

III Unterrichtsbeobachtung und Evaluation

Einführung

Ob und inwieweit Unterrichtsbeobachtung zur Evaluation eines Bildungsprogramms (Curriculum, Schulversuch, usw.) gehört, hängt davon ab, welches Erkenntnisinteresse zugrunde liegt. Das Erkenntnisinteresse ist wiederum abhängig von den Adressaten der Evaluation (Lehrern, Schülern, Eltern, Schulverwaltungen, Bildungspolitikern, Erziehungswissenschaftlern) und den Evaluatoren. Bei der Entwicklung eines Evaluationsplans wird man sich daher zunächst folgende drei Fragen stellen, von deren Beantwortung es abhängt, ob Unterrichtsbeobachtung stattfinden soll und wenn, in welcher Form und in welchem Ausmaß sie erfolgen soll:

- (1) Welche Informationen benötigen die Adressaten der Evaluation?
- (2) Welche Informationen hat man in ähnlichen Untersuchungen zu gewinnen versucht?
- (3) Welche personellen und materiellen Möglichkeiten bestehen für die Durchführung der Evaluation?

Zweifellos ist die Frage nach dem Erkenntnisinteresse und der Art der benötigten Informationen die entscheidende Frage zu Beginn einer Evaluation. So dürften einige Adressaten mehr an einer ideologiekritischen, d. h. intrinsischen Evaluation der Ziele, Inhalte und Materialien eines Curriculum interessiert sein, andere an einer reinen Ergebnisevaluation und wie der andere an einer Aufwands-Effektivitäts-Analyse (vgl. Alkin 146 ff.).

Für viele Adressaten einer Evaluation ist es wichtig, Informationen darüber zu erhalten, wie die Wirklichkeit eines Schulversuchs aussieht und ob die Intentionen einer Innovation im Unterricht realisiert werden, d. h. ob und inwieweit es zu einer Verbesserung der Unterrichtswirklichkeit kommt. Im amerikanischen Bereich ist es hier in den letzten Jahren zu einer intensiven Forschungsarbeit gekommen. Man begnügte sich nicht länger mit der bloßen Erfolgskontrolle in einer Ergebnisevaluation, sondern wollte verstärkt Informationen darüber erhalten, wie bestimmte Ergebnisse erreicht worden sind. Wurde dies bereits zu einem Anliegen derer, die mit Hilfe von Verhaltenszielen das Schwergewicht der Evaluation auf die Er-

gebnisevaluation legen, so mußte es für die Pädagogen zu einer zentralen Frage werden, die die besondere Wichtigkeit der Unterrichtsprozesse auch im Hinblick auf die Zieldefinition des Unterrichts betonten (Eisner 1969; Stenhouse 1971; Brügelmann 1972; Wulf 1972 b).

Im Rahmen einer Evaluationsuntersuchung kommt der Unterrichtsbeobachtung auch insofern erhebliche Bedeutung zu, als sie dazu beitragen kann, die Gründe für Unzulänglichkeiten bei der Realisierung eines Schulversuchs aufzudecken. Es lassen sich mit ihrer Hilfe »Pattern« im Unterrichtsverhalten von Lehrern und Schülern entdecken, die z. B. die Realisierung bestimmter Innovationen behindern. Darüber hinaus kann Unterrichtsbeobachtung unmittelbare Informationen über die Angemessenheit bestimmter, an einem Bildungsprogramm orientierter Lehr- und Lernstrategien liefern und bei etwaiger Unzulänglichkeit zu entsprechenden Modifikationen der im Rahmen der Realisierung der Innovationen erfolgenden Lehrerfortbildung führen. In den letzten Jahren dienten Unterrichtsbeobachtungssysteme in den USA außer zur Erforschung von Lehrer- und Schülerverhalten auch zur Curriculumevaluation (vgl. z. B. Oliver/Shaver 1966) und vor allem zur Lehrerausbildung. In der Lehrerausbildung kommen ihnen wenigstens drei Funktionen zu:

(1) Sie können dazu dienen, Lehrerverhalten und Lehrerausbildungsprogramme zu evaluieren;

(2) sie sensibilisieren den Lehrer, der sie zur Analyse von Unterrichtsverhalten im Rahmen seiner Ausbildung anwendet, für Unterrichtsinteraktionen und helfen ihm bei der Selbstevaluation seines Unterrichts;

(3) sie geben dem Lehrer in einem ihn unmittelbar berührenden Bereich eine Einführung in die Anwendung von Verfahren empirischer Forschung.

Daher haben Unterrichtsbeobachtungssysteme an vielen amerikanischen Universitäten einen festen Platz im Curriculum der Lehrerausbildung.

Unsere oben formulierten Fragen nach den erkenntnisleitenden Interessen der Adressaten der Evaluation und den dementsprechend in den verschiedenen Untersuchungen gewählten Schwerpunkten führen unmittelbar zu der Notwendigkeit, eine Bestandsaufnahme der bestehenden Unterrichtsbeobachtungssysteme zu machen. Dabei muß man das den Systemen zugrunde liegende Erkenntnisinteresse aufdecken und die dazu entwickelten Beobachtungsverfahren analysieren. Daraufhin kann man entscheiden, ob das System zur intendierten Evaluation der Unterrichtsprozesse beitragen kann oder ob man ein anderes verwenden oder gar ein eigenes Beobachtungssystem entwickeln muß. Eine systematische Analyse der vorliegenden Beobachtungssysteme kann außerdem Auskunft darüber geben, wo die bisherigen Schwerpunkte der Unterrichtsbeobachtung bzw. -evaluation lagen. Einen solchen Versuch machen die 15 umfangreichen Bände der

Mirrors for Behavior (Simon/Boyer, 1967; 1970), in denen eine Sammlung von 79 Unterrichtsbeobachtungsinstrumenten dargestellt wird, die nach 7 Kategorien klassifiziert worden sind. Ohne hier auf eine nähere Analyse dieser Anthologie eingehen zu können, sollen doch wenigstens kurz die Kategorien genannt und die Zuordnung der Instrumente wiedergegeben werden, aus der bereits die Schwerpunkte der bisherigen Unterrichtsbeobachtung deutlich werden und die zugleich eine Antwort darauf geben, inwieweit diese Systeme im Rahmen der Evaluation verwendet werden können und inwieweit neue Systeme entwickelt werden müssen, die neuen und anderen Erkenntnisinteressen gerecht werden.

Die sieben Dimensionen der Unterrichtsbeobachtung sind:

1. affektive – der emotionale Inhalt der Kommunikation (62 Systeme)
2. kognitive – der intellektuelle Inhalt der Kommunikation (48 Systeme)
3. psychomotorische – nicht sprachliche Kommunikation (17 Systeme) (mit Körperhaltung, Gesichtsausdruck, Gesten usw.)
4. Aktivität – ein Geschehen, das sich auf eine Person oder einen Gegenstand (z. B. Lesen und Schreiben) bezieht (37 Systeme)
5. Inhalt – worüber gesprochen wird (27 Systeme)
6. soziologische Struktur – Soziologie der Interaktion einschließlich der Frage, wer spricht zu wem in welcher Rolle (26 Systeme)
7. äußere Bedingungen der Schulumwelt – Beschreibungen des Schulraumes, in dem die Beobachtung stattfindet, einschließlich der benutzten Materialien und Einrichtungen (10 Systeme).

Deutlich liegt der Schwerpunkt der Unterrichtsbeobachtung in der affektiven Dimension. Ihr folgen die kognitive, die aktivitäts- und inhaltsbezogene und die soziologische Struktur-Dimension. Entsprechend den Bedürfnissen der Adressaten wird man also im Hinblick auf die Dimensionen der Untersuchung eine Auswahl treffen müssen. Darüber hinaus bietet sich eine weitere Spezifikation des Beobachtungssystems an. Es muß entschieden werden, ob die Unterrichtsbeobachtung sich auf den Lehrer (14 Systeme), die Schüler (7 Systeme) oder auf die Interaktion zwischen ihnen (67 Systeme) beziehen soll (12 Systeme für den außerschulischen Bereich).

So nützlich die Anthologie »Mirrors for Behavior« für den ist, der Unterrichtsprozesse evaluieren will, so wird doch ihr Wert durch die weitgehend deskriptive Zusammenstellung der Systeme eingeschränkt. Es erfolgt keine Bewertung und kritische Sichtung der vorliegenden Instrumente, so daß eine effektive Nutzung auf erhebliche Schwierigkeiten stößt. Bei einer sorgfältigen Analyse der Beobachtungsinstrumente ergibt sich sodann, daß viele Systeme m. E. nicht genügend evaluativ sind, d. h. für die Zwecke der Evaluation nicht in ausreichendem Maße die Dimension der Bewertung der Unterrichtsaktionen berücksichtigen (vgl. auch Scriven 1967, 49, 69).

Nach Aufarbeitung der wohl umfangreichsten Bibliographie zur Unterrichtsbeobachtung im »Summary«-Band der *Mirrors for Behavior* erschien der Abdruck dieser beiden Beiträge von Bellack und Nuthall sinnvoll, da sie eine gute Einführung in die Grundproblematik der Unterrichtsbeobachtung liefern und zugleich den Forschungsstand in diesem Bereich kritisch zusammenfassen und auf neue Aufgabengebiete hinweisen. Die beiden Beiträge machen nicht den Versuch, alle 79 Systeme zu beschreiben. Sie wählen einige der Systeme aus, die für die Entwicklung des ganzen Untersuchungsbereichs von zentraler Bedeutung sind. Die Auswahl erfolgt so, daß die Beobachtungssysteme für die verschiedenen Dimensionen der Unterrichtsbeobachtung exemplarisch sind.

Der Beitrag Bellacks, der selbst eines der am besten ausgearbeiteten Beobachtungssysteme entwickelt hat, versteht sich als eine Einführung in die Methoden der Beobachtung des Unterrichtsverhaltens von Lehrern und Schülern, der zugleich in didaktisch gelungener Weise einige konzeptuelle und methodische Probleme der Entwicklung von Beobachtungssystemen darstellt. Sodann bietet er die exemplarische Beschreibung des auf die affektive Dimension zielenden Flanderschen Systems der Interaktionsanalyse, des Bellackschen Systems der Analyse der kognitiven Dimensionen des Unterrichts und des Oliver/Shaverschen Systems zur Analyse der affektiven, der kognitiven und der Verfahrens-Dimension, das im Zusammenhang mit der Evaluation des Public Issues Harvard Social Studies Project entwickelt wurde. Schließlich werden noch einige Probleme der Entwicklung und des Gebrauchs dieser Analyse-Systeme dargelegt.

Die Ausführungen Nuthalls, der früher an der Entwicklung des bekannten Smith/Meuxschen Systems mitgearbeitet hatte, sind in vieler Hinsicht eine Ergänzung und Erweiterung des Bellackschen Beitrags. Der Autor beschreibt, vergleicht und kritisiert einige weitere Beobachtungssysteme. Ferner werden zahlreiche Ergebnisse der Unterrichtsbeobachtung im Hinblick auf ihren Beitrag zum Verständnis der Ursachen für bestimmte Lernvorgänge dargestellt und analysiert, die von erheblichem Interesse sein dürften. Abschließend wird versucht, die Forschungsergebnisse aufzuarbeiten, die die Schülerleistung in Beziehung zu den Unterrichtsinteraktionen und dem Lehrerverhalten setzen – eine für die Evaluation von Unterrichtsprozessen besonders wichtige Frage.

Die beiden Beiträge bieten zusammen eine Einführung in den Stand der Entwicklung von Systemen zur Unterrichtsbeobachtung und -evaluation, die bislang dem deutschen Leser nicht zugänglich waren, und dürften daher einen wesentlichen Beitrag zur Erweiterung unseres Verständnisses der Unterrichtsinteraktion und der Verfahren zu ihrer Erforschung und Evaluation liefern.

ARNO A. BELLACK

*Methoden zur Beobachtung des Unterrichtsverhaltens
von Lehrern und Schülern*

Während der letzten zehn Jahre wurde durch das aufkommende Interesse an der Erforschung des Unterrichtsverhaltens eine bedeutsame Entwicklung in der pädagogischen Forschung in Gang gesetzt. Es handelt sich hierbei um das Wiederaufgreifen eines seit langem existierenden Anliegens der Erziehungswissenschaftler. Bei den gegenwärtigen Unterrichtsforschungen, die sich von den Arbeiten vorangegangener Perioden unterscheiden, wird auf systematische Verhaltensbeobachtung von Schülern und Lehrern im Unterricht großer Wert gelegt.

Dieser Beitrag stellt eine Auseinandersetzung mit der neueren Entwicklung von Methoden der Verhaltensbeobachtung im Unterricht dar. Er gliedert sich in drei Abschnitte:

- I. Entwicklung von Systemen zur Verhaltensbeobachtung im Unterricht
- II. Beispiele für Beobachtungssysteme
- III. Probleme und Konsequenzen in der Entwicklung und im Gebrauch von Systemen zur Analyse des Lehrer- und Schülerverhaltens.

*I. Entwicklung von Systemen zur Verhaltensbeobachtung
im Unterricht*

Bei der Konstruktion eines Systems zur Beobachtung des Unterrichtsverhaltens muß der Forscher zwei fundamentale Fragen beantworten: (1) Welche Dimensionen des Unterrichtsverhaltens sollen beobachtet werden? (2) Auf welche Art und Weise sollen die Beobachtungen durchgeführt werden? Diese beiden Fragen sind natürlich eng miteinander verknüpft, denn Beobachtungsgegenstand und Beobachtungsmethode sind voneinander abhängig. Der Prozeß der Unterrichtsbeobachtung wird von den Konzeptionen des Forschers über das Verhalten, das er untersucht, und von den Verfahrensweisen bestimmt, mit denen er die ihm zur Verfügung stehenden Daten sammelt.

Welches Unterrichtsverhalten soll beobachtet werden?

Da der Forscher keinesfalls alles in der Klasse beobachten kann, muß er eine Entscheidung darüber fällen, welche Aspekte des Lehrer- und Schülerverhaltens er zum Gegenstand seiner Untersuchungen nimmt. »Jede Beobachtung wird gemacht«, schreibt Kaplan, »sie ist das Produkt einer aktiven Wahl, nicht eines passiven Geschehens. Keine einzige Interpretation ist notwendigerweise die Folge von dem, was beobachtet wurde; es gibt immer viele Möglichkeiten, Verhalten zu klassifizieren« (Kaplan 1964, 133).

Wie der Forscher bei der Klassifikation des Unterrichtsverhaltens verfährt, wird vom Ziel seiner Untersuchung bestimmt, ebenso vom konzeptuellen Rahmen, der den Beobachtungsgegenstand, die Beobachtungsmethode und den Umfang der Beobachtung angibt. Flanders (1965) z. B. entwickelte auf Grund seines theoretischen Interesses am Klassenklima eine Reihe von Kategorien, um die verschiedenartigen Einflüsse des verbalen Lehrerverhaltens zu beschreiben. Im Gegensatz dazu konstruierten B. O. Smith und Meux (1962) auf Grund ihres Interesses an den kognitiven Dimensionen des Unterrichtsgeschehens einen Katalog von Beobachtungskriterien, um logische Operationen im Unterricht zu kategorisieren. Kounin (1967), dessen Interesse sich auf Klassenführung und Disziplin konzentrierte, entwickelte ein Instrument zur Klassifikation von Dimensionen des Lehrerstils, wie er sich bei der Kontrolle des Schülerverhaltens zeigt.

Zur systematischen Beobachtung des Unterrichtsverhaltens, das für die Zielsetzung und den theoretischen Anspruch relevant ist, muß der Wissenschaftler ein zuverlässiges und gültiges Instrument in Form von Schätzskaleten (rating scales) bzw. ein Beobachtungssystem entwickeln, das in operationalisierter Form die zu kategorisierenden oder zu messenden Verhaltensdimensionen spezifiziert.

Schätzskaleten: Beim Gebrauch von Schätzskaleten besteht die Aufgabe des Beobachters darin, das beobachtete Verhalten an einem Ort auf einem Kontinuum oder in einem in einer bestimmten Reihenfolge vorgegebenen Kategoriensystem zu lokalisieren (vgl. Remmers 1963). Ein Beispiel für eine experimentell entwickelte Schätzskaleten ist die von Ryan (1960) für seine Erforschung von Lehrercharakteristika konstruierte Skala. In dieser Untersuchung bewertete der Beobachter, nachdem er einer Unterrichtsstunde beigewohnt hatte, jede der sechsundzwanzig Dimensionen des während des Unterrichts beobachteten Verhaltens und übertrug anschließend seine Bewertung in Form von Schätzungen auf 7-Punkte-Skaleten. Die 22 Dimensionen des Lehrerverhaltens bezogen sich u. a. auf folgende Kategorien: partiisch – fair; barsch – freundlich; zurück-

haltend – temperamentvoll; autokratisch – demokratisch und unsicher – sicher. Zu den Dimensionen des Schülerverhaltens gehörten: teilnahmslos – lebhaft; sich auf andere verlassend – die Initiative selbst ergreifend; Widerstand leistend – sich verantwortlich fühlend. Jede Dimension war in einem speziellen Wörterverzeichnis definiert, wobei die relevanten Verhaltensaspekte, auf denen die Schätzung basierte, gesondert angegeben wurden.

Die Verfahren, in denen die Schätzung *nach* der Beobachtung erfolgte – wie bei den von Ryan (1960) entwickelten – unterliegen allerdings bedeutsamen Einschränkungen, wenn mit ihrer Hilfe die Aktivitäten im Unterricht erfaßt werden sollen. Die Einschätzungen durch die Beurteiler liefern lediglich allgemeine Eindrücke und Erinnerungen an das tatsächliche Unterrichtsgeschehen, nicht jedoch exakte Aufzeichnungen des Lehrer- und Schülerverhaltens. Deshalb und wegen gewisser Unzulänglichkeiten der Schätzskalen als Meßinstrumente werden sie nur in wenigen gegenwärtigen Untersuchungen verwendet, die die direkte Beobachtung des Unterrichtsgeschehens zum Gegenstand haben, obwohl sie in früheren Untersuchungen häufig Anwendung fanden.

