisch-analytische Forschung, die Ursache-Wirkungs-Zusammenhänge aufklären will, wobei als interessierende Ursache die zu evaluierende schulische Neuerung auftritt. Solche Untersuchungen sind idealiter quasi-experimentell, mit vorher-nachher-Messungen in Versuchs- und Kontrollgruppen angelegt. Sie stützen sich auf das traditionelle Methodeninventar empirischer Sozialwissenschaften (Tests, Fragebogen, Statistik etc.). Zweifellos haben solche summative Evaluationen wichtige Ergebnisse gebracht, wie aus den Berichten von Fend (1982), Petri (1979), Teschner, Jochimsen & Stolz (1980) u.a. hervorgeht. Evaluationen dieser Art unterliegen jedoch so vielen Einschränkungen und stossen auf so viele Schwierigkeiten, dass sie nicht als die Evaluationsstrategie angesehen werden können. Dies wird durch die Nennung einiger Probleme im folgenden illustriert, ohne dass auf wissenschaftstheoretische Probleme eingegangen wird:

Empirisch-analytische Untersuchungen versuchen, auf formaler, zur Hauptsache statistischer Basis, die Verallgemeinerbarkeit ihrer Ergebnisse zu erreichen. Dazu müssen sie (a) relevante Populationen definieren, (b) konstante Interventionen

voraussetzen, die Wirkung der interessierenden Ursache von der anderer Ursachen durch (c) Konstanthaltung, (d) Einbezug oder (e) statistische Kontrolle trennen und (f) abhängige Variablen messen.

Die *Festlegung und empirische Erhebung der Kriterien* (f), nach denen Schulversuche beurteilt werden sollen, stellt, teilweise unabhängig vom Forschungsansatz, grosse Probleme. Zum Beispiel können zentrale Erziehungsziele wie die Bereitschaft zur Übernahme staatsbürgerlicher Verantwortung als Erwachsener aus zeitlichen Gründen und wegen ihrer Allgemeinheit nicht direkt in messbare Kriterien umgesetzt werden. Andere Kriterien, nach denen Schulsysteme verglichen werden sollen, decken sich nur mit den Zielen des einen Systems, so dass von vornherein Resultate zu erwarten sind, die zu dessen Gunsten ausfallen. Das Problem der Kriterien und der Messung stellt sich oft als Konflikt zwischen inhaltlicher Gültigkeit (Validität) und zuverlässiger Messbarkeit (Reliabilität) dar. Eine Gefahr dabei liegt darin, dass die Forschung dorthin kanalisiert wird, wo es gelungen ist, zu inhaltlich bedeutsamen Kriterien geeignete Messinstrumente zu entwickeln. Ein Uebergewicht einer solchen Teilfrage gefährdet dann die Validität der ganzen Evaluation (Teschner, 1981, S. 223).

Eine elementare Voraussetzung der empirisch-analytischen Methode ist, dass die eingeführten Behandlungen *zeitlich und zwischen den verschiedenen Einheiten konstant sind* (b). Ausser wenn die Evaluation Jahre nach Versuchsbeginn einsetzt, bedeutet die Forderung nach der zeitlichen Konstanz, dass im einzelnen Schulversuch aus den gemachten Erfahrungen nicht Verbesserungen abgeleitet werden dürfen. Dies würde das Gelingen der Innovation gefährden. Würde man andererseits tatsächlich solange mit der Evaluation warten, bis sich die Versuchsschule stabilisiert hat, würde man die ganze Information verlieren, die sich während dieser Entwicklung ergibt, und die bei einer Generalisierung der Innovation notwendig ist.

Um die Konstanz der Intervention über alle Einheiten hinweg zu wahren, müssen alle Versuchsschulen von der gleichen Art sein, ausser ihre Unterschiedlichkeit werde im Versuchsplan vorgesehen. Bei einer komplexen Intervention wie einer Strukturreform auf der Sekundarstufe I sind aber Adaptationen an die besonderen Verhältnisse jeder Schule und die Erprobung verschiedener Varianten erwünscht, da man sich nicht von vornherein begründet auf eine ganz spezifische Struktur festlegen kann. Damit die primär zur Diskussion stehende Strukturreform wirksam werden kann, sollte sie inhaltlich ergänzt werden (z.B. Schülerbeurteilung, innere Differenzierung). Die optimale Gestaltung dieser Ergänzungen muss noch gefunden werden. Auch sie sollten also von Pilotschule zu Pilotschule variieren dürfen. Diese notwendige Variabilität der Intervention ist ein erstes Element, das in einem analytischen Design nur mit unrealisierbarem Aufwand berücksichtigt werden könnte.

Bei der *Konstanthaltung konkurrierender Ursachen* (c) wird die Wirkung der Intervention nur unter einer bestimmten Ausprägung von Randbedingungen untersucht, z.B. nur in städtischen Verhältnissen. Dies schränkt natürlich die Verallgemeinerbarkeit ein. Je nach Art der konstantgehaltenen Variablen können

aber auch inhaltliche Verfälschungen auftreten. Petri (1981, S. 203) schlägt z.B. vor, «sich bei vergleichenden Untersuchungen der Lernerfolge in homogenen und heterogenen Schulklassen zunächst auf eine bestimmte Art des Unterrichts, wie etwa den 'herkömmlichen' Unterricht» zu beschränken. Dies kann für die Neuerung (heterogene Klassen) fatale Folgen haben, da eine notwendige inhaltliche Komponente der Neuerung, die Abstimmung der Art des Unterrichts auf die Heterogenität, unterbunden wird.