Beobachtungssysteme: Die Systeme zur Beobachtung des Unterrichtsverhaltens liefern dem Beobachter eine Reihe von Kategorien, denen das jeweilige Verhalten zugeordnet wird. Meistens sind Verhaltensstimuli oder operationale Definitionen der Kategorien und Kodierungsanweisungen vorhanden, um dem Beobachter die Entscheidung zu erleichtern, welcher der Kategorien das beobachtete Verhalten zuzuordnen ist. Während der letzten Jahre wurde eine große Anzahl von Beobachtungssystemen entwickelt. Simon und Boyer (1967) katalogisierten in ihrer Anthologie von Beobachtungsinstrumenten, »Mirrors for Behavior«, 26 Verfahren; doch enthält sie lediglich die Hälfte aller in der letzten Zeit entwickelten Systeme¹.

Zwischen den vorhandenen Beobachtungssystemen zeigen sich große Unterschiede. Es bestehen Differenzen

1. in den Dimensionen des zu klassifizierenden Unterrichtsverhaltens,
2. in der Art des zur Beobachtung entwickelten Kriterienkatalogs,
3. in dem Bezugsrahmen des Beobachters für das Kodieren,
4. in der beim Kodieren zu benutzenden Verhaltenseinheit,
5. im Anwendungsbereich.

Diese Unterschiede verdeutlichen die Grundproblematik bei der Konstruktion von Beobachtungsinstrumenten.

1. Dimensionen des Unterrichtsverhaltens

Wie bereits dargelegt, werden bei den vorliegenden Systemen zur Verhaltensanalyse bestimmte Aspekte besonders hervorgehoben, andere vernachlässigt. Die vorhandenen Beobachtungskataloge enthalten ein weites Spektrum von Verhaltensweisen. Biddle (1967) fand bei einer Aufarbeitung der Forschungen über Unterrichtsverhalten, daß kürzlich entwickelte Instrumente sich befassen

- a) mit dem Lehrerverhalten im Sinne von Handlungsweisen, Methoden und charakteristischen Rollen,
- b) mit dem Zuhörer- und Zielverhalten von Lehrern und Schülern,
- c) mit der Lehrer-Schüler-Interaktion,
- d) mit äußeren Strukturen – wie Lehrinhalten und administrativen Maßnahmen,
- e) mit inneren Strukturen – wie Kommunikationsstruktur, Aktivitätsstruktur, charakteristischen Rollen, sozialen Funktionen.

Im Gegensatz dazu haben Simon und Boyer (1967) verschiedene Beobachtungsverfahren in drei Gruppen nach folgender Systematik kategorisiert:

- a) affektive Systeme, mit deren Hilfe man das emotionale Klassenklima erfassen kann und die Art und Weise, wie es durch die Reaktion der Lehrer auf die Emotionen, Ideen oder Handlungsweisen der Schüler bedingt wird,
- b) kognitive Systeme, die sich mit Denkprozessen befassen und den diese zum Ausdruck bringenden verbalen Verhaltensmustern,
- c) multidimensionale Systeme, mit denen man sowohl die kognitiven als auch die affektiven Dimensionen des Verhaltens erfassen kann.

Das Flanderssche System der Interaktionsanalyse (1965) und das Hughessche System zur Klassifizierung der Funktionen des Lehrerverhaltens (1959) sind bedeutsame Beispiele für das erste System. Das System von Bellack und anderen zur Beschreibung des Sprachverhaltens im Unterricht (1966) und B. O. Smiths' und Meux' Verfahren zur Analyse logischer Operationen des Unterrichtens (1962) sind typisch für die zweite Gruppe. Olivers und Shavers Beobachtungssystem zur Beschreibung kontrastierender Unterrichtsstile in dem Fach Sozialkunde (social studies) (1966) sowie das System von Joyce zur Erfassung affektiver und kognitiver Aspekte der verbalen Lehrerkommunikation (vgl. Joyce/Harootunian 1967) sind Repräsentanten der dritten Kategorie.

Die vorhandenen Systeme zur Beobachtung des Unterrichtsverhaltens können ebenso nach der Art der Kommunikation klassifiziert werden, d. h. danach, ob sie die Aufmerksamkeit auf (a) verbale Aspekte, (b) nicht-ver-

bale Aspekte oder (c) verbale und nicht-verbale Aspekte der Kommunikation konzentrieren. Die meisten Systeme richten sich ausschließlich auf verbales Verhalten, während einige Systeme verbale und nicht-verbale Dimensionen beinhalten. Zwanzig der Systeme, die in »Mirrors for Behavior« (Simon und Boyer 1967) beschrieben sind, befassen sich mit verbaler Kommunikation, nur sechs beziehen sich sowohl auf verbales als auch auf nicht-verbales Verhalten. Soweit der Autor informiert ist, ist Galloway (1962) der einzige Forscher, der für seine Untersuchungen über nicht-verbale Kommunikation, wie die Gestik des Lehrers, den Ausdruck der Stimme und die Mimik, ein Instrument nur zur Beschreibung des nicht-verbalen Verhaltens entwickelt hat.

Die Entscheidung, welche Aspekte des Lehrer-Schüler-Verhaltens zu untersuchen sind, ist prinzipiell ein theoretisches Problem. Es sind jedoch auch wichtige methodische Erwägungen miteinbezogen, z. B. ist dafür zu sorgen, daß die Kategorien in einem vorgegebenen System sich gegenseitig ausschließen und einen definierten Bereich erschöpfend repräsentieren. Das Kriterium der Vollständigkeit bedeutet natürlich nicht, daß jedes Verhalten während einer bestimmten Beobachtungsperiode klassifiziert wird. Vielmehr ist es erforderlich, daß das zu beobachtende Verhaltensuniversum (z. B. soziales Klima, logische Operationen, Kommunikationsprozesse) klar definiert und eine repräsentative Stichprobe von Kategorien aus diesem Verhaltensuniversum entnommen wird. Zum Beispiel schließen sich die folgenden fünf von Aschner und Gallagher (1965) entwickelten allgemeinen Kategorien zur Klassifikation von Gedankenprozessen, die sich im verbalen Unterrichtsverhalten widerspiegeln, gegenseitig aus und sind auch repräsentativ für ein Universum kognitiver Funktionen: mechanisches Denken, Gedächtnistätigkeit, konvergentes Denken, evaluatives Denken, divergentes Denken.

In dem Stadium der Forschungsarbeit, in dem die relevanten theoretischen Konzepte noch nicht vollständig entwickelt sind, ist es häufig schwer, Kategorien aufzustellen, die sich gegenseitig ausschließen und die repräsentativ sind. Jedoch muß der Wissenschaftler im Verlauf seiner Forschungsarbeit diese Kriterien ständig beachten und sich nicht nur auf die Analyse der empirischen Beobachtungsdaten, sondern ebenso auf die Verbesserung der zur Analyse der Daten verwendeten Begriffe konzentrieren.

2. Verschiedene Arten von Beobachtungssystemen

In der Erforschung des Unterrichtsverhaltens werden heute weitgehend zwei Arten von Beobachtungssystemen angewandt; Medley und Mitzel (1963) bezeichnen diese als Kategoriensysteme und Zeichensysteme.

Ein Kategoriensystem beschränkt die Beobachtung auf spezifische Verhaltensdimensionen im Unterricht, wobei es eine Reihe von Kategorien liefert, in die jede beobachtete Verhaltenseinheit eingestuft wird. Das daraus resultierende Ergebnis zeigt für jeden Beobachtungszeitraum die vollständige Anzahl der aufgetretenen Verhaltenseinheiten an sowie die Anzahl innerhalb jeder einzelnen Kategorie. Die von Hughes (1959) in ihrer Untersuchung über die Interaktion in der Primarstufe entwickelte Reihe von Kategorien ist ein typisches Beispiel dafür. Der dieses System benutzende Beobachter klassifiziert das verbale und nicht-verbale Lehrerverhalten nach sechs Hauptfunktionen: Kontrollfunktionen, unterstützende Funktionen, Funktionen, die die Vermittlung von Unterrichtsinhalten zum Gegenstand haben, Funktionen, die als persönliche Antwort dienen, Funktionen von positivem und von negativem Gefühlswert. Im Gegensatz dazu liefert ein Zeichensystem dem Beobachter eine Reihe spezifischer Verhaltensweisen, die während einer Beobachtungsperiode auftreten oder nicht auftreten können. Die Beobachter werden angewiesen, auf diese spezifischen Verhaltensweisen zu achten, und das Ergebnis zeigt, welche dieser Verhaltensweisen während des Beobachtungszeitraums aufgetreten sind. Ein Beispiel für ein solches Beobachtungssystem ist die von Medley und Mitzel (1958) entwickelte Beobachtungsskala (OSCAR). Ein Beobachter, der dieses Instrument verwendet, registriert lediglich die Verhaltensweisen, die von den 71 Items erfaßt werden. Zu diesen in mehrere Abschnitte gruppierten Items gehören etwa solche: »der Lehrer trägt vor«, »der Lehrer beantwortet die Fragen der Schüler«, »der Lehrer veranschaulicht etwas an der Tafel«, »der Lehrer reagiert sarkastisch«, »der Schüler spricht zur Gruppe«, »der Schüler flüstert«, »der Schüler liest oder arbeitet an seinem Platz«.

Während Zeichensysteme gewöhnlich aus einer großen Itemanzahl bestehen, die sich auf konkrete, spezifische Verhaltensweisen beziehen und daher ein geringes Maß an Schlußfolgerungen seitens des Beobachters erfordern, sind Kategoriensysteme gewöhnlich aus einer geringeren Anzahl von Items zusammengestellt und besitzen ein höheres Abstraktionsniveau, das einen höheren Grad an schlußfolgerndem Denken beim Beobachter voraussetzt. Nicht alle Kategoriensysteme sind auf dem gleichen Abstraktionsniveau, aber in der Regel auf einem höheren als die Zeichensysteme konstruiert. Die in den beiden vorhergehenden Abschnitten erwähnten Beispiele veranschaulichen den Unterschied zwischen mehr oder weniger schlußfolgerndes Denken erfordernden Kategorien. Der Beobachter ist weniger auf Schlußfolgerungen angewiesen, wenn er die Handlung eines Lehrers zu interpretieren hat, wie z. B. »Beantwortung der Schülerfragen« oder »Veranschaulichung an der Tafel«. Im Gegensatz dazu er-

fordert es einen höheren Grad an schlußfolgerndem Denken, eine Aussage des Lehrers als Kontroll- oder Unterstützungsfunktion zu interpretieren. Es sollte ebenso festgehalten werden, daß Kategorien, die in einem hohen Maße schlußfolgerndes Denken erfordern und die die Verhaltensbeobachtung als Basis für Schlußfolgerungen über Motive oder Auswirkungen des Verhaltens beinhalten, den Forscher bei der Validitätsbestimmung mit schwierigeren Problemen konfrontieren, als es bei Kategorien der Fall ist, die sich mit deskriptiven Verhaltensbeobachtungen befassen und oft als solche bezeichnet werden, die Augenscheinvalidität besitzen.

Medley und Mitzel (1963) stellen fest, daß Kategoriensysteme häufiger in empirischen Untersuchungen angewandt werden, die auf hochdifferenzierten Theorien beruhen, während Zeichensysteme dann verwendet werden, wenn die Theorie nicht genügend aussagekräftig ist. Es ist selbstverständlich, daß die theoretische Orientierung des Forschers und der Grad der Entwicklung seiner Theorie das Niveau der Konzeptualisierung und folglich auch den Grad des erforderlichen Maßes an schlußfolgerndem Denken bestimmen.

3. Bezugsrahmen des Beobachters

Die Wissenschaftler versuchen, die Vorgänge im Schulunterricht wenigstens unter drei Aspekten zu kategorisieren:

- a) die Absicht oder das Motiv des Handelnden
- b) die Auswirkungen des Verhaltens auf den Adressaten
- c) die objektiven Verhaltenscharakteristika.

Beispiele dieser drei Aspekte finden sich in neueren Untersuchungen über Unterrichtsprozesse.

Das analytische System von Withall (1949), das er in Verbindung mit seinen Forschungen über sozial-emotionales Klassenklima entwickelt hat, erfordert vom Beobachter eine Interpretation der Aussagen des Lehrers dahingehend, ob ihre Intention eine schülerunterstützende, problemstrukturierende oder direktive ist. Hughes (1959) andererseits klassifizierte in ihrer Untersuchung über das Unterrichtsverhalten von Primarschullehrern die Funktionen des verbalen und nicht-verbalen Lehrerverhaltens nach ihren erwarteten Wirkungen und deren Bedeutung für die Schüler. Im Gegensatz dazu beschreiben Smith und Meux (1962) in ihrer Untersuchung über die logischen Aspekte des Unterrichtsgeschehens die logischen Bestandteile der Lehrer- und Schüleraussagen, wobei sie über Motive des Sprechers oder seine Wirkungen auf die Zuhörer keine Aussagen machen.

Die Erfahrungen dieser und anderer Forscher verdeutlichen, daß das Unterrichtsverhalten auf der Basis aller drei Aspekte zuverlässig kodiert werden kann. Welcher der drei Aspekte für eine Untersuchung angemessen

sen ist, kann nur entschieden werden, wenn man das Ziel der Untersuchung berücksichtigt. Biddle (1967) hat mehrere mögliche Zielsetzungen genannt, die für die Unterrichtsforschung in Frage kommen, sowie Konzeptionen entwickelt, die diesen Zielen angemessen sind. Zum Beispiel ist er der Auffassung, daß Urteile über die Intention des Lehrers angemessen sind, wenn man primär an den Determinanten des Lehrerverhaltens interessiert ist. Wenn man sich allerdings mit der Qualifikation des Lehrers befaßt, sind Urteile über die Wirkung des Lehrerverhaltens auf die Lernprozesse der Schüler angebrachter. Wenn andererseits die individuellen und sozialen Determinanten des Verhaltens erforscht werden sollen oder kontrastierende Modelle der Interaktion im Klassenraum zu testen sind, wäre es sinnvoll, objektive Verhaltenscharakteristika als Grundlage zu nehmen.

4. Verhaltenseinheiten

Eine entscheidende Aufgabe bei der Erstellung eines Beobachtungssystems besteht darin, die Verhaltenseinheit zu spezifizieren, die als Basis für das Kodieren benutzt wird. Für die vielen Möglichkeiten, die Verhaltenseinheit zu definieren, gibt es zwei grundsätzlich verschiedene Methoden: (a) Bestimmung einer willkürlichen Zeiteinheit, (b) genaue Beschreibung einer analytischen Einheit, die häufig durch das Kategoriensystem selbst vorgegeben wird.

Bei den Systemen, bei denen eine willkürliche Zeiteinheit benutzt wird, muß der Beobachter ein Protokoll über die in dieser festgesetzten Zeit auftretenden Verhaltensweisen erstellen. Flanders (1965) z. B. verlangt vom Beobachter alle drei Sekunden ein Urteil darüber, ob der Lehrer innerhalb der kurzen Periode direkten oder indirekten Einfluß auf die Schüler ausübt. In gleicher Weise schreibt Spaulding (vgl. Simon/Boyer 1967) Zeitperioden von 3–10 Sekunden vor, für die der Beobachter ein Urteil über das Verhalten des Lehrers hinsichtlich seiner Kontrollfunktion in der Klasse abgibt, wobei er Kategorien wie »Setzen von Verhaltenszielen« und »Anordnen bestimmter Aktivitäten« verwendet. Medley und Mitzel (1958) andererseits erwarten, daß der Beobachter Verhaltensweisen in 5-Minuten-Perioden protokolliert.

Der grundlegende Vorteil der Setzung beliebiger Zeiteinheiten liegt in ihrem mechanischen Charakter, der eine Hilfe für die Regulierung des Beobachtungsprozesses darstellt. Weiterhin kann diese Methode bei der Analyse des verbalen und des nicht-verbalen Verhaltens angewandt werden. Jedoch liegt die Problematik der Zerlegung des Unterrichtsgeschehens in willkürliche Zeiteinheiten darin, daß die Ergebnisse nicht die na-

türlich eintretenden Verhaltensmuster oder den Ablauf des Verhaltens widerspiegeln, wie es sich allmählich entwickelt.

Anstatt eine beliebige Zeiteinheit anzugeben, entscheiden sich viele Forscher für Analyse-Einheiten verschiedener Art. Diese Einheiten repräsentieren einzelne Elemente des verbalen und/oder nicht-verbalen Verhaltens, die beim Kodieren verwendet werden, um Unterrichtsgeschehnisse in Einzelkomponenten zu zerlegen. Analyse-Einheiten werden von Wissenschaftlern unterschiedlich definiert, und zwar im Sinne von:

- a) Aktivität im Unterricht,
- b) verbaler Kommunikation zwischen zwei oder mehreren Sprechern,
- c) Kommunikation oder Mitteilung eines einzelnen Sprechers,
- d) Ausdruck einer Gedankeneinheit durch einen Sprecher.

Diese Analyse-Einheiten können durch folgende Beispiele veranschaulicht werden:

a) In seiner Untersuchung über Unterrichtsverhalten und mangelndes Leistungsvermögen der Schüler führte Perkins (1964, 1965) als grundlegende Analyse-Einheiten sechs Arten von Klassenaktivitäten auf: Großgruppendifkussion, Klassenvortrag, Einzel- oder Projektarbeit, die nicht allen Schülern zugewiesen ist, Bearbeitung einer allen zugewiesenen Aufgabe, Kleingruppenarbeit, mündliche Berichterstattung. Aspekte des Schüler- und Lehrerverhaltens wurden im Kontext mit diesen Aktivitätsarten kodiert.

b) In ihrer Untersuchung der logischen Dimensionen des Unterrichtsverlaufs bezeichneten B. O. Smith und Meux (1962) als fundamentale Verhaltenseinheit den verbalen Austausch zwischen zwei oder mehreren Sprechern, wobei drei Phasen zu unterscheiden sind: Anfangs- oder Eröffnungsphase, Durchführungsphase und Endphase.

c) Jackson (1965) definierte drei Arten von verbalen Äußerungen als fundamentale Einheiten zur Beschreibung der Kommunikation in der Primarstufe: Äußerungen, die sich auf (a) inhaltliche Ziele, (b) Verfahrensweisen und Regeln bei der Gruppenarbeit, (c) Mitteilungen zur Aufrechterhaltung der Disziplin und Ordnung beziehen.

d) In ihrer Untersuchung über die Lehrstrategien zur Schulung kognitiver Fähigkeiten bei Primarschülern bestimmten Taba u. a. (1964, 1966) die fundamentale Analyse-Einheit als »Gedanken-Einheit«, definiert als »Äußerung oder Serie von Äußerungen, die einen mehr oder weniger vollständigen Gedanken zum Ausdruck bringen, eine spezifische Funktion besitzen und entsprechend dem jeweiligen Niveau des Gedankens klassifiziert werden können« (1966, 134).

5. Anwendungsbereich

Die Beobachtungssysteme unterscheiden sich auch dadurch voneinander, daß sie in verschiedenem Ausmaß auf andere Forschungsvorhaben und Populationen anwendbar sind.

Einige Systeme wurden in der Absicht entwickelt, sie für verschiedene Klassen, Altersstufen, Fächer und Schülergruppen verwenden zu können. Zum Beispiel ist das System der Interaktionsanalyse von Flanders in Primar- und Sekundarstufen anwendbar, in denen Lehrer mit Schülern verbal in Interaktion treten. Gleichermaßen wurde das Verfahren zur Analyse der strukturellen und funktionellen Aspekte des Kommunikationssystems im Unterricht von Biddle und Adam bewußt für die Anwendung in Primar- und Sekundarstufen entwickelt, in denen zahlreiche Fächer unterrichtet werden.

Im Gegensatz dazu sind einige Beobachtungssysteme zur Analyse von Unterrichtsprozessen nur für bestimmte Altersstufen geeignet. Das System von Hughes (1959) zur Beschreibung der Lehrerfunktionen ist in Primarstufen anwendbar, während das System zur Analyse logischer Operationen des Unterrichtens von B. O. Smith und Meux (1962) lediglich für Klassen der Sekundarstufe in Frage kommt, in denen Fächer mit stärkerem Wissenschaftscharakter unterrichtet werden.

Andere Systeme sind auf bestimmte Fächer oder Schülergruppen beschränkt. Das System von Wright (1959) wurde zur Untersuchung des Mathematikunterrichts entwickelt, während das Verfahren von Oliver und Shaver (1966) vornehmlich zur Erforschung des Sozialkundeunterrichts (social studies) eingesetzt werden kann.

Einige Forscher befassen sich nur mit bestimmten Schülergruppen: Kounin (1967) mit emotional gestörten Schülern, Perkins (1964, 1965) mit leistungsschwachen Schülern und Smith und Geoffrey (1968) mit Schülern von sozioökonomisch niedrigerem Status.

Ob die Forscher versuchen sollten, umfassende multidimensionale Systeme zu konstruieren, die auf einen großen Bereich von Unterrichtssituationen anwendbar sind, oder sich darauf beschränken sollten, Instrumente für eine begrenzte Anzahl von Unterrichtssituationen zu entwickeln, ist eine umstrittene Frage, auf die später näher eingegangen wird.

Wie werden Unterrichtsbeobachtungen durchgeführt?

Registrieren und Kodieren von Verhaltensweisen

Eine Vielzahl verschiedener Registrier- und Kodiertechniken zur Fixierung von Verhaltensweisen wurde im Zusammenhang mit den oben erörterten

Beobachtungssystemen angewandt. Einige Forscher (z. B. Flanders 1965; Jackson 1965; Medley/Mitzel 1958) fordern vom Beobachter, das Verhalten so zu kodieren, wie es sich tatsächlich ereignet hat; d. h. der Beobachter, dem die Aufgabe des Kodierens zukommt, transformiert die beobachteten Ereignisse in Symbole, die gezählt und tabuliert werden können. Der Hauptvorteil beim unmittelbaren Kodieren des Verhaltens liegt darin, daß man direkten Zugang hat zu visuellen Reizen wie Mimik und Gestik der Lehrer und Schüler und zu situationsgebundenen Faktoren während des Unterrichts, die für die genaue Interpretation des beobachteten Verhaltens relevant sein können. Ein schwerwiegender Nachteil liegt darin, daß es äußerst schwierig ist, die komplexen Verhaltensweisen unmittelbar in dem Augenblick zuverlässig zu kodieren, in dem sie während des Unterrichts auftreten. Daher protokollieren viele Forscher zunächst das Verhalten während des Unterrichts und kodieren es anschließend mit Hilfe des Protokolls.

Bei der Erforschung des Unterrichtsverhaltens wurden verschiedene Arten der Berichterstattung verwendet, u. a. schriftliche Protokolle, Tonbandaufnahmen und audiovisuelle Aufzeichnungen. Hughes (1959) sammelte in ihrer Untersuchung über den Unterricht in der Primarstufe Daten in Form von schriftlichen Protokollen, die in der Hauptsache aus einem chronologischen Bericht über das verbale und nicht-verbale Lehrerverhalten im Unterricht bestanden. Zwei geübte Beobachter protokollierten gleichzeitig das Unterrichtsgeschehen; das endgültige Protokoll über eine Beobachtungsperiode bestand lediglich aus Beschreibungen, denen beide Beobachter zustimmten. Es wurde kein Versuch unternommen, das Verhalten zu kategorisieren; die Beobachter berichteten lediglich in Kurzfassung über die Aussagen und Handlungen der Lehrer. Beim Kodieren klassifizierte man später die protokollierten Daten in sieben Hauptkategorien hinsichtlich der Lehrerfunktionen.

Hughes vertritt die Auffassung, daß das Sammeln von Unterrichtsdaten in Form von schriftlichen Protokollen verschiedene Vorteile hat, z. B. ermöglicht dieses Vorgehen den Forschern, den Unterrichtsverlauf so zu fixieren, daß er hinsichtlich seiner spezifischen Qualität sowie seiner charakteristischen Merkmale untersucht werden kann; die Kontinuität des Lehrerverhaltens wird festgehalten; die Protokolle sind insofern neutral, als während der Beobachtung des Unterrichtsgeschehens keine Bewertung abgegeben wird. Diese Neutralität der schriftlichen Berichte muß jedoch ernsthaft in Frage gestellt werden, da die Protokolle über das Unterrichtsverhalten aus zweiter Hand durch den Beobachter abgefaßt werden, dessen Subjektivität der Wahrnehmung sich unvermeidbar in seinem Bericht über das Geschehen widerspiegelt.

Diese Unzulänglichkeit von schriftlichen Berichten kann weitgehend überwunden werden durch den Gebrauch von Tonbändern oder audiovisuellen Aufzeichnungen, die fortlaufende, objektive Verhaltensprotokolle liefern. In mehreren Untersuchungen zur Erforschung des verbalen Unterrichtsverhaltens (z. B. Bellack u. a. 1966; Taba u. a. 1964; Taba 1966; und B. O. Smith/Meux 1962) sind die Ereignisse mit Hilfe von Fernsehkameras aufgezeichnet worden. Häufig werden von Tonbandaufzeichnungen maschinenschriftliche Protokolle angefertigt; derjenige, dem die Aufgabe des Kodierens zufällt, hat bei der Kategorisierung des Verhaltens Zugang zu den Tonbandaufzeichnungen und den maschinenschriftlichen Protokollen. In Anbetracht der Schwierigkeit, das komplexe verbale Verhalten im Unterricht der Realität entsprechend zu beobachten und zu analysieren, haben diese Aufzeichnungen über den Unterrichtsverlauf offensichtlich Vorteile, da sie wiederholt von verschiedenen Perspektiven her analysiert werden können. Die prinzipielle Unzulänglichkeit von Tonbandaufnahmen liegt darin, daß sie keine Informationen über nicht-verbales Verhalten liefern, das als solches sehr bedeutsam sein oder als Hinweis auf die adäquate Interpretation des verbalen Verhaltens dienen kann.

Dieser Schwierigkeit kann durch die Verwendung von audiovisuellen Aufzeichnungen begegnet werden, die ein umfassendes Protokoll der Unterrichtsabläufe liefern, da sie verbales und nicht-verbales Verhalten erfassen. Kounin (1967) und Biddle und Adams (1967) haben Untersuchungen abgeschlossen, in denen Beobachtungsdaten während des regulären Unterrichts mit transportablen Videorekordern aufgezeichnet wurden. Die Hauptprobleme liegen in den damit verbundenen Kosten und den technischen Schwierigkeiten beim Gebrauch der komplizierten Apparatur. Der Leser, der am Einsatz von Videorekordern in der Unterrichtsforschung interessiert ist, wird auf den Bericht von Biddle und Adams verwiesen (1967).

Beziehungen zwischen Beobachter und Beobachteten

Ein häufig erhobener Einwand gegen Beobachtungsverfahren liegt darin, daß die Gegenwart eines Beobachters oder eines Aufnahmeinstruments derart ablenkend ist, daß das beobachtete Verhalten nicht als typisches Verhalten angesehen werden kann. Dies scheint jedoch nach den Urteilen erfahrener Forscher keine stichhaltige Kritik zu sein. Heyns und Lippitt stellen in ihrem umfangreichen Überblick über systematische Beobachtungstechniken in der sozialpsychologischen Forschung fest, daß die Wissenschaftler, die Beobachter zur Erforschung des Unterrichtsgeschehens einsetzen, die allgemeine Auffassung teilen, daß »die Beobach-

ter, wenn überhaupt, nur sehr geringe Wirkung ausüben. Dieser Überzeugung sind auch experimentell arbeitende Forscher, die in einer Vielfalt von Situationen und auf verschiedenartigen Fachgebieten arbeiten« (Heyns/Lippitt 1954, 399).

Die meisten Forscher im Bereich der Erziehungswissenschaft scheinen diese Auffassung zu teilen. Biddle und Adams (1967), deren komplizierte audiovisuelle Apparatur aus zwei Kameras und zwei Mikrofonen bestand, die in den Klassenräumen aufgestellt wurden, in denen sie ihre Beobachtungen vornahmen, kommentieren die Auswirkungen dieser Apparatur auf Schüler und Lehrer folgendermaßen:

»Alle beteiligten Lehrer wurden informell über den Effekt der Aufnahmeapparatur interviewt. Einige berichteten über gewisse Spannungsgefühle zu Beginn der ersten protokollierten Unterrichtseinheit, stellten jedoch fest, daß diese Spannung dann aufhörte, wenn sie sich im Unterricht engagierten. Die Forscher bemerkten, daß einige Lehrer sich an den Tagen besonders sorgfältig zu kleiden schienen, an denen die Aufnahmen stattfanden, obwohl ein derartiger Effekt bei den Schülern nicht festgestellt wurde, die – es soll noch einmal daran erinnert werden – nicht wußten, an welchen Tagen die Aufnahmen gemacht wurden. Gelegentlich zeigten sich deutliche Anzeichen dafür, daß die Schüler sich der Aufnahmekameras bewußt waren, da sie auf die vermuteten Kameras starrten oder in der Pause vor ihnen Theater spielten. Jedoch wurde überall bestätigt, daß die Kameras die Schüler offensichtlich kaum ablenkten« (Biddle/Adams 1967, 217).

Ähnliche Auffassungen, daß die von Beobachter und Aufnahmegerät ausgehenden Wirkungen auf das Unterrichtsverhalten als unbedeutend angesehen werden können, vertraten u. a. auch: B. O. Smith und Meux (1962), Hughes (1959), Flanders (1965) und Bellack u. a. (1966).

Einige Wissenschaftler trafen besondere Vorkehrungen, um die Wirkungen zu verringern, die von Beobachtern und von der Aufnahmeapparatur ausgehen. Zum Beispiel weisen B. O. Smith und Meux (1962) darauf hin, daß die Informationen für die mit ihnen zusammenarbeitenden Lehrer die Versicherung enthielten, daß die Untersuchung keinerlei Bewertung ihrer Lehrfähigkeit bezwecke, daß bei der Veröffentlichung der Forschungsergebnisse völlige Anonymität gewährleistet sei und daß nur Mitglieder des Forscherteams Einblick in die Tonbänder und Manuskripte hätten. Außerdem wurden die Aufnahmegeräte einige Tage vor der Unterrichtsstunde, in der die Beobachtungen stattfanden, im Klassenraum installiert, damit sich Schüler und Lehrer daran gewöhnen konnten.

Ohne die Probleme zu unterschätzen, die durch Beobachter und Aufnahmegerät entstehen, scheint es, daß diese Einflüsse zum großen Teil

ausgeschaltet werden können, wenn entsprechende Vorsichtsmaßnahmen getroffen werden. Da jedoch nicht angenommen werden kann, daß Lehrer und Schüler durch die Anwesenheit eines Beobachters und durch das Vorhandensein eines Aufnahmeapparates trotz derartiger Maßnahmen völlig unbeeinflusst bleiben, sollte man sich an die einfache, offensichtliche Tatsache erinnern, daß es besser ist, »etwas über das Lehrer-Schüler-Verhalten in Erfahrung zu bringen, während sie unter Beobachtung stehen, als überhaupt keine Kenntnisse über das Lehrer-Schüler-Verhalten gewinnen zu können« (Medley/Mitzel 1963, 248).

Die Zuverlässigkeit beim Kodieren

Die Zuverlässigkeit eines Beobachtungsinstruments für das Unterrichtsgeschehen kann definiert werden als »das Ausmaß, in dem die Messung, unter konstanten Bedingungen wiederholt, konstante Ergebnisse erbringt« (Kaplan 1964, 200). Unter den relevanten Bedingungen bei der Verhaltensmessung während des Unterrichts sind von besonderer Bedeutung die Beobachter, die die Messungen vornehmen, sowie die Stabilität des zu messenden Verhaltens. Daher kann die Zuverlässigkeit von Instrumenten zur Beobachtung von Unterrichtsabläufen geschätzt werden durch den Grad der Übereinstimmung zwischen voneinander unabhängigen Beobachtern (Übereinstimmungskoeffizient) und auf Grund der Stabilität der Verhaltensdimensionen unter Beobachtung (Stabilitätskoeffizient). Obwohl beide Arten der Zuverlässigkeit offensichtlich von Bedeutung sind, wird der Stabilitätskoeffizient bei der Erforschung des Unterrichtsverhaltens gegenwärtig kaum berechnet; größere Aufmerksamkeit wird der Schätzung des Ausmaßes an Übereinstimmung zwischen den Beobachtern geschenkt. Die gebräuchlicheren statistischen Indizes für diese Schätzungen sind der Prozentsatz der Übereinstimmung zwischen den kodierenden Personen und der Korrelationskoeffizient; von einigen Wissenschaftlern wird auch die Varianzanalyse angewandt.

Zur Sicherung einer adäquaten Zuverlässigkeit beim Gebrauch von Beobachtungsskalen haben Wissenschaftler bestimmte Verfahrensweisen entwickelt. Von zentraler Bedeutung ist die sorgfältige Entwicklung der Beobachtungsskala selbst, indem man besonders auf die präzise Definition der Kategorien und der Analyse-Einheit und auf die Formulierung von Kodier-Regeln zur Anleitung der kodierenden Personen achtet. Wahrscheinlich resultieren die größten Schwierigkeiten beim zuverlässigen Kodieren aus der unklaren Definition der zu kodierenden Verhaltenseinheit; ohne eine derartige präzise Definition kann kein hoher Grad an Übereinstimmung zwischen den Beobachtern erzielt werden. Die sorgfältige Schu-

lung der Beobachter ist zweifellos ein wesentliches Mittel zur Sicherung der Zuverlässigkeit; die Erfahrung der Forscher bestätigt die Wichtigkeit intensiver Schulungsprogramme für die beobachtenden und kodierenden Personen und für die Notwendigkeit eines systematischen Vorgehens zur Kontrolle der Beobachtungen und Kodierungen.

Statistische Analyse der Daten

Eine detaillierte Diskussion über die Methoden einer statistischen Analyse der Beobachtungsdaten aus dem Unterrichtsgeschehen geht über den Rahmen dieses Beitrags hinaus. Dennoch ist darauf hinzuweisen, daß neuere Theorien und Techniken es ermöglichen, bestimmte Dimensionen von Unterrichtsprozessen zu erforschen, die zuvor nur schwer, wenn überhaupt, statistisch zu erfassen waren. Biddle und Adams (1967) und Bellack u. a. (1966) haben z. B. die zeitliche Verlaufsscharakteristik von Ereignissen im Unterrichtsgeschehen statistisch mit Hilfe einer Markoff-Kette beschrieben. In der letzteren Untersuchung versuchten die Wissenschaftler festzustellen, ob bestimmte pädagogische Verhaltensmuster dahin tendieren, andere pädagogische Verhaltensmuster nach sich zu ziehen. Dies wurde statistisch durch die Markoff-Kette beschrieben, mittels derer es möglich ist, die bedingte Wahrscheinlichkeit des Übergangs von einem Verhaltensmuster auf ein anderes zu bestimmen (Kemeny/Snell 1962). Nimmt man die verschiedenen pädagogischen Verhaltensmuster als Einheiten an, dann wurde die Wahrscheinlichkeit des Übergangs von einem Verhaltensmuster zu einem anderen erforscht, um festzustellen, ob ein pädagogisches Verhaltensmuster die Tendenz hat, das jeweils unmittelbar nachfolgende Verhalten zu beeinflussen.