Die informativste Art der Berücksichtigung konkurrierender Variablen ist, sie *systematisch variierend in die Untersuchung einzubeziehen* (d). Die Vielzahl wichtiger konkurrierender Variablen ist jedoch ein zweites Element, das nur mit einem unrealistisch komplexen Untersuchungsplan angemessen bearbeitet werden könnte. Faktisch muss sich diese Art der Kontrolle auf wenige Störvariablen beschränken. Die Strategie, Wechselwirkungen mit solchen weiteren Variablen aufzuklären, stösst rasch auf prinzipielle Grenzen. Dies wird schon durch die Ergebnisse der ATI-Forschung belegt, obwohl dort meist bescheidenere Fragestellungen angegangen werden. Selbst Cronbach, der eigentliche Promotor dieses Forschungsansatzes, kommt im Rückblick auf ein gutes Jahrzehnt entsprechender Forschung zum Schluss, Untersuchungen nach diesem varianzanalytischen Paradigma sollten zugunsten der Berücksichtigung der idiosynkratischen Besonderheiten, des Kontextes und des Verlaufs der einzelnen Fälle zurückgestellt werden (Cronbach, 1975).

Die prinzipiell korrekte Lösung des Problems der Vielzahl von mitbeeinflussenden Variablen innerhalb des empirisch-analytischen Ansatzes liegt darin, durch Zufallszuordnung der Untersuchungseinheiten diese *Einflüsse statistisch zu kontrollieren* (e). Wo aber gerade bedeutsame Variablen wie Lehrereinfluss und Unterrichtsgestaltung so neutralisiert werden müssen, kann die Untersuchung nur noch eine sehr bescheidene Erklärungskraft haben. Es ist dann zwar möglich, zu Aussagen mit weitem Gültigkeitsbereich zu kommen. Diese Aussagen sind jedoch sehr schwach und erlauben nicht, z.B. für eine einzelne Schule eine bedeutsame Voraussage zu machen. Da in üblichen summativen Evaluationen auch alle zeitlichen Abläufe, vom Verlauf des konkreten Unterrichts bis zum Verlauf eines ganzen Versuchs, zu den bestenfalls statistisch kontrollierten Variablen gehören, liegen Zusammenhänge zwischen Faktoren und Effekten meist nur in Form von black-box-Modellen vor; man weiss also nicht, welche Prozesse zu den Effekten führen und kann folglich nicht optimal eingreifen.

In summativen Evaluationen zu Strukturen auf der Sekundarstufe I ist es unvermeidlich, dass die notwendigen Kontrollen der «Störvariablen» teilweise fehlschlagen. Durch die daraus folgenden Konfundierungen ist besonders die externe Validität, die Uebertragung auf das Regelschulsystem bedroht. Zu denken ist etwa an konkurrierende Faktoren wie die besonders zusammengesetzte, reformoffene Lehrerschaft, die unrepräsentative Schülerschaft bei freiwilligen Schulversuchen und die häufige Ausgestaltung von Gesamtschulen als Tagesschulen (vgl. Wottawa, 1982). Auch die Tatsache, dass Neuerungen erprobt werden und Neues noch wenig ausgereift ist, aber auch Schüler und Lehrer eher

zur Mitarbeit motiviert (Hawthorne-Effekt), bedroht die externe Validität. Eine weitere, unvermeidliche Bedrohung der Übertragbarkeit liegt darin, dass sich Schulversuche an den Berührungspunkten dem übrigen Schulsystem anpassen müssen auch wenn dadurch die eigenen Zielsetzungen tangiert werden. Bei Versuchen auf der Sekundarstufe I zeigt sich das deutlich an der Auflage, den Anschluss an die höheren Mittelschulen zu gewährleisten. Für die Wahl der Lehrstoffe und die Realisierung neuer Lehrziele sind für die gesamte obligatorische Schulzeit damit enge Grenzen gesetzt, da Schüler nicht in der Lage sind, beliebig viel neues Können ohne Abstriche am alten zu erwerben.

Die formale Generalisierung in der empirisch-analytischen Forschung geht grundsätzlich von der beliebigen Wiederholbarkeit der Intervention und von *gleichwertigen, austauschbaren Einheiten* aus, von denen eine Zufallsstichprobe näher betrachtet wird (a). Dieser Ansatz ist eminent ahistorisch und sieht von der konkreten Einbettung des Einzelfalles in seine Umgebung und von seiner individuellen Entwicklung ab bzw. behandelt Auswirkungen als bloße Fehlvarianz. Die Bedeutung der ökologischen Lage der einzelnen Schule und der schulinternen Prozesse (Rutter, Maughan, Mortimore & Ouston, 1979) liegt aber heute auf der Hand! Haenisch & Lukesch (1980, S. 260) ziehen z.B. aus zahlreichen Vergleichen von Schulsystemen das Fazit, für die Schülerleistung sei es wichtiger, welche konkrete Schule der Schüler besuche, als welchem Schulsystem diese zugehöre. Da grossangelegte empirisch-analytische Vergleiche von Schulsystem die so möglichen Erkenntnisse aufgezeigt haben, fordern auch deren Repräsentanten die vermehrte Beachtung schulspezifischer Variablen und den Einbezug von Fallstudien (Fend, 1982, s. 500f).