Die statistische Auswertung von Beobachtungsdaten erfuhr eine wesentliche Erleichterung durch den Einsatz von Computern. Gage (1969) berichtete über wichtige Ergebnisse aus der Arbeit von Allen und Snow am Stanford Center for Research and Development in Teaching, die durch den Einsatz von Computern ermöglicht wurden. Allen und Snow entwickelten eine Taxonomie von Verhaltensweisen im Unterricht, indem sie Computer zum Speichern und zum Reproduzieren von Items verwendeten, die sich auf Lehrer- und Schülerverhalten beziehen. Gage beschreibt die Ziele und das Verfahren dieses Projektes:

»Es wurden mehr als 1000 Items gespeichert; diese können vom Computer in einer endgültigen Fassung reproduziert und gedruckt werden, so daß sie den Beobachtern, Bewertern und Wissenschaftlern für die Inhaltsanalyse zur Verfügung stehen. Die Items, die reproduziert werden sollen, können entsprechend vielen differierenden Dimensionen exakt angegeben werden. Die von den Beobachtern oder

den für die Inhaltsanalyse verantwortlichen Wissenschaftlern gewonnenen Daten, die die Häufigkeit, Intensität oder die Verhaltenskorrelate betreffen, die durch diese Items gekennzeichnet werden, können zusammen mit den Items im Computer gespeichert werden. Auf diese Weise werden die Erfahrungswerte gesammelt, die für die Gültigkeit und Zuverlässigkeit der Items bei verschiedenen Lehrertypen, Beobachtern, Lehrinhalten, Effektivitätskriterien und dergleichen relevant sind. Kurz, das Ideal einer universalen Taxonomie des Unterrichtsverhaltens, die für viele Zwecke nützlich ist, kann durch ein auf dem Computer basierendes System – wie in Stanford – erreicht werden« (Gage, 1969, 1452).

II. Beispiele für Beobachtungssysteme

In diesem Abschnitt werden kurz drei Analyse-Systeme beschrieben, die drei Haupttypen von vorhandenen Beobachtungssystemen charakterisieren: (1) Das auf das affektive Geschehen zielende System der Flandersschen Interaktionsanalyse, das sich mit dem sozialen Klima im Klassenraum befaßt (1965); (2) das System von Bellack u. a. zur Analyse des Sprachverhaltens; es enthält vor allem kognitive Dimensionen des Unterrichts (1966); (3) das System von Oliver und Shaver zur Beschreibung von Erziehungsstilen im Sozialkunde-Unterricht, ein multidimensionales System, das sich auf affektive und kognitive Aspekte der Aktivitäten im Unterricht konzentriert (1966).

Ein System zur Analyse affektiver Dimensionen des Unterrichts

Konzeptueller Rahmen: Mit Flanders' System zur Interaktionsanalyse können vornehmlich affektive und interpersonelle Komponenten der Unterrichtsprozesse erfaßt werden. Flanders behauptet, daß die Art und Weise, in der der Lehrer durch seine verbale Kommunikation das Verhalten von Schülern zu beeinflussen sucht, das wichtigste Kennzeichen der Lehrer-Schüler-Beziehung im Unterricht ist. Seine Bemühungen sind darauf gerichtet, zwei vom Lehrer angewandte kontrastierende Beeinflussungsmethoden zu analysieren:

(a) Direkte Beeinflussung, die »aus den verbalen Äußerungen des Lehrers besteht und die die Handlungsfreiheit einschränken, indem entweder die Aufmerksamkeit auf ein Problem gerichtet wird oder der Lehrer sich auf seine Autorität beruft oder beides zusammen erfolgt.

(b) Indirekte Beeinflussung, die sich in solchen verbalen Äußerungen des Lehrers zeigt, die die Handlungsfreiheit des Schülers durch Ermutigung zur verbalen Teilnahme und Initiative am Unterricht vergrößern« (Flanders 1965, 9).

Kategorien und Methode der Analyse: Das Kategoriensystem von Flanders, das für Situationen im Unterricht entwickelt wurde, in denen Lehrer und Schüler aktiv in verbaler Interaktion stehen, beschreibt verbale Handlungen des Lehrers, die direkten und indirekten Einfluß ausüben. Das Beobachtungsschema besteht aus den folgenden zehn Items:

Flanders Kategorien der Interaktionsanalyse

Lehrer- äußerungen	Indirekte Beeinflussung	1. akzeptiert Gefühle 2. lobt oder ermutigt 3. akzeptiert oder bewertet Ideen der Schüler 4. stellt Fragen
	Direkte Beeinflussung	5. trägt etwas vor 6. gibt Anweisungen 7. kritisiert oder rechtfertigt seine Autorität
Schüleräußerungen		8. Schüleräußerungen – Antwort 9. Schüleräußerungen – Eigeninitiative
		10. Ruhe oder Unruhe

Sieben Kategorien sind zur Klassifikation der Lehreräußerungen, zwei zur Klassifikation der Schüleräußerungen vorgesehen. In die zehnte Kategorie fallen Pausen, kurze Ruheperioden oder Phasen der Unruhe. Von den sieben auf die Lehreräußerungen bezogenen Kategorien repräsentieren die Items 1–4 indirekte, die Items 5–7 direkte Einflußnahme.

Die Daten werden von einem Beobachter während des Unterrichts gewonnen. Am Ende jeder 3-Sekunden-Periode entscheidet der Beobachter, welche der zehn Kategorien die Kommunikationsereignisse dieser Periode am besten beschreibt. Er notiert die Numerierungen der jeweiligen Kategorien, während er gleichzeitig die Kommunikation in der nächsten 3-Sekunden-Periode feststellt. Insgesamt erhebt er also ca. 20 Beobachtungen pro Minute. Seine Notierungen bestehen folglich aus einer von oben nach unten in einer Kolonne geschriebenen Sequenz von Zahlen, so daß die ursprüngliche Sequenz von Ereignissen beibehalten bleibt. Wenn sich eine Veränderung im Unterrichtsablauf ergibt, zieht der Beobachter einen Doppelstrich und vermerkt die Zeit. Handlungsperioden auf diese Weise festzustellen ist somit eine zweite Art der Kategorisierung, die zur Ergänzung des Systems zur Klassifizierung verbaler Aussagen

dient. Flanders nennt fünf Handlungsarten: Einführung von neuem Material, Bewertung von Hausaufgaben oder Testen, Klassendiskussion, Beaufsichtigung und Förderung individueller Unterrichtsarbeit, sonstige übliche Schularbeit.

Nach einer bestimmten Beobachtungsperiode überträgt der Beobachter die Numerierungen, die er in Zeitreihen registriert hat, auf eine 10 x 10-Matrix. Die Numerierungen sind in der Matrix jeweils paarweise registriert. Wenn z. B. die Reihenfolge der Numerierungen auf Grund der Beobachtungen lautet: 10, 6, 7, 5, 4, 8, 1, 4, 8, besteht das erste Paar aus den Numerierungen 10-6, das zweite Paar lautet 6-7, das dritte 7-5 etc. Für das erste Paar, 10-6, wird die Registrierung in die Zelle der Reihe 10, Spalte 6 eingetragen, für das zweite Paar, 6-7, in die Zelle der Reihe 6, Spalte 7. Diese Art der Analyse hilft bis zu einem gewissen Grad, die Sequenz der Ereignisse im Unterrichtsablauf festzuhalten. Die Analyse der Matrixdaten liefert verschiedene Informationen über den Interaktionsprozeß im Unterricht: Der Umfang der Lehrer- und Schüleräußerungen, die Zeitspanne, die in Ruhe oder Unruhe verbracht wurde, das Verhältnis von indirektem und direktem Einfluß, den der Lehrer ausübt (I/D-Quotient), die Sequenzen der verschiedenen Arten von Lehrer- und Schüleräußerungen u. ä.

Zuverlässigkeit: Flanders und seine Mitarbeiter entwickelten systematische Verfahrensweisen zur Schulung der Beobachter, um die Zuverlässigkeit des Kodierens zu sichern. Die Zuverlässigkeit der Beobachtungen, die sich durch die Übereinstimmung der Ergebnisse der einzelnen Beobachter ergibt, wird mit Hilfe des Scott-Koeffizienten geschätzt. In einer von Flanders (1965) berichteten neueren Untersuchung lagen die Reliabilitätskoeffizienten von Scott für geschulte Beobachter ausnahmslos über .86.

Anwendungsbereich: Das System der Interaktionsanalyse wurde häufig bei einer Vielfalt von deskriptiven und experimentellen Untersuchungen angewandt, kürzlich auch bei der Lehrerausbildung und -fortbildung, um den Lehrern eine Rückmeldung über ihr eigenes Verhalten im Unterricht zu ermöglichen.

Ein System zur Analyse kognitiver Dimensionen des Unterrichts

Konzeptueller Rahmen: In ihrer Untersuchung »The Language of the Classroom« (1966) konzipierten Bellack und seine Mitarbeiter den Unterrichtsablauf als eine Art Sprachinteraktion (language game), wobei sie Wittgensteins Auffassung von der Sprache als einem durch Regeln gelenkten verbalen Verhalten übernahmen. Die Grundeinheit zur Beschreibung dieser Interaktionen im Unterricht stellt der pädagogische Impuls

(paedagogical move) dar. Diese Impulse werden entsprechend den pädagogischen Funktionen, die sie in der Unterrichtsdiskussion erfüllen, in vier Hauptkategorien eingeteilt:

Strukturieren: Strukturierende Impulse dienen den pädagogischen Funktionen der Setzung des Kontexts für nachfolgendes Verhalten bei entweder erfolgreich beginnender oder schleppend anlaufender Interaktion zwischen Schülern und Lehrern. Die Lehrer beginnen z. B. häufig eine Unterrichtsstunde mit einem strukturierenden Impuls, mit dem sie die Aufmerksamkeit auf das während dieser Stunde zu diskutierende Thema oder Problem lenken wollen.

Auffordern: Pädagogische Impulse in dieser Kategorie sollen verbale Antworten hervorlocken, um die Schüler, von denen Aufmerksamkeit oder eine physische Reaktion erwartet wird, zu ermutigen. Alle Fragen, Befehle, Anweisungen und Bitten sind zu den Aufforderungen zu zählen.

Antworten: Diese pädagogischen Impulse stehen in einer reziproken Beziehung zu den Aufforderungen und treten nur im Zusammenhang mit ihnen auf. Ihre pädagogische Funktion liegt darin, die durch die Aufforderungen ausgedrückten Erwartungen zu erfüllen; so werden Schülerantworten auf Lehrerfragen als Antworten auf die Aufforderungen klassifiziert.

Reagieren: Diese pädagogischen Impulse werden durch ein strukturierendes, aufforderndes, antwortendes oder der Reaktion vorausgehendes Verhalten veranlaßt, sind aber nicht direkt durch sie verursacht. Aus pädagogischer Sicht dienen diese Verhaltensweisen der Modifikation (durch Klärung, Synthetisierung oder Erweiterung) und/oder der Bewertung (positiv oder negativ) dessen, was vorher gesagt wurde. Reaktionen unterscheiden sich von Antworten folgendermaßen: Während eine Antwort immer direkt durch eine Aufforderung hervorgerufen wird, werden die Reaktionen durch vorausgehende pädagogische Impulse lediglich verursacht; z. B. ist die Bewertung einer Schülerantwort durch den Lehrer als eine Reaktion gekennzeichnet.

Die das verbale Lehrer- und Schülerverhalten beschreibenden Impulse treten im Unterrichtsablauf in bestimmten periodischen Verhaltensmustern oder Kombinationen auf, die als Unterrichtszyklen (teaching cycles) bezeichnet werden. Ein typischer Unterrichtszyklus besteht z. B. aus der Aufforderung eines Lehrers, auf die der Schüler mit einer Antwort reagiert, auf die wiederum eine Lehrerreaktion folgt. Der sequentielle Ablauf der Unterrichtszyklen wird durch eine Markoff-Kette beschrieben.

Die Analyse der sprachlichen Interaktionen im Unterricht wäre ohne Beschreibung des Inhalts der pädagogischen Impulse unvollständig. Zwei grundlegende Arten des Inhalts können unterschieden werden:

(a) Inhalte mit fachspezifischer Bedeutung, die sich auf den zur Diskussion stehenden Lehrstoff beziehen (in dieser Untersuchung: Internationaler Handel);

(b) Inhalte mit unterrichtsspezifischer Bedeutung, die sich auf die zugewiesene Aufgabe, auf die Materialien und Verfahrensweisen im Unterrichtsgeschehen beziehen.

Die Inhalte mit fach- und unterrichtsspezifischer Bedeutung werden zusammen mit ihren assoziierten logischen Bedeutungen beobachtet und registriert, die sich auf die mit den fachspezifischen und unterrichtsspezifischen Inhalten – wie definieren, interpretieren, Fakten darlegen, erklären, Meinungen äußern und begründen – in Verbindung stehenden kognitiven Prozesse beziehen. Wenn ein Schüler also die Fragen eines Lehrers beantwortet, indem er eine Definition von »Preis« gibt, ist der Impuls des Schülers als eine *Antwort* zu kodieren, die fachspezifische Bedeutung als *Preis*, und der logische Prozeß als *definieren*. Zusätzlich wird der Schüler als der Sprecher gekennzeichnet und die Länge seiner Antwort nach der Anzahl der Schreibmaschinenzeilen gemessen.

Kodierungsverfahren: Das Kodieren erfolgt nach Gesichtspunkten des Beobachters, der den pädagogischen Bedeutungsgehalt dem verbalen Verhalten des Redners entnimmt. Die mit der Kodierung beauftragten Personen hören die Bandaufnahmen ab, die vom Unterrichtsgeschehen aufgenommen wurden, und verwerten ebenso die Niederschriften der Tonbandaufnahmen. Jeder pädagogische Impuls wird folgenden Analysekatégorien entsprechend kodiert:

- (1) Sprecher
- (2) Art der pädagogischen Impulse
- (3) fachspezifische Bedeutung
- (4) fachspezifisch-logische Bedeutung
- (5) Antwortenlänge bei (3) und (4)
- (6) unterrichtsspezifische Bedeutung
- (7) unterrichtsspezifisch-logische Bedeutung
- (8) Antwortenlänge bei (6) und (7)

Das folgende Beispiel stellt einen kodierten pädagogischen Impuls dar:

L / STR / IE / ERK / 4 / V / F / 2
 (1) (2) (3) (4) (5) (6) (7) (8)

Diese Angaben werden wie folgt interpretiert: Ein *Lehrer* (1) gibt einen *strukturierenden* (2) Hinweis, indem er etwas über *Import* und *Export* (3) in einer Länge von *vier* (5) Schreibmaschinenzeilen *erklärt* (4); ferner informiert er die Klasse *faktisch* (7) darüber, wie sie im weiteren Verlauf des Unterrichtsgeschehens *verfahren* (6) soll, und zwar in einer Länge von *zwei* (8) Zeilen. Im folgenden wird eine kurze Zusammenfas-

sung der Analysekategorien gegeben, wobei die acht Numerierungen mit den oben genannten übereinstimmen:

1. Sprecher

Lehrer (L)

Schüler (S)

audiovisuelles Gerät (A)

2. Art des pädagogischen Impulses

strukturieren (STR)

auffordern (AUF)

antworten (ANT)

reagieren (REA)

3. Fachspezifische Bedeutungen (Lehrinhalt – in dieser Untersuchung: Internationaler Handel; die Kategorien basieren auf einer Inhaltsanalyse eines Textes über den internationalen Handel, der in den an der Untersuchung teilnehmenden Klassen zugrunde gelegt wurde)

Handel (HA)

Produktions- und/oder Spezialisierungsfaktoren (PSF)

Importe und/oder Exporte (IE)

ausländische Investitionen (AI)

Handelsbarrieren (HB)

Förderung des freien Handels (FFH)

Kriterien, die für den Handel von Bedeutung sind (KHB)

Gründe, die gegen den Handel sprechen (GH)

4. Fachspezifisch-logische Bedeutungen (kognitive Prozesse, die bei der Behandlung von Lehrinhalten miteingeschlossen sind)

Analytische Prozesse

definieren (DEF)

interpretieren (INT)

Empirische Prozesse

formulieren von Faktoren (F)

erklären (ERK)

Evaluative Prozesse

Meinungen äußern (MÄ)

begründen (BEG)

5. Antwortenlänge in Schreibmaschinenzeilen bei 3. und 4.

6. Unterrichtsspezifische Bedeutungen (Faktoren, die auf die Klassenführung bezogen sind)

Zuweisung einer Aufgabe (ZA)

Material (M)

Person (P)

Verfahren (V)

Aussage (A)
 logischer Prozeß (LP)
 Sprachtechnik (ST)
 allgemeine Handlungen (AH)
 verbale Handlungen (VH)
 physische Handlungen (PH)
 kognitive Handlungen (KH)
 emotionale Handlungen (EH)

7. Unterrichtsspezifisch-logische Bedeutungen (kognitive Prozesse, die mit unterrichtsspezifischen Bedeutungen assoziiert sind)
 analytische, empirische, evaluative Prozesse (wie bei 4.)
 Einschätzen (Bezug zur Metakommunikation, gewöhnlich eine evaluative Reaktion)
 positiv (POS)
 zustimmend (ZU)
 wiederholend (WDH)
 genau beschreibend (GB)
 nicht zustimmend (NZU)
 negativ (NEG)
 außer-logische Prozesse
 ausführen (AUS)
 anleiten (ANL)

8. Antwortenlänge in Schreibmaschinenzeilen bei 6. und 7.

Zuverlässigkeit: Um die Zuverlässigkeit des Kodierens zu schätzen, kodierten zwei Gruppen des Forscherteams Stichproben von Protokollen, die nach dem Zufallsprinzip ausgewählt worden waren, und verglichen die Ergebnisse miteinander. Der Prozentsatz der Übereinstimmung wurde für jede der fundamentalen Kategorien des Analyse-Systems berechnet, und zwar hinsichtlich der Anzahl und Dauer der pädagogischen Impulse. Das Ergebnis erbrachte einen übereinstimmend hohen Zuverlässigkeitsgrad für alle Kategorien. Die Übereinstimmung lag zwischen .84 und .96.

Anwendungsbereich: Mit geeigneten Modifikationen in den Kategorien, die als »fachspezifische Bedeutung« klassifiziert sind, liefert dieses Verfahren eine nützliche Technik zur Erforschung des Unterrichtsverhaltens in einer Vielzahl von Fächern der Primar- und Sekundarstufe. Vor kurzem wurden am »Teachers College« Dissertationen beendet, in denen diese Methode der Analyse zur Erforschung des Mathematikunterrichts in der Sekundarstufe und des Leseunterrichts in der Primarstufe verwendet wurde.

*Ein System zur Analyse affektiver und kognitiver Dimensionen
des Unterrichts*

Konzeptueller Rahmen: Im Zusammenhang mit ihrer Untersuchung »Teaching Public Issues in the High School« (1966) entwickelten Oliver und Shaver ein Beobachtungssystem zur Analyse von zwei Unterrichtsstilen: Lehrervortrag und sokratisches Gespräch. Der theoretische Rahmen, in dem diese beiden Unterrichtsweisen definiert sind, beinhaltet drei Dimensionen der Interaktion:

- a) kognitive Dimension,
- b) affektive oder sozial-emotionale Dimension,
- c) Verfahrensdimension.

Was die kognitiven Aspekte betrifft, »tendiert der Lehrervortrag zur Darstellung. Hierbei wird angenommen, daß der Tatbestand eindeutig dargestellt werden kann und daß der Lehrer nur Informationen vorzutragen und zu erklären hat oder eine analytische Struktur vorgibt, durch die die Informationen gegliedert werden können« (Oliver/Shaver 1966, 178). Im Gegensatz dazu »ist der sokratische Stil zweifellos dialektisch. Er setzt voraus, daß das Problem nur in einem entgegengesetzte Aspekte beinhaltenden Kontext geklärt werden kann, in dem verschiedene Standpunkte dargestellt und verteidigt werden« (a. a. O., 178). Hinsichtlich der affektiven Aspekte der Interaktion ist es naheliegend, daß sowohl der sokratisch arbeitende als auch der vortragende Lehrer den Schüler unterstützen muß. Da jedoch »die offene Kontroverse auf dem kognitiven Niveau in den affektiven Bereich übergeht, tendiert die sokratische Methode dahin, daß eine Diskussion stark mit negativen Affekten aufgeladen wird« (a. a. O., 179). Da der Lehrer jedoch im Zentrum des Dialogs mit den Schülern steht, werden keine systematischen Unterschiede zwischen vortragenden und sokratisch unterrichtenden Lehrern in dieser Dimension erwartet.

Kategorien für die Analyse: Oliver und Shaver entwickelten eine Reihe von Kategorien, die für die oben beschriebenen Unterrichtsstile des Lehrervortrags und des sokratischen Gesprächs bedeutsam sind. Ihr Untersuchungsinstrument besteht aus den folgenden drei Gruppen von Kategorien:

Affektive und sozial-emotionale Kategorien

1. Solidarität
2. geringer positiver Affekt
3. Entspannung
4. Spannung
5. geringer negativer Affekt

6. Widerspruch
neutral (keine affektive Mitteilung)

Kognitive Kategorien

7. Hinweis auf Inkonsistenz
8. Beschreibung
9. Evaluation
10. Wiederholung, Zusammenfassung, Herausstellung des Wesentlichen
11. Klärung
12. Hinweis auf Analogien
nicht-kognitiv

Verfahrenskategorien

13. Anordnung aufgabenorientierten Verhaltens
14. Kontrolle abweichenden Verhaltens

Der Beobachter, der dieses Instrument benutzt, muß auf die kognitive oder sonstige für den Unterrichtsverlauf relevante Bedeutung in jeder Aussage oder in jeder Gedankeneinheit schließen, die vom Lehrer zum Ausdruck gebracht wird. Jede derartige einstuftbare Handlung wird in einer kognitiven oder Verfahrenskategorie und in einer affektiven Kategorie eingetragen. Die Funktionen der kognitiven Kategorien werden folgendermaßen beschrieben: »Die primäre Funktion der kognitiven Kategorien (7–12) liegt darin, Fragen über Unterschiede in bezug auf den intellektuellen oder logischen Inhalt in den nach den beiden verschiedenen Stilen stattfindenden Unterrichtsabläufen zu beantworten. Von zentraler Bedeutung für die Unterscheidung zwischen den beiden Stilen ist das Ausmaß, in dem sich die Lehrer mit *deskriptiver Information* bei kontroversen Fragen im Verhältnis zu *Werturteilen* befassen, die über diese Fragen abgegeben werden. Die Kategorien 8 und 9 sind speziell dafür vorgesehen, Unterschiede dieser Art aufzudecken. Die Kategorie 7 soll die Versuche des Lehrers zum Ausdruck bringen, bei den Schülern persönliche Wertkonflikte dadurch hervorzurufen, daß er den Eindruck erweckt, gegensätzliche Urteile in gleichen Situationen abzugeben (das kontroverse Problem und eine Analogie). Die Kategorie 12 (Analogie) hat so eine offensichtliche Bedeutung. Die Kategorien 10 und 11 wurden miteinbezogen, um das kognitive Subsystem zu vervollständigen« (Oliver/Shaver 1966, 291–292).

Zuverlässigkeit: Die Beobachter wurden anhand von Tonbandaufnahmen der Klassendiskussionen sorgfältig in der Anwendung des Verfahrens geschult. Eine Bearbeitung des Chi-Quadrat-Verfahrens wurde zur Schätzung der Übereinstimmung zwischen den Beobachtern verwendet. Alle Chi-Quadrat-Werte lagen deutlich unter dem 5 %-Niveau, das die Wissenschaftler als Kriterium für die Untersuchung angenommen hatten.

Anwendungsbereich: Obwohl dieses analytische Verfahren zur Verwendung im Fach Sozialkunde entwickelt wurde, in dem öffentliche Probleme im Zentrum der Diskussion stehen², nehmen die Autoren an, daß es ebenso zur Analyse des Unterrichtsgeschehens in anderen Fächern, z. B. im naturwissenschaftlichen und Literaturunterricht, verwendet werden kann.

III. Probleme und Konsequenzen in der Entwicklung und im Gebrauch von Systemen zur Analyse des Lehrer- und Schülerverhaltens

In diesem Abschnitt wendet sich die Diskussion zwei Problemen und Konsequenzen zu, die den Erziehern bei der Entwicklung und Anwendung von Beobachtungssystemen im Unterricht begeben.

Beziehungen zwischen den Beobachtungssystemen

Das erste Problem erwächst aus dem Bestreben der Forscher, ein neues Analyse-System für nahezu jede neue Untersuchung zu entwickeln, wohl in der Annahme, die von anderen entwickelten Verfahren seien ungeeignet oder inadäquat. Es mag in der Tat zutreffen, daß Aspekte untersucht werden sollen, für die eine Beobachtungsskala noch nicht entwickelt wurde, oder daß die vorhandenen Systeme nur eine geringe Gültigkeit und Zuverlässigkeit besitzen. Die Entwicklung zahlreicher verschiedener Beobachtungssysteme hat zur Folge, daß den Wissenschaftlern nur sehr wenige Daten zur Verfügung stehen, die einen Vergleich verschiedener Untersuchungen zulassen.

Die umfangreiche Konstruktion von Beobachtungssystemen ist zweifellos das Ergebnis unterschiedlicher Konzeptionen der Forscher. In diesem Zusammenhang sei daran erinnert, daß »in den frühen Stadien der Entwicklung jeder Wissenschaft verschiedene Menschen, die mit dem gleichen Bereich von Phänomenen, nicht aber mit denselben Einzelphänomenen konfrontiert sind, diese unterschiedlich beschreiben und interpretieren« (Kuhn 1962, 17). So kann von den Wissenschaftlern auf dem Gebiet der Unterrichtsforschung – einem Sektor, der sich gerade als wissenschaftlicher Forschungsbereich abzuzeichnen beginnt – nicht erwartet werden, daß sie ihren Forschungen ein gemeinsames theoretisches Konzept zugrunde legen. In diesem Stadium der Entwicklung sind in Widerspruch stehende und konkurrierende Ansätze zur Beschreibung von Unterrichtsabläufen zu erwarten und sogar zu unterstützen.

Jedoch schließt dieser Stand der Dinge nicht die Möglichkeit von Un-

tersuchungen zur Bestimmung von Faktoren oder Dimensionen aus, die den verschiedenen Beobachtungssystemen zugrunde liegen, so daß Differenzen und Ähnlichkeiten zwischen ihnen klarer unterschieden werden können. Ein faktorenanalytisches Verfahren, das dem Vergleich von Beobachtungssystemen zum Zweck der Feststellung von Unterschieden zwischen den ihnen zugrunde liegenden Dimensionen dient, wurde von Gage vorgeschlagen (1969, 1450):

»Die Verhaltensweisen von Lehrern in einer großen Stichprobe können aufgrund vieler Variablen solcher Art, wie sie von Flanders (1965), Smith u. a. (1962), Bellack u. a. (1966), Medley und Mitzel (1958) u. a. entwickelt wurden, erfaßt werden. Dann können die Interkorrelationen der aus den Variablen gewonnenen Punktwerte der Faktorenanalyse unterzogen werden. Die ermittelten Faktoren können die Dimensionen in relativ knapper oder einfach strukturierter Form definieren.«

Wenn vergleichende empirische Untersuchungen von vorhandenen Systemen verfügbar werden, wird es möglich sein, viele Aspekte berücksichtigende Beobachtungsskalen zu konstruieren, die Konzepte einer Vielfalt verschiedener Systeme in sich vereinen. Meux (1967) hat z. B. die Entwicklung eines neuen multidimensionalen Systems vorgeschlagen, um Einheiten mit lernzielorientierten Inhalten und innerhalb dieser Einheiten Perioden, in denen initiiert und reagiert wird, sowie fachspezifische und unterrichtsspezifische Aspekte, Gruppenprozesse und Denkprozesse einzubeziehen. Er ist der Auffassung, daß Unterrichtsstrategien konstruiert werden können, die weit effektiver als die besten der zur Zeit von unseren Lehrern verwendeten sind, wenn Kategorien wie die oben genannten von verschiedenen Beobachtungssystemen in dem vorgeschlagenen multidimensionalen System kombiniert werden. »Diese Möglichkeit neuer Kombinationen von gegenwärtigen Praktiken ist von einigen Autoren übersehen worden, die der Auffassung sind, daß Beobachtungsverfahren uns prinzipiell niemals über die gegenwärtige Praxis hinausbringen werden« (Meux 1967, 549–550).

Mit ähnlichen Zielsetzungen hat Biddle (1967) eine andere Variante eines Idealsystems entworfen, das drei umfassende Konzepte miteinbeziehen soll:

(a) Übergreifende Konzepte für das Verhalten im Unterricht, wobei strukturelle und funktionelle Aspekte des sozialen Systems – dargestellt durch den Klassenverband – mitberücksichtigt werden sollen;

(b) Konzepte für die beobachtbaren, sich gegenseitig beeinflussenden Handlungen der Unterrichtsteilnehmer;

(c) Konzepte für den Sprachgebrauch in der Klasse einschließlich Syntax und Phonologie.

Ob in diesem Entwicklungsstadium der Unterrichtsforschung die Wissenschaftler ihr Augenmerk mehr auf die Konstruktion von multidimensionalen Verfahren wie die von Meux und Biddle vorgeschlagenen richten sollen, oder ob das primäre Bemühen auf die Konstruktion von Beobachtungssystemen von begrenztem Rahmen konzentriert werden soll, ist eine umstrittene Frage. Es ist jedoch wenig sinnvoll, wenn Forscher sich auf das eine oder andere Vorgehen beschränken. In Anbetracht unseres begrenzten Wissens über den Unterrichtsprozeß könnte man vernünftigerweise erwarten, daß beide Ansätze zu einem besseren Verständnis der Unterrichtsprozesse beitragen.

Mögliche Anwendungsbereiche für Beobachtungsverfahren

Die Beobachtung von Lehrer- und Schülerverhalten ist zu einer anerkannten Methode für die Sammlung von Daten in der Unterrichtsforschung geworden. Die mit Hilfe der Beobachtungstechniken gewonnenen Daten geben Anlaß zu der Hoffnung, daß sich unser Verständnis für das komplexe Leben im Klassenraum vertieft. Der zunehmende Gebrauch von Beobachtungstechniken in der Unterrichtsforschung hat Pädagogen, die in der Lehrerbildung und Schulaufsicht tätig sind, zu dem Vorschlag angeregt, die gleichen Beobachtungsverfahren für die Lehrerausbildung und -fortbildung einzusetzen. Dieser Vorschlag wirft mehrere Probleme auf, die wenigstens kurz erörtert werden sollen.

Für den Einsatz von Beobachtungssystemen in der Lehrerausbildung und -fortbildung wird von Simon und Boyer (1967) folgende Begründung gegeben:

»Zunächst stellen die Systeme einen Spiegel für den Lehrer dar, um eine Rückmeldung über sein eigenes Lehrverhalten in den im angewandten System berücksichtigten Dimensionen zu erhalten. Dieses Feedback stellt für die Lehrer eine günstige Gelegenheit dar, ihr Verhalten aufgrund der Informationen darüber, wie sie sich im Unterricht verhalten, selbst zu modifizieren. Ferner – und dies ist eventuell ein noch wichtigerer Aspekt – sind viele dieser Systeme aufgrund einer theoretischen Dimension konstruiert worden, die Verhaltensweisen beinhaltet, die bei der Verwendung im Unterricht eine Hilfe zur Förderung der Entwicklung der Schüler darstellen sollen, wie es zur Zeit in den Schulen in der Weise noch nicht der Fall ist. Wenn ein Lehrer eines dieser Systeme anwendet, erhält er ein Feedback über die Verhaltensweisen, die er *nicht* zeigt, wie auch über die, die er zeigt. Dem Lehrer bietet sich somit eine Möglichkeit, neue Verhaltensweisen zu lernen und auf diese Weise sein Verhaltensrepertoire mit Hilfe von Methoden zu erweitern, die den Lehrern im allgemeinen nicht zur Verfügung stehen« (Simon/Boyer 1967, 19–20).

Kaum einer würde die Wichtigkeit konzeptueller Instrumente für die

Ausbildung und Fortbildung von Lehrern leugnen, die sie befähigen, den Unterrichtsprozeß zu analysieren und auf diese Weise das Verständnis für ihre Aufgaben als Lehrer zu vertiefen. Wenn die verschiedenen Analyseverfahren, die im Zusammenhang mit Forschungsprojekten entwickelt wurden, den Lehrern als Instrumente vorgestellt werden, mit denen sie den Unterricht von verschiedenen Perspektiven her betrachten können, sind die Lehrer in der Lage, aus den Analysen große Vorteile zu ziehen. In einem kürzlich erschienenen Bericht der American Association of Colleges for Teacher Education (1968) wird der Vorschlag gemacht, eine Phase in der Lehrerausbildung einer solchen Analyse des Unterrichts zu widmen.

Jedoch ist ein Wort der Vorsicht angebracht. Die Versuchung, Beschreibungen des Lehrerhaltens in Vorschriften umzuwandeln, die von Lehrern zu befolgen sind, muß vermieden werden. Denn bisher konnten nur wenige Korrelationen zwischen den Leistungen der Schüler und vielen Unterrichtsvariablen in den vorhandenen Beobachtungssystemen festgestellt werden. Obwohl die Beobachter, die die verschiedenen Analyseverfahren anwenden, dem Lehrer zuverlässige Beschreibungen vieler schwieriger Dimensionen der Unterrichtsabläufe liefern können, dürfen sie selbstverständlich nicht vorschreiben, wie Lehrer auf der Basis dieser Beschreibungen unterrichten sollen. Denn Entscheidungen hinsichtlich des sich daraus ergebenden Unterrichtsverhaltens sind vom Wissen um Konsequenzen verschiedener Verfahren und vom Urteil hinsichtlich des Wertes dieser Konsequenzen abhängig.

GRAHAM A. NUTHALL¹

*Ausgewählte neuere Untersuchungen
zur Unterrichtsinteraktion und zum Lehrverhalten*

Ein kritischer Bericht

Den meisten Lesern wird aufgefallen sein, daß in letzter Zeit die Zahl der Untersuchungen über das Lehrerverhalten und die Lehrer-Schüler-Interaktion im Unterricht zugenommen hat. Einige neuere Untersuchungen machen das Ausmaß dieser Zunahme deutlich (Lawrence 1966; Amidon/Simon 1965; Gage/Unruh 1967; Biddle 1967). Obwohl derartige Untersuchungen nicht neu sind, haben viele Studien noch immer Versuchscharakter. Daher wurden bisher nur wenige von ihnen in größerer Auflage veröffentlicht; und es wurde selten versucht, die bisherigen oder noch erreichbaren Ergebnisse dieser neuen Forschungsrichtung kritisch einzuschätzen. Dieser Beitrag soll einen kritischen Überblick über eine Auswahl neuerer Untersuchungen geben, die für diesen Forschungszweig sehr wichtig zu sein scheinen. Besondere Aufmerksamkeit soll dabei jenen Befunden zukommen, die unsere Kenntnis über die Beziehungen zwischen den Verhaltensweisen des Lehrers und den Lernprozessen bzw. Leistungen des Schülers verbessern.