Die Quintessenz dieser Kritik der empirisch-analytischen Evaluation ist, dass sie von einer Forschungslogik ausgeht, die kausale Aussagen verspricht, dass aber in der Praxis der Evaluation bei allen Untersuchungsschritten so viele Einschränkungen und Fehler in Kauf genommen werden müssen, dass die in Aussicht gestellte Sicherheit und inhaltliche Differenziertheit der Aussagen nicht erreicht werden kann. Dies wird dadurch verschärft, dass die Evaluationen in einem stark politisch und sozial geprägten Raum agieren und *eine bloss ergänzende bzw. kontrollierende Funktion* für die Versuche haben: Die Versuche werden in aller Regel nicht als Intervention in einem geeigneten Untersuchungsdesign konzipiert, sondern die Evaluation hat sich den Versuchsgegebenheiten anzupassen. Bei der Übertragung des experimentellen Forschungsparadigmas auf Schulevaluation bleibt davon nicht viel mehr als eine bloße Denkfigur übrig. Diese Denkfigur hat zudem den prinzipiellen und oft kritisierten Mangel (z.B. Gröben & Scheele, 1977), in einem sozialen Handlungsfeld die Handelnden nur als Forschungsobjekte und nicht als Subjekte zu sehen. Gerade wenn man allen in Abschnitt 2 genannten Informationsbedürfnissen genügen will, ist es daher angebracht, sich nicht auf das formale analytisch-empirische Entscheidungsverfahren auszurichten, sondern eine «vernünftige Argumentation» einzusetzen, die die beschränkten Forschungsmöglichkeiten und besonders die Bedeutung der ökologischen Lage und der internen Prozesse der einzelnen Schulen bewusst



in Rechnung stellt. Eine solche «weiche» Konzeption ist vorzuziehen, obwohl sie keine konzise, allgemein anerkannte Lösung für die aufgeworfenen Probleme hat.

### *3.2 Evaluation als Handlungsforschung*

Evaluation als Handlungsforschung ist ein konsequenter Gegenpol zur Evaluation auf empirisch-analytischer Basis. Handlungsforschung will gemeinsam mit Praktikern, d.h. Lehrern, die Schule erneuern. «Das Selbstverständnis der Handlungsforschung orientiert sich an der für eine Veränderung von Alltagssituationen und Lebenswelten konstitutiven Bedingungen bzw. an deren Deutung durch die Betroffenen. Forschungssubjekte und Forschungsgegenstände werden von dieser Grundorientierung aus als Personen interpretiert, die durch eine gemeinsame Kultur verbunden sind und während des Forschungsprozesses in einer kommunikativen Beziehung zueinander stehen. Der Forschungsprozess ist demnach ein Kommunikationsprozess zwischen Personen in einem komplexen und dynamischen sozialen Feld, in dem die Selbstreflexion aller Beteiligten auf dem Hintergrund der dem Forschungsprozess zugrundeliegenden gesellschaftlichen Sinndeutungen zur Sprache kommt» (Gstettner & Seidl, 1977, S. 87). Schulreform kommt nach dieser Vorstellung dadurch zustande, dass Praktiker zusammen mit Wissenschaftlern ihr pädagogisches Handeln hinterfragen und zu einer neuen Konzeption ihrer Tätigkeit gelangen.

Problematisch an diesem Ansatz ist die weitgehende Beschränkung auf Entwicklungsarbeiten und der Verzicht auf summative Evaluation mit ihren traditionellen Hilfsmitteln zur Sicherung der Intersubjektivität. Durch ihre Identifikation mit der Entwicklung unterliegen die Forscher in spezifischer Weise den in der Sozialpsychologie bekannten Determinanten der Urteilsbildung und produzieren Daten mit fraglicher Glaubwürdigkeit (Hubermann, 1983). Die Verallgemeinerbarkeit der Ergebnisse bleibt unklar. Sie wird von dem Forscher zwar gar nicht explizit angestrebt, von den Abnehmern jedoch selektiv postuliert. Eine weitere Gefahr liegt darin, dass die Forscher ganz unter den Druck der schulpraktischen Arbeiten geraten, ihre Tätigkeit weitgehend mit der des Lehrers zusammenfällt und die beiden nicht mehr ihren Qualifikationen entsprechend arbeitsteilige Beiträge leisten. Dadurch ist eine genügende Berichterstattung über die Projektarbeit gefährdet.

## **4. Ein Vorschlag zur Evaluation von Schulversuchen**

Der folgende Vorschlag zur Evaluation von Schulversuchen beruht auf Überlegungen im Zusammenhang mit Reformen, die in Richtung einer Integration der Schultypen und einer Individualisierung des Unterrichts auf der Sekundarstufe I gehen. Sie lassen sich auf thematisch anders ausgerichtete Versuche übertragen, sofern Gemeinsamkeiten mit der nachfolgenden Konzeption des Versuchsprogramms gegeben sind:

- Es werden ganze Schulen in die Versuche einbezogen.
- Es werden nicht nur organisatorische Änderungen angestrebt, sondern



- Reformen, die sich auch direkt im Unterricht auswirken.
- Das Programm umfasst eine beschränkte Zahl von Versuchsschulen (etwa 2-8).
  - Welche Schulen wann in den Versuch eintreten, zeigt sich erst im Verlaufe des Versuchsprogramms.
  - Die einzelnen Schulen haben eine beträchtliche Autonomie. Die je getesteten Strukturen fallen entsprechend den örtlichen Verhältnissen unterschiedlich aus.
  - Über die Unterschiede in den Versuchselementen hinaus streuen die Schulen in wichtigen Variablen stark (z.B. Schulgrösse, Stadt-Land usw.).