Von ein paar Ausnahmen abgesehen, wurde das Lehrerverhalten und die Unterrichtsinteraktion nach dem klassischen Modell aller wissenschaftlichen Untersuchungen erforscht. Nach einer Reihe erkundender und abtastender Felduntersuchungen führten gezielte Versuche, Taxonomien des Unterrichtsverhaltens zu erarbeiten, zu mehreren hochentwickelten deskriptiven Untersuchungen über Lehrer und Schüler *in situ*. Im Anschluß an diese ersten »ökologischen« Überblicke wurde dann versucht, die beobachteten Lehrvariablen mit der gemessenen Schülerleistung zu korrelieren. Und neuerdings gibt es auch sorgfältig kontrollierte Untersuchungen halbexperimenteller Art. Da die Wirklichkeit jedoch stets verworrener ist als das Modell, finden immer noch interessante, wichtige Entwicklungen auf jeder der verschiedenen Entwicklungsstufen statt.

Daher werden die Untersuchungen, die analysiert werden sollen, unter drei Rubriken betrachtet: Zuerst soll die Entwicklung der Methoden behandelt werden, die zur Beobachtung und Analyse des gesamten Unter-

richtsgeschehens dienen. Dabei wird nicht versucht, alle der mehr als dreißig verschiedenen Systeme zu behandeln; es sollen lediglich einige interessante neue Verfahren sowie einige Modifikationen alter Systeme beschrieben werden. Danach sollen einige Ergebnisse der Unterrichtsbeobachtung dargestellt werden; dabei soll besonders berücksichtigt werden, wie weit sie zu unserem Verständnis der Ursachen des Lernens beitragen. Im letzten Teil werden dann Untersuchungen behandelt, in denen versucht wird, Schülerleistung in Beziehung zu Lehrerverhalten oder Unterrichtsinteraktion zu bringen. Sie reichen von einfachen Korrelationsuntersuchungen der Beziehungen in normalen Klassen, die ihren üblichen Unterricht erhielten, bis hin zu Versuchen, die Lehrvariablen experimentell zu manipulieren.

Verfahren zur Analyse von Unterrichtsverhalten

Revision eines alten Systems

Das bei weitem am häufigsten gebrauchte Verfahren für die Analyse des Unterrichtsverhaltens ist das Flanderssche Verfahren der Interaktionsanalyse (Amidon/Flanders 1963). Nach diesem Verfahren erstellt ein Beobachter ein fortlaufendes Protokoll der Unterrichts-Aktivitäten von Lehrer und Schülern. Alle drei Sekunden (bei wichtigen Veränderungen noch öfter) notiert der Beobachter die Kategorie, die das augenblickliche Geschehen am besten beschreibt. Das Ergebnis ist ein kontinuierliches Protokoll, das in bezug auf die relative Häufigkeit verschiedener Verhaltensarten und hinsichtlich der paarweisen Verhaltensfolgen anhand einer zwei-dimensionalen Matrix analysiert werden kann. Der ursprüngliche Zweck der Interaktionsanalyse bestand darin, Schätzwerte zu erhalten, bis zu welchem Grad die Lehrer ihre Schüler »indirekt« oder »direkt« beeinflussen. Das ursprüngliche, von dieser Analyse abgeleitete Maß (I/D Verhältnis) sollte diesen Aspekt des »affektiven Klimas« (affective climate) im Unterricht ausdrücken.

Während dieses Verfahren in seiner ursprünglichen Form noch immer sehr häufig benutzt wird, haben auch mehrere Verbesserungsversuche stattgefunden. In seinem Bericht über neue Entwicklungen bei der Interaktionsanalyse wies Amidon (1966) auf das Fehlen »kognitiver« Kategorien hin. Er beschrieb eine Reihe von Kategorien zur Identifikation verschiedener kognitiver Ebenen bei Fragen und Antworten. Teilweise basierten sie auf der Arbeit Aschners und Gallaghers (Aschner/Gallagher u. a. 1965), die auf Guilfords fünf kognitive Prozesse zurückgreift (Gedächtnis, Er-

kenntnis [cognition], konvergierende und divergierende Produktion und Bewertung). Bisher gibt es offenbar noch keine Untersuchungen, die eine Modifikation des Flandersschen Verfahrens in bezug auf kognitive Variablen benutzen; einige empirische Untersuchungen deuten darauf hin (Furst 1967), daß das Konzept des »kognitiven Klimas« im Unterricht weiter ausgearbeitet werden sollte. Ein solcher Meßwert des »kognitiven Klimas« im Unterricht könnte zeigen, welches »kognitive Niveau« (cognitive level) für einen Lehrer charakteristisch ist. L. Siegel und L. C. Siegel (1967) entwickelten ein ähnliches Konzept für das »intellektuelle Klima« (intellectual climate).

Eine weitere Modifikation des Flandersschen Verfahrens, typisches Lehrerverhalten zu messen, versuchte Honigman (1968). Er berichtete über die Entwicklung eines dreidimensionalen Kategorien-Systems und fügte der ursprünglichen affektiven Dimension eine kognitiv-inhaltliche (cognitive-substantive) sowie eine Verfahrensdimension (class management) hinzu. Dieses System wird als »Multidimensionale Analyse der Unterrichtsinteraktion« bezeichnet (Multidimensional Analysis of Classroom Interaction; MACI). Dieser Versuch, die Skala der Verhaltensweisen, auf die der Beobachter zu achten hat, auszuweiten, hat Honigman offenbar dazu gebracht, einige einfallsreiche Aufzeichnungstechniken zu entwickeln, die das Protokollieren erleichtern sollen. Honigmans Ausführungen sind auch bemerkenswert, weil sie einen Versuch beschreiben, die Validität (concurrent validity) der »Multidimensionalen Analyse der Unterrichtsinteraktion« zu prüfen. Bei dieser Untersuchung wurden 75 Lehrer, die an einem Sommerkurs teilnahmen, nach einem Zufallsstichprobenverfahren in drei gleich große Gruppen geteilt. Die eine Gruppe sah die Fernsehaufzeichnung einer Unterrichtsstunde, die zweite las nur das nach der »Multidimensionalen Analyse der Unterrichtsinteraktion« von einem Beobachter geführte Protokoll derselben Stunde, und die dritte Gruppe las ein Protokoll, das nach dem Flandersschen Verfahren geführt worden war. Der Kriteriumswert (criterion measure) wurde auf Grund der Angaben aller drei Gruppen in einem Fragebogen ermittelt, der aus 32 Fragen über verschiedene (affektive, kognitive und Verfahrens-)Aspekte der Stunde bestand. Folgende Hypothese wurde aufgestellt: Die Antworten der Gruppe, die das nach Honigmans »Multidimensionaler Analyse der Unterrichtsinteraktion« angefertigte Protokoll gelesen hatte, müßten mehr Ähnlichkeiten mit den Antworten der Gruppe haben, die die Life-Übertragung der Stunde gesehen hatte, als mit denen der Gruppe, die das nach dem Flandersschen Kategorien-System angefertigte Protokoll gelesen hatte. Ein solches Ergebnis war bei dem Teil des Fragebogens bestimmt zu erwarten, der sich auf die kognitive Dimension der Unterrichtsstunde bezog. Die Er-

gebnisse zeigten jedoch, daß die »Multidimensionale Analyse der Unterrichtsinteraktion« zwar besser die affektiven Aspekte, nicht aber die kognitiven und die Verfahrens-Aspekte der Unterrichtsstunde herausarbeitete. Es konnte noch nicht nachgewiesen werden, daß Honigmans Beobachtungssystem wirklich multidimensional ist.

Die Entwicklung neuer Systeme

Während nach dem Flandersschen Verfahren der Interaktionsanalyse und Honigmans »Multidimensionaler Analyse der Unterrichtsinteraktion« der Beobachter seine Aufzeichnungen in gleichmäßigen Zeitabständen machen muß, gibt es andere Analyse-Systeme, nach denen Beobachtungen in natürlichen Zeiteinheiten oder zu natürlichen Übergängen im Unterrichtsverhalten gemacht werden. Weil die nötige Identifikation und Analyse dieser Zeiteinheiten oder Übergangspunkte oft recht schwierig ist, werden in der Regel Tonband- oder Videorecorder-Aufzeichnungen des untersuchten Verhaltens gemacht. Bellack und andere (1966) berichteten z. B. von einer genauen Analyse der von ihnen als »Unterrichtszyklus (teaching cycles) bezeichneten Einheit. Die Analyse beruht auf Transkriptionen von Tonbandaufnahmen mehrerer Unterrichtsstunden, die über das gleiche ökonomische Thema von verschiedenen Lehrern gehalten wurden. Ein »Unterrichtszyklus« ist von allen natürlichen Einheiten im Ablauf der Diskussion im Unterricht am leichtesten als Einheit zu identifizieren; er besteht in der »Frage-Antwort-Reaktions-Sequenz«, oder in einer ihrer vielen Varianten. Dies war auch die Grundeinheit einer Untersuchung von Nuthall und Lawrence (1965).

In zwei sehr verschiedenen Untersuchungen wird von der Entwicklung neuer analytischer Systeme berichtet, die auf natürlichen Einheiten beruhen. Die erste Untersuchung beschreibt den Versuch, mit Hilfe eines brauchbaren psychologischen Konzepts das Unterrichtsverhalten zu erforschen. In der anderen Untersuchung wird der weit komplexere Versuch gemacht, Methoden zur Aufzeichnung aller denkbaren wichtigen Verhaltensweisen im Unterricht zu finden. MacDonald und Zaret (1967) berichteten von einem Versuch, den Begriff der »Offenheit« in den menschlichen Beziehungen (wie er von Rogers und anderen in der Psychotherapie entwickelt wurde) auf die Interaktion im Unterricht zu beziehen. Sie entwickelten ein Verfahren, den Grad an Offenheit in den protokollierten verbalen Reaktionen der Lehrer gerade zu den Zeitpunkten im Unterrichtsgespräch einzuschätzen, zu denen die Lehrer Antworten von Schülern kommentieren und die weitere Diskussionsrichtung bestimmen. Sie stellten die Hypothese auf, daß die Lehrer, die sich während dieser wichtigen Situati-

on in der Unterrichtsdiskussion offener verhielten, mehr produktive Antworten ihrer Schüler provozieren würden. Die Lehrer dagegen, die verschlossener sind oder sich stärker an den Rollenerwartungen orientieren, erhalten dadurch wahrscheinlich weniger produktive Antworten von ihren Schülern und erhöhen gleichzeitig auch die Zahl der nur reproduktiven Antworten. Durch Tonbandaufzeichnungen von Unterrichtsstunden in der Sozialkunde (social studies) in neun Klassen einer Elementarschule fanden sie in acht der neun Klassen statistisch signifikante Beziehungen zwischen dem Grad der Offenheit des Lehrers und der Zahl der produktiven Antworten der Schüler. Obwohl sicherlich noch mehr sorgfältig kontrollierte Daten nötig sind, scheint doch der Wert des Begriffs der Offenheit für die Unterrichtsinteraktion nachgewiesen worden zu sein. Es wäre höchst interessant zu wissen, ob es zwischen Flanders' Begriff der »Indirektheit« und MacDonalds und Zarets Begriff der »Offenheit« eine enge Beziehung gibt.

In ihrem Bericht über ihre umfangreiche Untersuchung über Unterrichtsaktivitäten beschrieben Biddle und Adams (1967) die Entwicklung eines Beobachtungssystems, das wohl das umfangreichste aller vorliegenden Systeme ist. Im Gegensatz zu dem System von MacDonald und Zaret beruht das System von Biddle und Adams nicht auf einer spezifischen Theorie oder einem theoretischen Konstrukt, obwohl es etwas soziologisch orientiert ist. Es scheint der wohlüberlegte Versuch zu sein, die Voraussetzungen dafür zu schaffen, daß möglichst viele verschiedene, leicht identifizierbare Aspekte aller Aktivitäten im Unterricht in einem Protokoll aufgenommen werden können. Der Umfang des Systems hängt offenbar davon ab, ob eine Videorecorder-Anlage vorhanden ist, die eine detaillierte Analyse verbaler und verhaltensbezogener Aspekte des Unterrichtsgeschehens ermöglicht. Das System von Biddle und Adams basiert auf der Unterscheidung zwischen strukturellen und funktionellen Aspekten der Unterrichtsaktivitäten. Der strukturelle Aspekt umfaßt Positionen und Rollen aller Unterrichtsteilnehmer (was sie tun, mit wem, wo, usw.). Der funktionelle Aspekt umfaßt die Art des Inhalts (oder der Bedeutung) der Interaktion und die Art der Behandlung dieses Inhalts. Um also Strukturen beschreiben zu können, wird identifiziert: Wer spricht (Sender), zu wem (Empfänger), und wer hört zu (Zuhörer), also die Kommunikationsstruktur und Gruppierung aller Individuen und das Verhältnis der Individuen zu den physikalischen Dimensionen der Klasse. Um Funktionen zu beschreiben, identifizieren sie die drei in Honigmans »Multidimensionaler Analyse der Unterrichtsinteraktion« enthaltenen Dimensionen, und differenzieren weiter zwischen relevanter und irrelevanter Thematik. Eine Sonderkategorie erfaßt nicht-sprachliche Vorgänge (z. B. Drill, praktische Übungen, usw.).

Im allgemeinen hat man bisher dazu geneigt, möglichst viele verschiedene Aspekte des Unterrichtsverhaltens zu erfassen, ohne die feineren Unterscheidungen und Unterordnungen zu berücksichtigen, die unter jedem dieser Aspekte vielleicht zu treffen wären. Das Ergebnis ist ein System, das in gewissem Sinne unbegrenzt erweiterbar ist, vorausgesetzt, daß die Video-recorder-Aufnahmen der Unterrichtsvorgänge ausreichend präzise sind.

Die Zukunft der Systeme zur Analyse der Unterrichtsinteraktion

Während der frühen Stadien der Entwicklung der ersten wichtigen Systeme zur Analyse des Unterrichtsverhaltens äußerten mehrere Autoren die Befürchtung, die Vielzahl verschiedener Systeme könne zu einer beträchtlichen Verwirrung führen. Wenn jeder sein eigenes Analyse-System aufbaue, so wurde befürchtet, werde eine gewaltige Informationssammlung über Unterrichtsverhalten zustande kommen, die nicht systematisiert werden könne und aus der noch nicht einmal die einfachsten Thesen über Unterrichtsgeschehen entwickelt werden können, die die meisten Forscher benötigen. Schon jetzt könne man die Situation verwirrend finden und vielleicht den Wert so vieler scheinbarer Wiederholungen in Frage stellen. Dazu führt Komisar aus:

Die Produktion »neuer« Kategoriensysteme bringt uns rasch dem Chaos näher. Scheinbar ist keiner der Forscher willens oder in der Lage, uns zu erklären, *warum* gewisse Kategorien gewählt werden oder in welchem Verhältnis die Kategorien eines Forschers zu denen eines anderen stehen (Komisar 1968, 22).

Andererseits wies Biddle darauf hin, daß jeder Forscher unbedingt sein eigenes System entwickeln müsse, um den Bedingungen seines eigenen Forschungsprojekts gerecht zu werden, zumal so vielen der neuen Systeme jegliche zuverlässige oder systematisch-theoretische Grundlage fehle (Biddle 1968, 31).

Mit dem Erscheinen einer Anthologie der 26 »bekanntesten und gebräuchlichsten Beobachtungssysteme für den Unterricht« (Simon/Boyer 1967) ist hoffentlich ein Wendepunkt erreicht worden². In Zukunft wird der Konstrukteur eines »neuen« Kategoriensystems sein Werk theoretisch rechtfertigen müssen. So wird vielleicht die nötige Aufmerksamkeit auf die Erklärung und das Verständnis des Unterrichtsverhaltens gelenkt. Wie in den folgenden Abschnitten dieses Beitrags gezeigt werden soll, ist trotz der vielen protokollierten Details über die Geschehnisse in zahlreichen Unterrichtssituationen immer noch sehr wenig über die Ursachen oder Wirkungen der Geschehnisse bekannt.

Beziehungen zwischen Beobachtungssystemen

Es wurde bereits mehrmals versucht, die Beziehungen zwischen verschiedenen Systemen zu erforschen. In einigen Untersuchungen wandte man dazu beispielsweise zwei verschiedene Systeme parallel an. Medley und Hill (1968) berichteten über die Ergebnisse eines Vergleichs des Flandersschen Verfahrens der Interaktionsanalyse mit ihrer eigenen, unlängst abgeänderten Fassung des »Beobachtungsplans und -protokolls« (Observation Schedule and Record; OSCAR 4V). Bei dieser Untersuchung wurden 70 Lehrer und Lehrerinnen der Sekundarschule während ihres ersten Berufsjahres von je zwei Beobachtern mindestens viermal systematisch beobachtet. Jedesmal protokollierte ein Beobachter nach dem Flandersschen System und der andere nach OSCAR 4V. Entsprechend diesen parallelen Protokollen wurden für jeden Lehrer insgesamt 75 Meßwerte interkorreliert (38 nach Flanders und 37 nach OSCAR). Da es offenbar schwierig ist, in eine so große Korrelationstabelle (75×75) Sinn hineinzubringen, wurde eine Faktorenanalyse durchgeführt. Es gelang, zehn Faktoren zu identifizieren, die die Unterschiede zwischen den Lehrern beschrieben. Von diesen zehn Faktoren wurden fünf mit beiden Systemen gemessen, drei nur mit OSCAR, und zwei nur mit dem Flandersschen System. Die beiden Systeme scheinen sich bis auf einige unabhängige Merkmale zu überschneiden. Es ist allerdings sehr schwierig, die Besonderheit jedes Systems aus den untersuchten Faktoren abzuleiten.

In einer von Pearson³ geleiteten Untersuchung wurde das Flanderssche Verfahren als zusätzliches Verfahren zur Analyse der »evaluativen Abschnitte« (evaluative ventures) bei Gesprächen in Klassen der Sekundarstufe I (intermediate school) benutzt. Ein »evaluativer Abschnitt« ist einer von neun Abschnitten oder themen-zentrierten Gesprächseinheiten (topic-centered units), die Smith und Meux (1967) identifiziert haben, unter dem die Diskussion über eine Wertfrage oder die Bedeutung eines Sachverhalts verstanden wird. Die vorläufige Analyse der Daten weist darauf hin, daß die Meßwerte des Flandersschen Systems mit der Zahl der evaluativen Gespräche, mit ihrer Dauer und ihrer relativ logischen Komplexheit in Beziehung stehen. Mit anderen Worten: Das »affektive Klima« ist erwartungsgemäß nicht unabhängig von den logischen und semantischen Aspekten der Diskussion in einer Schulklasse.

Durch Beobachtungssysteme gewonnene Ergebnisse

Einige neuere Untersuchungen brachten Daten über die Art des Lehrverhaltens und der Unterrichtsinteraktion; zu ihnen gehören auch die Untersuchungen von Bellack u. a. (1966) sowie von Smith und Meux (1967), deren Ergebnisse sich auf umfassende Forschungen stützen, die hauptsächlich verbale Interaktionen, semantische und thematisch-inhaltliche Aspekte betreffen.

Die Sequenz der Unterrichtsinhalte

Smith und Meux (1967) berichteten über die Ergebnisse eines Versuchs, die Organisation und sequenzielle Anordnung der Unterrichtsinhalte mit Hilfe von Aufzeichnungen von Diskussionen in Sekundarschulen zu analysieren. Sie definierten zunächst den Begriff »Unterrichtsstrategie« (teaching strategy) als eine besondere Sequenz von Teilthemen innerhalb des vorgegebenen Unterrichtsthemas. Davon ausgehend analysierten und beschrieben sie die »Strategien« in acht verschiedenen, zuvor identifizierten Kategorien. Diese acht Kategorien waren folgende: Ursachen, Begriffe, Bewertungen, Interpretationen, Verfahren, Regeln, Gründe und besondere Informationen (causes, concepts, evaluations, interpretations, procedures, rules, reasons and particular information). Dadurch fanden Smith und Meux z. B. heraus, daß Diskussionen über Begriffe zu der häufigsten Art von Diskussionen gehören. Nachdem sie diejenigen Teile des Gesprächs isoliert hatten, in denen ein genannter Begriff Mittelpunkt der Diskussion war, isolierten sie auch die Informationen über Begriffe, in denen die Begriffe ganz allgemein beschrieben oder besprochen wurden. Die kurzen Abschnitte der Diskussion, in denen nur eine einzige Information über einen Begriff beschrieben oder besprochen wurde, bezeichneten sie als »Begriffs-Impulse« (conceptual moves). So betrachtet, besteht der Begriff Unterrichtsstrategie also aus einer bestimmten Reihenfolge verschiedener »Begriffs-Impulse«.

Drei Hauptarten der »Begriffs-Impulse« ließen sich auf Grund umfangreicher Tonbandaufzeichnungen von Unterrichtsstunden in Sekundarschulen identifizieren: (1) Gesprächsteile, in denen Informationen über *Beispiele* des Begriffs besprochen wurden (»Beispiels-Impulse«) (instantial moves); (2) Gesprächsteile, in denen der Begriff mit anderen *verglichen* oder kontrastiert wurde (»Vergleichs-Impulse«) (comparative moves); und (3) die Gesprächsteile, in denen die Kriterien für den Begriff festgesetzt oder direkt *beschrieben* wurden (»Beschreibungs-Impulse«) (descriptive moves). Tabelle 1 enthält eine Liste dieser verschiedenen Arten der Impulse.

Tabelle 1
Impulsarten in »Begriffs-Abschnitten«*
(Types of Moves in Concept Ventures)

<p>I Beschreibungs-Impulse</p> <ol style="list-style-type: none"> 1. Beschreibung eines Merkmals 2. Definition ausreichender Bedingungen 3. Klassifikation des Begriffs 4. Klassifikatorische Beschreibung 5. Definition der Beziehungen zwischen Merkmalen 6. Zerlegung in Einzelteile <p>II Vergleichs-Impulse</p> <ol style="list-style-type: none"> 7. Vergleich durch Analogie 8. Vergleich durch Differenzierung 9. Vergleich von Beispielen eines Begriffs 	<p>III Beispiel-Impulse</p> <ol style="list-style-type: none"> 10. Beschreibung positiver Beispiele 11. Aufzählung aller Beispiele 12. Beschreibung eines Nicht-Beispiels 13. Beschreibung der Produktion von Beispielen 14. Bestätigung einer Sache als Beispiel <p>IV Gebrauchs-Impulse (usage moves)</p> <ol style="list-style-type: none"> 15. Metaunterscheidungen bei Verwendung eines Begriffs
--	---

* aus: B. O. Smith, *The Strategies of Teaching* (1967)

Verschiedene Lehrer wurden hinsichtlich der in ihren Klassen üblichen Sequenzen der Impulse beschrieben und verglichen. Dabei stellte sich z. B. heraus, daß eine wechselnde Sequenz der Erörterung von Kriterien und die positiven Beispiele eines Begriffs eine der verbreitetsten Unterrichtsstrategien ist.

Unterschiede im Lehrerverhalten beim gleichen Curriculum

Die Praxis vieler Forscher, irgendwelche leicht zugänglichen Stichproben von Lehrern oder Klassen zu untersuchen, hat einige Unzufriedenheit geschaffen. Neuere Untersuchungen berichten dagegen von sorgfältig erhobenen Daten. Danach kann nun einiges darüber ausgesagt werden, wie das Verhalten im Unterricht durch gewisse, als unabhängig bekannte Variablen verändert werden kann. Gallagher (1966) berichtete über einen interessanten Versuch, die Unterschiede im Unterrichtsverhalten von Lehrern zu beschreiben, die dasselbe Thema desselben Curriculum mit vergleichbaren Klassen behandeln. Gallagher und andere (1966) entwickelten ein System der »Themenanalyse« (topic analysis), das die unterschiedlichen Gesprächsgänge über das gleiche Thema in verschiedenen Klassen identifizieren sollte. Als »Thema« wurde der Teil der Diskussion bezeichnet, in dem »sich die Klassendiskussion auf eine gegebene Handlung, einen Begriff

oder ein Prinzip konzentriert«. Jedes Thema wurde nach der Art des Inhalts klassifiziert (Informationswissen, Verfahrensfertigkeiten), nach dem Grad der Abstraktheit in der Diskussion (Daten, Begriff, Generalisierung) und nach dem Denkstil, der dem Gespräch zu entnehmen ist (Beschreibung, Expansion, Erklärung, Rechtfertigung, Bewertung usw.). Mit diesem System wurden die Tonbandprotokolle der Diskussionen in den Klassen von sechs Biologielehrern analysiert, die das Thema »Photosynthese« mit Hilfe der Materialien der neuen Biological Science Curriculum Study (»blue book version«) behandelt hatten. Gallagher fand signifikante Unterschiede zwischen den Lehrern bezüglich des Inhalts der Themen und im Grad der Abstraktheit, nicht aber im Denkstil. Auch der Anteil der Zeit, in der der Lehrer sprach, variierte signifikant von Klasse zu Klasse. Signifikante Unterschiede gab es auch für drei der vier Denkart: Beschreibung, Erklärung und Expansion. Gallaghers Analyse machte deutlich, daß auch bei konstant gehaltenem Curriculum eine große Vielfalt in den behandelten Themen und in der Art ihrer Behandlung auftritt.

*Unterschiede im Unterrichtsverhalten in bezug auf das Schuljahr,
das Alter des Lehrers und die Unterrichtsinhalte.*

Weitere Informationen über die Art der Unterschiede zwischen Lehrern liefert der Bericht einer Untersuchung von Biddle und Adams (1967). Diese Wissenschaftler machten Fernsehaufzeichnungen von 32 Unterrichtsstunden, die von 16 Lehrern in drei verschiedenen Schuljahren (im ersten, sechsten und elften) gehalten wurden. Diese Unterrichtsstunden wurden gleichmäßig auf Mathematik und Sozialkunde verteilt; ebenso bildete man aus den Lehrern zwei gleich große Gruppen (älter oder jünger als 30 Jahre). Die Zusammenfassung einiger ausgewählter Ergebnisse soll hier wiedergegeben werden:

(1) *Schuljahr*. Viele Unterschiede ergaben sich als eine Funktion des Schuljahrs. Die ersten Klassen verbrachten viel Zeit mit Handlungen (wie Singen, Vorlesen) und mit Anweisungen zur Organisation des Unterrichts. Die sechsten Klassen wirkten am wenigsten traditionell, da es in ihnen mehr Gruppenarbeit und viele Interaktionen gab. Erwartungsgemäß waren die elften Klassen stärker thematisch orientiert und verbrachten mehr Zeit für die intellektuelle Auseinandersetzung mit relevanten Unterrichtsinhalten.

(2) *Alter des Lehrers*. Die Klassen der älteren Lehrer schienen insofern traditioneller zu sein, als ihre Handlungen öfter vom Lehrer gesteuert wurden. Sie beschäftigten sich auch mehr damit, Informationen weiterzugeben, als zu intellektuellen Auseinandersetzungen (z. B. zu klärenden und bewer-

tenden Diskussionen u. ä.) hinzufügen. Die Klassen mit jüngeren Lehrern zeigten weniger Lehrer-Kontrolle; hier waren die Lehrer öfter mit kleinen Arbeitsgruppen beschäftigt oder überhaupt nicht mit einbezogen.

(3) *Fachspezifische Inhalte.* Die Diskussion in den Mathematikstunden neigte dazu, in den unteren Klassen einen geringeren und in der elften Klasse einen höheren »intellektuellen« Gehalt zu haben als in den entsprechenden Unterrichtsstunden der Sozialkunde. Bei den letzteren gab es dagegen mehr Diskussionen über irrelevante Themen.

Interessante Zusatzbefunde:

(4) Wenn der Lehrer als Sender (Redner) vor der ganzen Klasse spricht, besteht der Inhalt seiner Rede aller Wahrscheinlichkeit nach in der Vermittlung einfacher Informationen über das Thema oder aus Anweisungen zur Organisation des Unterrichts.

(5) Wenn der Lehrer etwas zum Thema vorträgt, steht er wahrscheinlich vor der Klasse; wenn er dagegen etwas zur Organisation des Unterrichts äußert, steht er oft seitlich zur Klasse.

(6) Die Schüler, die in der Klasse entlang der Mittellinie sitzen, nehmen eher an der Diskussion teil und werden häufiger vom Lehrer angesprochen. Die Schüler, die weiter von der Mittellinie weg sitzen, sprechen entschieden seltener und werden auch seltener angesprochen.

Diese kurze Zusammenfassung kann den vielen Daten, die Biddle und Adams sammelten, zwar kaum gerecht werden; sie soll jedoch auf die allgemeine Bedeutung dieser umfassenden Dokumentation struktureller und funktioneller Aspekte des Unterrichtsverhaltens hinweisen.

Verbale Lehrerreaktion

Deutlich unterscheidet sich von der Studie von Biddle und Adams eine Untersuchung, über die Zahorik (1968) berichtete. Während Biddle und Adams möglichst viele Aspekte des Unterrichtsverhaltens berücksichtigten, ohne ihrer Analyse dabei eine bestimmte Richtung zu geben, richtete Zahorik seine Aufmerksamkeit auf einen einzigen Aspekt, der zweifellos von beträchtlicher Bedeutung ist. Zahorik entwickelte ein System zur Analyse und Bewertung der Reaktionen des Lehrers auf Schülerantworten. Zahoriks Daten wurden aus Tonbandprotokollen gewonnen, die in acht Klassen des dritten und in sieben Klassen des sechsten Schuljahrs aufgezeichnet worden waren. Man hatte die Lehrer gebeten, den Inhalt eines aktuellen Nachrichtenmagazins zu behandeln, wodurch man eine gewisse Kontrolle über den Unterrichtsstoff in den beiden protokollierten Stunden erhielt. Die erste Stunde sollte eine Einführung in die Lektüre des Nachrichtenmagazins geben, und die zweite Stunde sollte dann daran anknüpfen.

Aus den Reaktionen der Lehrer auf die Antworten ihrer Schüler bildete Zahorik 14 verschiedene Kategorien. Sie werden in Tabelle 2 aufgeführt:

Tabelle 2
Kategorien der verbalen Lehrer-Reaktionen *

-
1. Lobende Bestätigung
 2. Tadelnde Verneinung
 3. Lobende Bestätigung + Tadelnde Verneinung
 4. Positive Antwort
 5. Negative Antwort
 6. Positive Antwort + Negative Antwort
 7. Positive Erklärung
 8. Negative Erklärung
 9. Erweiterung der Reaktion: Entwicklung
 10. Erweiterung der Reaktion: Verbesserung
 11. Wiederholung der Aufforderung: mehrere Antworten
 12. Wiederholung der Aufforderung: eine Antwort
 13. Fortsetzung des Unterrichts: neues Thema
 14. Verschiedenes
-

* aus: J. A. Zahorik, Classroom Feedback Behavior of Teachers, *Journal of Educational Research* 62, 1968, 147-150.

Die protokollierten Lehrerreaktionen bestanden aus einer oder einer Sequenz solcher Äußerungen. Außer den Angaben über die Häufigkeit dieser verschiedenen Reaktionen verschaffte sich Zahorik auch noch Lehrer-Einschätzungen (ratings) über die Richtigkeit der Schülerantworten und Schüler-Einschätzungen (ratings) über den Wert einer Auswahl von Lehrerreaktionen. Seine Ergebnisse wiesen darauf hin, daß von den 175 verschiedenen Arten von Lehrerreaktionen nur sechzehn häufiger wiederkehrten. Am häufigsten kamen folgende vor:

- (1) Positive Antwort mit anschließender Überleitung zu neuem Thema (8,5 Prozent)
- (2) Bitte um Erweiterung der Antwort, ohne irgendwelche Hilfen zu geben (8,3 Prozent)
- (3) Einfache lobende Bestätigung mit anschließender Überleitung zu neuem Thema (7,8 Prozent)
- (4) Einfache lobende Bestätigung, dann positive Antwort (Wiederholung), dann Überleitung zu neuem Thema (5,8 Prozent)
- (5) Keine Antwort, Überleitung zu neuem Thema (5,1 Prozent)

Die Häufigkeit einer Reaktionsart variierte mit der Schulklasse, mit der

Phase des Unterrichts und mit den Urteilen der Lehrer über die Richtigkeit der Antwort. Durch diese Untersuchung konnte deutlich gemacht werden, daß die Rückmeldung durch den Lehrer keine Verstärkung (reinforcement) in dem Sinne ist, in dem Psychologen wie Skinner diesen Begriff verwenden. Auch betrachten die Lehrer Schülerantworten nicht als ein Verhalten, das Verstärkung braucht. Diese Untersuchung hat ein überraschendes Ergebnis: »Sie zeigt, daß es dem Lehrer nicht in erster Linie auf das sofortige Lernen ankommt und daß es ihm nicht das Wichtigste ist, was das Kind während der Interaktion sagt und tut« (Zahorik 1968).

Diese Ergebnisse sollten keinen Lehrer verwundern, es sei denn, er hätte sich von den Schriften einiger behavioristischer Theoretiker davon überzeugen lassen, daß Schüler nur das lernen, was sie tun, und daß ihr Tun aktiv verstärkt werden muß. Zur Verstärkung des Unterrichtsverhaltens sind viele vage Generalisierungen vorgeschlagen worden, besonders in Hinsicht auf das verbale Verhalten von Schülern. Zahoriks Untersuchung sollte dazu beitragen, entstandene Mißverständnisse zu beseitigen.

Damit soll freilich nicht gesagt werden, die Reaktionen des Lehrers auf die Antworten der Schüler seien unwichtig. Eine Untersuchung von Emmer (1968) weist darauf hin, daß einige Kategorien der Lehrerreaktionen (z. B. Flanders' Kategorie 3: »akzeptiert oder verwendet die Ideen des Schülers«) damit zusammenhängen, wie oft die Schüler durch Fragen die Initiative ergreifen. Emmers Ergebnisse, die aus der Untersuchung von sechzehn Lehrern des zweiten Schuljahres stammen, sind nicht unwidersprochen geblieben; sie machen jedoch deutlich, daß die Schüler um so freier mitmachen, je mehr der Lehrer ihre Ideen akzeptiert und darauf aufbaut. Wenn man davon ausgeht, daß es eine Beziehung zwischen der Beteiligung der Schüler an der Unterrichtsdiskussion und ihrer Leistung gibt – und einiges weist darauf hin (z. B. Gallagher 1966) –, dann sollte man auch nach der Beziehung zwischen der Art der verbalen Lehrerreaktionen und der Schülerleistung fragen. Das bedeutet jedoch nicht einfach, daß Schüler sich besser daran erinnern, was sie im Unterricht getan oder gesagt haben.

Unterschiede im Lehrerverhalten während des ersten Berufsjahrs

Medley berichtete über Beobachtungen an 70 Lehrern der Sekundarschule während des ersten halben Jahres ihrer Berufspraxis. Das verwendete Beobachtungssystem war das System »Observation Schedule and Record« (OSCAR 4V), das Medley und seine Mitarbeiter im Laufe mehrerer Jahre entwickelt hatten. Als signifikante Veränderung im Verhalten dieser Lehrer ergab sich: (1) sie neigten dazu, weniger ergänzende Fragen (d. h., anders formulierte oder zusätzliche Fragen) zu stellen, (2) weniger Reaktionen auf

Schülerantworten zu geben, und (3) mehr Antworten als falsch abzulehnen, gleichzeitig jedoch die Schüler zum Antworten aufzumuntern. Diese Ergebnisse geben vielleicht die zunehmende Geschicklichkeit der Lehrer bei der Formulierung und bei der Auswahl des rechten Zeitpunkts ihrer Fragen wieder. Sie sehen immer seltener einen Anlaß, sich selbst zu verbessern, weil sie immer eindeutiger und angemessenere Fragen stellen.