Dieses Versuchsprogramm ergibt sich einerseits aus realpolitischen Erwägungen (z.B. gestaffelte Einführung). Andererseits soll mit der angestrebten Autonomie und Diversität das Engagement der Beteiligten gewonnen werden, verschiedene Varianten und die Bewährung in unterschiedlichen Verhältnissen untersucht werden, und dank einer geeigneten lokalen Streuung entstehen Kristallisationspunkte der Reform in verschiedenen Regionen.

Die Evaluation einer solchen Versuchsreihe hat sich immer allen fünf in Abschnitt 2 genannten Informationsbedürfnissen zu stellen. Weder ein rein summativer, empirisch-analytischer noch ein rein handlungsorientierter Forschungsansatz können demnach den Anforderungen genügen. Ergiebiger kann ein «idiosynkratisch-systematisches» Vorgehen sein. Darin wären mehrere, methodisch teilweise stark divergierende Teiluntersuchungen zu kombinieren. Der Gesamtzusammenhang der Evaluation wird nicht durch ein Design gestiftet, das über ein standardisiertes Entscheidungsverfahren zu einer Gesamtbeurteilung führen soll, sondern durch eine Konzeption, die es erlaubt, die Teilelemente in vernünftiger Argumentation und aufgrund der verfügbaren allgemeinen Kenntnisse zur Versuchsthematik aufeinander zu beziehen.

Ein zentrales Element eines solchen Evaluationsansatzes sind Fallstudien der einzelnen Versuchsschulen (idiographisches Vorgehen). Solche Einzelfallanalysen sind wegen der Bedeutung der spezifischen ökologischen Situation und der Eigendynamik jeder einzelnen Schule und wegen den je unterschiedlichen Realisierungen der Versuchselemente notwendig. Sie ergeben Informationen, die der einzelnen Schule direkt bei ihrer Weiterentwicklung von Nutzen sind. Entscheidend ist, dass diese Fallstudien nicht voneinander isoliert, sondern systematisch aufeinander abgestimmt sind. Ihnen muss ein gemeinsamer Raster unterliegen und sie sollen durch übergreifende Abklärungen zu einzelnen Fragen ergänzt werden (systematischer Bezug). Da es bei der beschränkten Zahl von Versuchsschulen immer zu Konfundierungen zwischen schulspezifischen Versuchselementen und ökologischen Aspekten kommt, erlaubt nur die differenzierte und koordinierte Beschreibung, die Auswirkungen einzelner Komponenten plausibel zu klären. In den Einzelfallanalysen wird die Interpretation des schulischen Geschehens durch die Beteiligten eine wichtige Stelle einnehmen, es gehören aber auch die Erhebung äusserer Schulmerkmale und Beobachtungen dazu. Bei diesen Analysen können die in Entwicklung begriffenen Methoden der

Fallstudien (Fischer, 1982) und der systematischen Behandlung qualitativer Daten angewandt werden (Huberman & Miles, 1983, Miles & Huberman, 1984). Als Beispiel einer nützlichen Strategie sei die «Triangulation» erwähnt (Huberman, 1983, S. 48). Dabei werden mehrere voneinander unabhängige Methoden, von denen jede ihr eigenes Fehlerrisiko hat, eingesetzt, um wichtige Aussagen abzusichern.

Wenn die Gesamtanlage der Evaluation nicht einem empirisch-analytischen Design folgt, so sollen dennoch einzelne Fragestellungen mit solchen Forschungsmethoden abgeklärt werden. Wie schon früher festgestellt (z.B. Perret, 1981), ist nicht die Wahl zwischen globalen Forschungsstrategien, sondern die differenzierte Betrachtung dieser Ansätze interessant. Auf verschiedenen Ebenen (Design, Intervention, Datenerhebung, Datenreduktion, Interpretation) sollen, soweit dies sinnvoll ist, unabhängig voneinander Elemente der divergierenden, nicht formalen Gesamtargumentation zusammengeführt werden. Zur Abklärung einzelner, beschränkter Fragestellungen, abgeleitet von gemeinsamen Zielen oder allgemein befürchteter Nebenwirkungen, sollen die Vorzüge quantitativer und quasi-experimenteller Methoden (Bemühung um klare Operationalisierung, Verwendung geprüfter Erhebungsinstrumente, Sicherungen gegen die Ueberschätzung einzelner Unterschiede und gegen subjektive Einflussnahme) genutzt werden. Solche Erhebungen geben das Gerüst, in dem die Einzelfallstudien verankert werden können. Umgekehrt liefern die Einzelfallstudien jene Informationen, die notwendig sind, um eine plausible Wahl zwischen den vielen alternativen Erklärungsmöglichkeiten treffen zu können und ein differenziertes Gesamtbild zu gewinnen.

Schulversuche sollen der Verbesserung des Regelschulsystems dienen. Die Evaluation kann daher nicht nur belegen, dass eine Änderung möglich ist, sie muss auch mit Schulen des Regelsystems vergleichen - trotz aller Vorbehalte bezüglich der Gültigkeit solcher Aussagen. Wenn eine Vergleichsuntersuchung schon unumgänglich ist, so lohnt es sich, sie gleich breit anzulegen und wahrscheinliche Variationen im Regelschulsystem mitzuerfassen. Soweit dabei schriftliche Befragungen, Schulleistungstests usw. eingesetzt werden können, ist der Aufwand bei der Vergrößerung des Umfangs der Untersuchung relativ bescheiden. Wird eine solche Erhebung noch mit einer standardisierten Beschreibung der inneren Struktur und der ökologischen Lage der einbezogenen Schulen ergänzt, so lassen sich die Versuchsergebnisse trotz der beschränkten Zahl von Versuchsschulen sinnvoll dem Regelschulsystem, dessen üblichen Resultaten und dessen Unterschiedlichkeit, gegenüberstellen. Vor allem trägt eine solche Untersuchung fast beiläufig zur Problemanalyse des Regelschulsystems bei. Neben der Versuchsevaluation im engeren Sinne wird mit einer solchen Problemanalyse der zweite wesentliche Beitrag geleistet, den die Wissenschaft zur Reform der Schule beisteuern kann.

Wenn bei der vorgeschlagenen Evaluation mehrere Einzelfallstudien durchgeführt werden sollen, wenn Teilfragen mit verschiedenen Methoden bearbeitet werden und wenn aus diesen Teilen in einer Gesamtargumentation verallgemei-

nernde Aussagen angestrebt werden, so sollten all diese Einzelschritte auf einem gemeinsamen theoretischen Konzept aufbauen können. Ein Analyserahmen, der die notwendige Weite hat, ist am ehesten mit der Betrachtung der Schule als System gegeben. Heute liegen im Rahmen der Schultheorie einige einschlägige Konzeptionen vor (z.B. Isenegger, 1977; Fend, 1980). Aus diesen muss vorbereitend und begleitend zur empirischen Evaluation ein auf die besonderen Verhältnisse bezogenes theoretisches Konzept entwickelt werden. Der folgende Exkurs soll die Thematik und Breite eines solchen Ansatzes illustrieren. Er stellt noch nicht den erwünschten Analyserahmen dar.

### ***Exkurs: Schule als komplexes soziales System***

Die Schule ist ein Subsystem der Gesellschaft, das an der für das Gesamtsystem entscheidenden Tradierung der Kultur massgeblich beteiligt ist und die spezifischen Funktionen der Qualifikation, Selektion und Integration / Legitimation übernimmt. Die Schule veranstaltet gesellschaftlich kontrollierte, breit angelegte Sozialisation. Diese hat immer ein Doppelgesicht. Sie ist einerseits auf die Gesellschaft, andererseits auf das zu sozialisierende Individuum ausgerichtet. Zwischen diesen Polen kann es immer wieder zu Konflikten kommen. Versuchsziele und Auseinandersetzungen um Schulversuche sind auf dem Hintergrund dieser gesellschaftlichen Funktionen zu sehen. Die Schule ist ihrerseits ein komplex gegliedertes System. Sie soll einer grossen Zahl von Schülern eine vergleichbare Bildung vermitteln. Vieles an ihrer eigenen Struktur ist denn auch auf Stabilität, Planbarkeit und Kontrollierbarkeit ausgerichtet, z.B. die äussere Gliederung in Schultypen und Jahrgangsstufen, die Lehrpläne und Stundentafeln. Zur Schule gehören mehrere ergänzende Systeme wie Verwaltung, Inspektorat, Schulkommission, Beratungsdienste, Lehrerbildung, Lehrmittelproduktion.

Auch die einzelne Schule ist als soziales System mit definierten Rollen wie Schüler, Lehrer, Leiter, Abwart zu verstehen. Sie ist in soziale Einheiten wie Klassen und Lehrerkonferenz gegliedert und durch eine Vielzahl von Aspekten (z.B. Grösse, Grad innerer und äusserer Autonomie, Schulklima, Infrastruktur) gekennzeichnet. Wichtig ist ihre Einbettung in die örtliche, soziale und geographische Umgebung, die durch lokal dominierende Wertungen und Einstellungen, Bevölkerungsstruktur, Einzugsbereiche, Wegverhältnisse, Infrastruktur der Gemeinde / Region, lokale Wirtschaftsstruktur usw. gekennzeichnet ist. All dies setzt die Randbedingungen, unter denen üblicherweise ein Lehrer mit rund 13-35 Schülern studenweise in einen Raum gewiesen wird mit dem Auftrag, einen in groben Zügen vorgegebenen Stoff zu vermitteln, Lernprozesse anzuregen und den Lernerfolg der Schüler zu prüfen. Solcher Unterricht ist ein recht genau bestimmter Verhaltensrahmen, der zusätzlich zu den offiziellen Randbedingungen durch (vermutete) Erwartungen der Eltern, Lehrerkollegen, Behörden und durch nicht hinterfragte Gewohnheiten, z.B. aus der eigenen Erfahrung des Lehrers als Schüler, bestimmt ist.

Die individuellen Lernprozesse werden durch den Unterricht geprägt, sie sind dadurch jedoch keineswegs festgelegt. Individuelle kognitive und motivationale Voraussetzungen, die Struktur des Lehrgegenstandes und vielfältige, variable Randbedingungen wie der Verlauf der vorherigen Lektion, bevorstehende Prüfungen, aber auch die Raumtemperatur usw. haben ebenfalls bestimmenden Einfluss darauf. Manche Aktivitäten im Unterricht sind zudem nicht auf Lernziele ausgerichtet; ein Teil von ihnen wird als Störung wahrgenommen. Gelernt werden nicht nur fachliche Ziele, sondern teils gewollt, teils «heimlich» (Zinnecker, 1976) allgemeinere Haltungen wie Konzentration, Zurückstellen eigener Bedürfnisse, Akzeptieren von Erfahrungen aus dritter Hand, Erfüllen fremdbestimmter Aufgaben, Einordnung in soziale Gruppen, Konkurrenz um knappe gute Noten und Plätze in anspruchsvollen Bildungsgängen, Leistung als Wert an sich.

Auch wenn Unterricht die zentrale schulische Tätigkeit ist, sollte die Schule nicht nur unter diesem Aspekt betrachtet werden. Einmal ist das «Schulleben» ausserhalb des Unterrichts für manche Lernprozesse wesentlich. Die Schule hat auch weitere Funktionen (z.B. Hortfunktion), die dann wahrgenommen und wesentlich werden, wenn sie durch eine Änderung in Frage gestellt werden. Schule kann auch weniger unter einem pädagogisch-didaktischen Gesichtspunkt als vielmehr als Organisation oder als Arbeitsplatz für Lehrer und Schüler betrachtet werden.

Der Exkurs demonstriert trotz seiner Unvollständigkeit (z.B. historische Dimension) und Unsystematik, die Komplexität des Phänomens «Schule». Schulversuche sind Interventionen in diesem Feld. Ein theoretischer Ansatz, der diese Komplexität berücksichtigt, ermöglicht es, Versuchselemente im Gesamtzusammenhang zu sehen, Hypothesen über ihre gewollten und ungewollten Auswirkungen aufzustellen, entsprechende Untersuchungsfragen zu formulieren und Teiluntersuchungen zu kombinieren. Er relativiert vor allem auch die Erwartungen an die Wirksamkeit einzelner Versuchsmassnahmen in der Schule, deren übrige Bedingungen konstant bleiben.

Die vorgeschlagene Art der Evaluation hat sowohl formative als auch summative Aspekte; sie soll die Entwicklung der einzelnen Pilotschulen fördern und Grundlagen zur Beurteilung liefern. Damit sie diese Doppelfunktion, die z.B. Fend (1977, S. 57) für ungünstig hält, wahrnehmen kann, müssen Abgrenzungen vor allem gegenüber den Entwicklungsaufgaben getroffen werden. Die Evaluation soll die Entwicklung und Verbesserung der Pilotschule unterstützen, indem sie beobachtet, den Ist-Zustand erhebt und mit Zielsetzungen vergleicht. Wo Mängel festgestellt werden, können «Evaluatoren» wohl beratend an ihrer Behebung mitwirken; die eigentliche Entwicklungsarbeit verlangt aber in der Regel andere Kompetenzen als die Auswertung. Sie sollten in erster Linie von den beteiligten Lehrern, dann von einem ständigen pädagogischen Berater, der z.B. die Lehrerfortbildung zu organisieren hat, und von je nach Gegenstand zugezogenen Fachleuten geleistet werden. Der Beitrag, den die Evaluation durch gezielte Rückmeldungen an die Lehrer an die Entwicklung leisten kann, darf dabei nicht unterschätzt werden.



Die Evaluation ist darauf angewiesen, möglichst unverfälschte Informationen aus den Pilotschulen zu erhalten. Sie braucht dazu das Vertrauen der Lehrer. Ein solches Vertrauen ist nur dann möglich, wenn die Auswertung vollständig von Aufsichtsfunktionen bezüglich des Schulbetriebs getrennt ist, so dass die Lehrer nie befürchten müssen, dass allfällige Probleme, die sich aus der Versuchsarbeit ergeben können, ihnen als persönliches Versagen ausgelegt und für sie entsprechende Konsequenzen haben könnten.

Das Vertrauen der an einer Pilotschule beteiligten Lehrer in die Auswertung setzt wohl auch voraus, dass die Lehrer vom guten Willen der Auswertenden ausgehen können, d.h. davon, dass « Aktive » und « Beobachter » im gleichen Boot sitzen und ein Interesse daran haben, dass der Versuch zum Erfolg wird. Es fragt sich damit aber, ob die Auswertenden überhaupt noch die Aufgabe der unverfälschten Berichterstattung gegenüber der Öffentlichkeit wahrnehmen können. Tatsächlich kann die Berichterstattung nicht wertfrei sein; sie ist der Verbesserung des Bildungswesens verpflichtet. Sie muss aber einerseits so ausführlich sein, dass der gründliche Leser seine eigenen Wertmassstäbe ansetzen kann, und andererseits klarmachen, an welchen Werten und Zielen sie sich orientiert. Die Glaubwürdigkeit der Evaluation kann auch durch eine Organisationsform gestützt werden, die ein geeignetes Mass der Verwaltungs- und Schulunabhängigkeit demonstriert. Geeignet kann z.B. ein die Evaluation leitendes Organ sein, in dem Lehrerschaft, Wissenschaft und Verwaltung gleichberechtigt vertreten sind. Solange die Evaluation einerseits tiefen Einblick in die Versuchsschule haben muss und andererseits möglichst objektiv gegenüber politischen Entscheidungsträgern informieren muss, ist der Konflikt zwischen Schule, Evaluation und politischer Instanz vorprogrammiert. Es gilt nicht, ihn zu unterdrücken, sondern ihn in geeigneter Form als « kontrollierten Konflikt » (Teschner, 1981, S. 230), z.B. im genannten Gremium, auszutragen.

Die vorgeschlagene « idiographisch-systematische » Evaluation erweckt sicher den Eindruck eines Wunschprogramms: Wer hätte nicht gerne neben Vergleichsuntersuchungen auf Test- und Befragungsebene auch mehrere Einzelfallstudien gemacht, Beobachtungen ausgewertet und die Komplexität des Untersuchungsfeldes berücksichtigt! Tatsächlich wird die verfügbare Arbeitskapazität jeder Evaluation enge Grenzen setzen; eine selektive Problembearbeitung ist notwendig. Der Vorschlag zielt jedoch darauf, dass die Konsequenzen aus dieser Begrenzung neu zu ziehen sind: Wenn es für die Wirksamkeit der Schulversuche notwendig ist, fünf Informationsbedürfnisse (Abschnitt 2) zu decken, so nützt es nichts, nur im Hinblick auf eines davon mit wissenschaftlicher Akribie zu arbeiten und die andern beiseite zu lassen. Angewandte Forschung (und nicht nur sie) kann es sich nicht leisten, eine Teilfrage präzise zu beantworten und dabei den Gesamtzusammenhang aus den Augen zu verlieren. Zugunsten der Breite der Untersuchung sollen Einzelfragen also nur soweit untersucht werden, als dies praktische Konsequenzen hat und für die Glaubwürdigkeit der Untersuchung notwendig ist. Es gilt ja nicht, Hypothesen zu falsifizieren, sondern sie aufgrund bewusster Erfahrungen neu zu formulieren. Neben dieser grundsätzli-



chen Prioritätensetzung relativiert auch die Konzeption der Evaluation als Kombination abgestimmter, jedoch weitgehend selbständiger Teiluntersuchungen die Frage des Aufwands: Eine solche Organisation erlaubt es, in einem Masse universitäre Abschlussarbeiten einzuordnen, wie dies bei einer einheitlichen Untersuchungskonzeption nicht so leicht möglich wäre. Wo die Arbeitskapazität dennoch nicht ausreicht, sollte die Evaluation von allen Beteiligten gemeinsam eingegrenzt werden, damit die notwendigen Beschränkungen auch allgemein verstanden und verantwortet werden. Wenn nötig muss die Zahl der Versuchsschulen reduziert werden, um eine ausreichende Evaluation zu ermöglichen und vor allem, um durch ausreichende Begleitung ein Scheitern einzelner Schulen zu verhindern.

Eine auf Einzelfälle und ihren Kontext bezogene Evaluation ist nicht neu. Eine Untersuchung über kooperative Gesamtschulen von Tillmann, Bussigel, Philipp & Rösner (1979) realisiert diese Vorstellungen unter besonderer Betonung der Verlaufsanalyse weitgehend und auch die Begleitung der zürcherischen abteilungsübergreifenden Versuche an der Volksschuloberstufe (Oertel, 1981) und, auf summative Fragestellungen beschränkt, die Auswertung des Schulversuchs Manuel (Ramseier, 1982) sind ähnlich konzipiert.

Eine «idiographisch-systematische» Evaluation hat einige Vorteile für sich. Zum einen ist sie sehr flexibel. Sie geht zwar von einem Gesamtrahmen aus, ohne aber von Anfang an ein einheitliches Programm vorzulegen. Sie kann daher auf die akuten Informationsbedürfnisse in den Pilotschulen und in der interessierten Öffentlichkeit reagieren. Die Evaluation liefert auch Zwischenergebnisse, die der frühzeitigen Einführung einzelner Versuchselemente in weiteren Schulen dienlich sein können. Nicht unwichtig ist, dass eine solche Evaluation, die detaillierte Informationen über einzelne Versuchsteile, über die Prozesse in den Versuchsschulen und über das Regelschulsystem beschafft, auch dann wertvoll ist, wenn die Versuche später nicht ins Regelschulsystem übergeführt werden sollten. Wenn in der Evaluation verschiedene Teiluntersuchungen und Einzelfälle auf dem Hintergrund eines Analyserahmens zueinander in Beziehung gebracht werden, sind schliesslich günstige Voraussetzungen gegeben, um in ähnlicher Weise auswärtige Ergebnisse einzubeziehen und aktiv eine Koordination anzustreben. Die vorgeschlagene Evaluationsstrategie könnte so eine Basis für die Bildung von Projektverbänden (Petri, 1981, S. 212) werden.

## LITERATURVERZEICHNIS

- Cronbach, L. J.* Beyond the two disciplines of Scientific Psychology. In: American Psychologist, 1975, 30, 116-127.
- Fend, H.* Wissenschaftssoziologische Perspektiven für eine Analyse der Begleitforschung im Rahmen von Modellversuchen im Bildungswesen. In: Mitter W. & Weishaupt H.: Ansätze zur Analyse der wissenschaftlichen Begleitung bildungspolitischer Innovationen, Weinheim, 1977.
- Fend, H.* Theorie der Schule. München, 1980.

- Fend, H.* Gesamtschule im Vergleich. Bilanz der Ergebnisse des Gesamtschulversuchs. Weinheim, 1982.
- Fischer, D.* Fallstudien in der Pädagogik. Konstanz-Litzelstetten, 1982.
- Gotthelf, J.* Leiden und Freuden eines Schulmeisters. Bd. 1, Bern, 1838.
- Gröben, N. & Scheele, B.* Argumente für eine Psychologie des reflexiven Subjekts. Darmstadt, 1977.
- Gstettner, P. & Seidl, P.* Sozialpsychologische Aspekte wissenschaftlicher Begleitung. In: Mitter W. & Weishaupt H.: Strategien und Organisationsformen der Begleitforschung. Weinheim, 1979.
- Haenisch, H. & Lukesch, H.* Ist die Gesamtschule besser? Gesamtschulen und Schulen des gegliederten Schulsystems im Leistungsvergleich. München, 1980.
- Huberman, M.* S'évaluer pour s'illusionner? Promesses et écueils de l'évaluation « adaptive / interactive » des innovations scolaires. Cahier du GCR prov., Institut romand de recherche et de documentation pédagogiques, Neuchâtel, 1983.
- Huberman, M. & Miles, M.* L'analyse des données qualitatives: quelques techniques de réduction et de représentation. Cahier du GCR Nr. 6 prov., Institut romand de recherche et de documentation pédagogiques, Neuchâtel, 1983.
- Isenegger, U.* Schulen und Schulsysteme. Grundlagen organisatorischer Neugestaltung. München, 1977.
- Miles, M. & Huberman, M.* Analyzing Qualitative Data: a source-book of new methods. London, 1984 (in Vorbereitung bei Sage).
- Oertel, L.* Abteilungsübergreifende Versuche an der Oberstufe (AVO), Kanton Zürich. In: Meylan J. R.: Innovation und Evaluation von Schulversuchen. Bern, 1981.
- Perret, J.-F.* Evaluation et modalités de recherche empiriques. Contribution au débat sur l'élargissement des méthodes d'évaluation pédagogiques. In: Bildungsforschung und Bildungspraxis, 1981, 3, 65-75.
- Petri, G.* Evaluation der Schulversuche im Bereich der Schulen der Zehn- bis vierzehnjährigen. Gesamtbericht über die erste Evaluationskohorte. Teil 1: Lernerfolgsvergleiche, Schülerlaufbahnen. Bundesministerium für Unterricht und Kunst, Zentrum für Schulversuche und Schulentwicklung, Abteilung II, Graz. 1979.
- Petri, W.* Bedingungen spezifischer und allgemeiner Relevanz von Evaluation. In: Meylan, J. R.: Innovation und Evaluation von Schulversuchen. Bern. 1981.
- Ramseier, E.* Schulversuch Manuel. Ziele, Massnahmen, Ergebnisse. Bern, 1982.
- Rutter, M. Maughan, B. Mortimore, P. & Ouston, J.* Fifteen Thousand Hours. Secondary Schools and their effects of Children. London, 1977.
- Teschner, W. -P.* Drei Fallstudien zur (wissenschaftlichen) Begleitung von Innovationen in der Sekundarstufe I. In: Mitter W. & Weishaupt H.: Strategien und Organisationsformen der Begleitforschung. Weinheim, 1979.
- Teschner W. -P.* Organisatorische und methodologische Problem der Verknüpfung von Entwicklung und Evaluation bei Innovation. In: Meylan J. R.: Innovation und Evaluation von Schulversuchen. Bern, 1981
- Teschner, W. -P. Jochimsen, I. D. & Stolz, G. E.* Mobilität in verschiedenen Differenzierungsmodellen. Wissenschaftliche Auswertung der Schulversuche mit Gesamtschulen. Projektbericht 4/1977, Kiel, 1980.
- Tillmann, K. -J. Bussigel, M. Philipp, E. & Rösner, E.* Kooperative Gesamtschule-Modell und Realität. Weinheim, 1979.
- Wottawa, H.* Gesamtschule: Was sie uns wirklich bringt. Düsseldorf, 1982.
- Zinnecker, J.* Der heimliche Lehrplan. Weinheim, 1976.

## RÉSUMÉ

L'évaluation des expériences scolaires : proposition d'une méthode flexible.

*L'évaluation des expériences scolaires doit être conçue selon ses fonctions pour la réforme scolaire. Ni l'évaluation sommative, ni l'évaluation formative ne peuvent fournir les informations nécessaires. Nous proposons un programme d'évaluation, qui combine des méthodes différentes et qui délaisse les critères traditionnels de précision. Le centre de ce programme consiste en analyses de cas, qui se basent sur un cadre commun et qui démontrent la complexité du système scolaire de manière complémentaire. Pour des questions spécifiques il faut comparer les écoles expérimentales et traditionnelles en utilisant des méthodes quantitatives. Si on réussit à saisir la variabilité des écoles traditionnelles, ces comparaisons contribuent à l'analyse des problèmes du système scolaire.*

## SUMMARY

The evaluation of school experiments : Arguments for a flexible approach

*In evaluating school experiments, different types of informational demands have to be taken into account. Neither by evaluating results on the basis of an action research paradigm, nor by their evaluation on the basis of a quasi experimental design is it possible to respond to these demands. We use an approach combining different methods and looking for precision only as far as absolutely necessary. This approach is characterized by case studies of individual experimental schools in trying to document their environmental situation and their internal processes. They are based on a common observational framework which is supposed to capture the complexity of the school as a system. With respect to specific problems analytical and empirical comparisons with the traditional school system are required in order to represent systematically the variations of the traditional system. Thus the problems of the traditional school system are analyzed at the same time.*