Die Beziehung zwischen Lehrerverhalten und Schülerleistung

Affektive Dimensionen der Unterrichtsinteraktion

Das Unbehagen nach früheren Mißerfolgen und der Glaube daran, daß alle wichtigen Variablen im Unterricht noch entdeckt werden müssen, haben in neuerer Zeit Forscher dazu geführt, an das Problem der Ursachen von Schülerleistung nur mit beträchtlicher Vorsicht heranzugehen. Ohne Widerspruch zu fürchten, behauptete Bloom in einer Rede über den Stand der pädagogischen Forschung (1966), die Erforschung von Lehrmethoden habe gezeigt, daß die meisten etwa gleich wirkungsvoll sind. Dabei waren einige bedeutende Unterschiede bereits damals bekannt. Flanders (1965) hatte schon die Ergebnisse seiner umfassenden Untersuchung berichtet, in der er eine eindeutige Beziehung zwischen seinem Maß »Indirektheit« in der Unterrichtsinteraktion und der Schülerleistung in der Sozialkunde und in der Mathematik hatte zeigen können. Seinen Untersuchungen zufolge schien es so, als-ob die »indirekteren« Lehrer ein Unterrichtsklima schufen, in dem die Schüler Sympathien für ihre Lehrer empfanden und in standardisierten Leistungstests besser abschnitten.

In einem kleinen Kontrollversuch bestätigte La Shier (1967) diese Ergebnisse. La Shier gebrauchte das Flanderssche Verfahren zur Bewertung der Leistung von zehn Schulpraktikanten (student teachers), die Schüler der achten Klasse in einer sechswöchigen Arbeitseinheit über »tierisches Verhalten« aus dem Biological Science Curriculum Study (BSCS) unterrichteten. Den Schülern wurde in den entsprechenden Inhalten der Biologie ein Vor- und ein Nachtest gegeben, und sie wurden gebeten, einen Fragebogen über ihre Einstellungen zu dem Unterricht auszufüllen. Als Kriteriumswert diente die durchschnittliche Verbesserung der mittleren Leistung jeder Klasse im Vergleich zur Ausgangsleistung (California Test of Mental Maturity). Die Ergebnisse wiesen darauf hin, daß die »Indirektheit« (I/D Verhältnis) der Praktikanten in signifikanter Beziehung zum Leistungsgewinn und den positiven Einstellungen bei den Schülern stand.

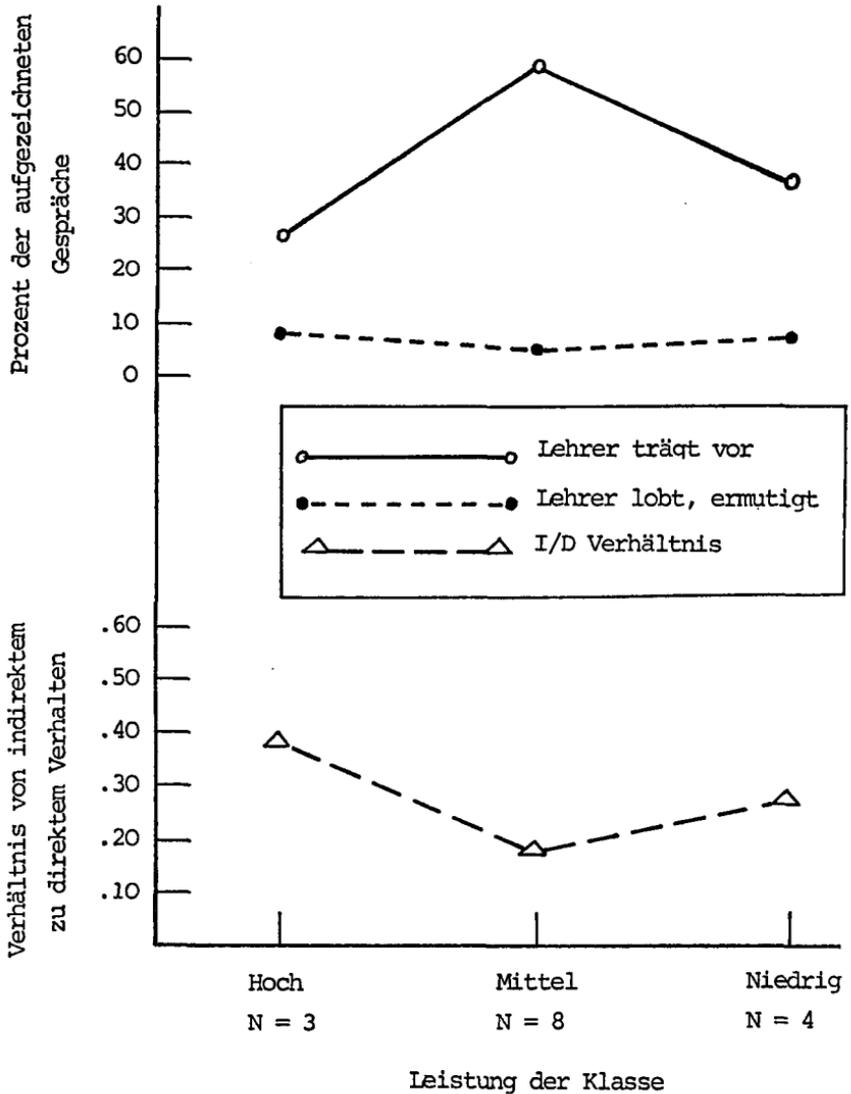
Das Verhältnis zwischen der Einstellung der Schüler und ihrem Leistungszuwachs war ebenfalls signifikant.

In einer noch umfassenderen Untersuchung benutzte Furst (1967) Daten von Bellack und anderen (Bellack u. a. 1966) in dem Bemühen, die Schülerleistung mit Maßen des Flandersschen Verfahrens der Interaktionsanalyse und mit Maßen des Bellackschen Analyse-Systems in Beziehung zu bringen. Bellacks Daten bestanden aus schriftlichen Protokollen von je vier Unterrichtsstunden, die fünfzehn Lehrer in ihren New Yorker Sekundarschulen gehalten hatten. Der Unterrichtsstoff war ein standardisierter vierstündiger Kurs in Wirtschaftswissenschaften. Auf Grund der Ergebnisse der Leistungstests (korrigiert nach Schülerintelligenz und Klassengröße) wurden die Lehrer aufgeteilt. Drei Lehrer wurden den Klassen mit den besten Leistungen, vier denen mit den schwächsten Leistungen und acht denen mit den mittleren Leistungen zugeordnet. Diese drei Lehrergruppen wurden nach drei Maßen des Flandersschen Verfahrens und nach drei des Bellackschen Analyse-Systems miteinander verglichen. Die Ergebnisse zeigten, daß die drei Klassen mit der höchsten Leistung sich von den anderen in folgendem unterschieden: mehr ausführliche, indirekte Gesprächsbeiträge der Lehrer, mehr positive als negative unmittelbare Reaktionen auf Schülerantworten, intensivere Unterrichtseteiligung seitens der Schüler. Bei näherer Betrachtung zeigen Fursts Ergebnisse freilich keine einfache Beziehung zwischen diesen Maßen und der Schülerleistung. Eine Auswahl der Ergebnisse wird in Abbildung 1 dargestellt. Hieraus ist zu ersehen, daß eine Kurvilinearität zwischen der Leistung und dem Prozentsatz der verschiedenen Arten von protokolliertem Lehrer- und Schülerverhalten besteht. Ob diese deutliche Kurvilinearität in Fursts Daten allerdings einmalig ist oder ob dieser Sachverhalt von anderen Forschern nur übersehen worden ist, ist nicht bekannt. Aber das weist darauf hin, daß bei einer anderen Lehrer-Stichprobe möglicherweise ganz andere Ergebnisse gefunden worden wären.

Noch einige Ergebnisse von Furst müssen erwähnt werden. Sie legten die Hypothese nahe, Lehrer mit hohen Leistungsergebnissen gäben in ihren Unterrichtsdiskussionen eine mittlere Anzahl pädagogischer Impulse zur verbalen Strukturierung des Unterrichts (verbal structuring moves), zeigten ein mittleres Verhältnis im Frage-Antwort-Austausch und einen hohen Grad an Verschiedenheit in den logischen Prozessen. Ein zusammengesetzter Meßwert, der aus den Einzelwerten dieser »kognitiven« Variablen bestand, bestätigte diese Hypothese. Ohne weitere Details über die Art dieser Meßwerte und ihre Beziehung zur Leistung ist es jedoch unmöglich, die Bedeutung dieser Ergebnisse richtig einzuschätzen.

Powell (1968) versuchte zu zeigen, daß man auf Grund der Beobachtun-

Abbildung 1. Meßwerte der Klassen mit hohen, mittleren und niedrigen Leistungen, gewonnen mit dem Flandersschen Verfahren der Interaktionsanalyse (aus: Furst, 1967).



gen von Lehrern in ihrer normalen schulischen Umgebung auf signifikante Beziehungen zwischen »Indirektheit« und besseren Schülerleistungen bei standardisierten Leistungsprüfungen schließen kann. Er verwendete in verschiedenen Klassen des dritten Schuljahres das Flanderssche Verfahren und wiederholte das Ganze im folgenden Jahr in denselben Klassen (viertes Schuljahr). Er wählte die Schüler aus, die während ihrer ersten drei Schuljahre dieselben »indirekten« bzw. »direkten« Lehrer hatten, und im vierten Jahr dann umgekehrt bei »direkten« bzw. »indirekten« Lehrern waren. Die Ergebnisse zeigen, daß die Leistung im Rechnen, nicht aber im Lesen, mit der »Indirektheit« des Lehrers während der ersten drei Schuljahre eng zusammenhängt. Und im vierten Schuljahr hatte es offenbar überhaupt keine Wirkung, ob die Schüler bei einem »direkten« oder »indirekten« Lehrer waren. Der Autor schloß daraus:

Von den Ergebnissen dieser Untersuchung her scheint es eindeutig, daß der indirekte Unterricht keinen deutlichen Gesamtvorteil bringt (Powell 1968, 4).

Das Problem bei solchen Untersuchungen liegt in der Neigung des Forschers, eine Eigenschaft wie »Indirektheit« als feststehendes Attribut zu betrachten. Wenn der eine Forscher die beobachteten Lehrer in die »indirektesten« und die »direktesten« einteilt, kann seine Trennungslinie an einer anderen Stelle liegen als die eines anderen Forschers, der eine andere Lehrerstichprobe untersucht. Die Lehrer, die in der einen Untersuchung als »indirekt« bezeichnet werden, können in einer anderen Untersuchung als »direkt« bezeichnet werden. Dies macht vielleicht nicht sehr viel aus, sofern das Verhältnis zwischen den Faktoren des Unterrichtsklimas und der Leistung einfach linear ist. Aber wenn dieses Verhältnis kurvilinear ist, wie die Daten in Fursts Untersuchung andeuten, dann wird eine sorgfältigere Analyse der Daten erforderlich.

Der vielleicht sorgfältigste und umfangreichste Versuch, Aspekte des Unterrichtsklimas und des Lehrereinflusses mit der Schülerleistung in Beziehung zu bringen, wurde von Soar (1967; 1968) berichtet. Soar begann seine Untersuchung mit der Sammlung umfangreicher Daten über die Schüler in 55 Klassen (drittes bis sechstes Schuljahr) in vier verschiedenen Elementarschulen. Er ergänzte die standardisierten Leistungstests für Wortschatz, Lesen und Rechnen mit Kreativitätstests (Minnesota), Angsttests (Children's Manifest Anxiety Scale), Tests über Neigung zur Abhängigkeit (Flanders 1960) und Fragebogen, mit deren Hilfe die Einstellungen der Schüler gegenüber Lehrern und Mitschülern abgeschätzt werden sollten. Die gleichen Tests wurden nach einem Jahr und am Ende des zweiten Jahres wiederholt. Die Werte des Lernzuwachses der Schüler über den Zeitraum von einem bzw. zwei Jahren wurden als Kriteriumswerte verwendet.

In jeder der 55 Klassen wurden Beobachtungen angestellt, und zwar mit dem Flandersschen Verfahren der Interaktionsanalyse und mit einem anderen Verfahren, das Teile von Medley und Mitzels OSCAR und Fowlers Ablehnungs-Zustimmungs-Skala (Hostility Affection Schedule) einschloß. Alle Meßwerte aus diesen Beobachtungen wurden faktorenanalysiert; es wurde eine Neun-Faktoren-Lösung erarbeitet, die »das Bild vom Unterricht am deutlichsten wiedergab«.

Für jede Klasse wurde ein Faktorenwert für jeden dieser neun Faktoren errechnet. Diese Faktorenwerte wurden dann mit dem durchschnittlichen Lernzuwachs korreliert, den die Schüler in den Tests erreicht hatten. Die Ergebnisse können so zusammengefaßt werden:

(1) Es gab keinen signifikanten Zusammenhang zwischen den beiden Faktoren, die die Ursache für die höchsten Prozentwerte der Varianz waren, und einem der Maße des Leistungszuwachses. Diese Faktoren schienen vielmehr Maße für die Gesamtdauer der sprachlichen Äußerungen der Lehrer und für die Möglichkeit der Schüler zu sein, eine Diskussion anzufangen.

(2) Der Faktor mit der engsten Beziehung zur Leistungsverbesserung der Schüler wurde als »ausgedehnter Lehrervortrag im Gegensatz zu schneller Lehrer-Schüler-Interaktion« bezeichnet (extended discourse vs. rapid teacher-pupil interchange). Soar bemerkt dazu:

»Hier spiegelt sich offenbar ein Zyklus wider, in dem zuerst der Lehrer 15 bis 20 Sekunden lang redet, eine Frage stellt oder eine Anweisung gibt, und dann die Schüler eine Weile lang reden, . . . kein längerer Vortrag im üblichen Sinn also, sondern eine Reihe von Lehreräußerungen, die der Problem-Strukturierung dienen können, oder die Darbietung einer begrenzten Zahl von Informationen, die die Schüler weiterentwickeln und verwenden sollen.« (Soar 1967, 7).

Die Parallele zwischen dieser Beschreibung und Fursts »mittlerem Grad der Steuerung der Lernaktivitäten« ist evident.

(3) Die Häufigkeit verbaler Schärfe oder Kritik durch den Lehrer stand in negativer Beziehung zum Lernzuwachs der Schüler im Rechnen und in einiger Beziehung zu ihrer Angst und ihrer Neigung zur Abhängigkeit.

(4) Ein Faktor zeigte viel Ähnlichkeit mit Flanders Maß der »Indirektheit«. Er stand jedoch weniger zur »Direktheit« des Lehrers in negativer Beziehung als vielmehr zum Ausmaß an »Ruhe und Unruhe« im Unterricht. Als Faktor korrelierte er zwar nicht mit der Verbesserung der Schülerleistung, stand aber in positiver Beziehung zur Steigerung der Kreativitätswerte und der Interessen der Schüler am Unterricht. Soar fiel es schwer zu erklären, warum der Faktor, der am engsten mit der »Indirektheit« des Lehrers zusammenhing, keine Korrelation mit der Leistung zeigte. Er

korrelierte jedoch mit dem jeweiligen Schuljahr, wobei Soar davon ausging, daß bei einigen früheren Untersuchungen man vielleicht das Ausmaß der Leistung mit der Höhe des Schuljahrs vermischt habe.

Im zweiten Teil seiner Untersuchung erforschte Soar den Einfluß des Lehrerverhaltens auf die Entwicklung der Schüler im Laufe eines zweiten Jahres. Er interessierte sich für das Ausmaß, in dem sich die Wirkungen des Lehrerverhaltens eines bestimmten Jahres vielleicht auf nachfolgende Jahre übertragen. Die Ergebnisse zeigten, daß die Kritik der Lehrer nur in geringem Maße von einem Jahr auf das nächste übertragen wird:

Es scheint, als ob Ausdrücke negativer Affekte im Unterricht ihren stärksten Einfluß in der ersten Zeit danach hätten. Diesjährige Kritik ist von Bedeutung, die letztjährige nicht mehr.

Die Ergebnisse für die anderen Faktoren, die die Schülerleistung beeinflussen, zeigen kontinuierliche Auswirkungen und eine gewisse Interaktion zwischen den beiden Jahren. Der wohl auffallendste Befund im Hinblick auf die oben berichteten Daten von Furst besteht darin, daß ein mittleres Ausmaß von Lehrerkontrolle und »Indirektheit« die größten Vorteile zu bringen schien:

... in diesen Klassen erzeugte ein mittleres Maß an Kontrolle, entweder in Form von Kritik oder indirektem Unterricht, mehr wünschenswerte Veränderungen im Schülerverhalten als der übertriebene Mangel an Lehrerkontrolle. Vielleicht läßt sich daraus die Notwendigkeit ersehen, daß der Lehrer ein Minimum an Strukturierung schaffen muß, innerhalb derer die Schülerleistung maximiert werden kann (Soar 1967, 10).

Zweifellos wird es weitere Untersuchungen geben, die den Versuch machen werden, Schülerleistung und Wert für das affektive Klima im Unterricht miteinander in Beziehung zu bringen. Aber schon jetzt scheint genügend Beweismaterial vorzuliegen, um einige allgemeine Schlüsse ziehen zu können.

Erstens: Beobachtungsverfahren mit nur einem einzigen Kriterium wie dem Ausmaß der »Indirektheit« des Lehrers sind wahrscheinlich in ihrem Ansatz zu ungenau, um jemals klare Beziehungen mit Kriterien der Schülerleistung aufzeigen zu können. In einer Reihe von Beobachtungen werden sie wahrscheinlich zufällig mit mehreren anderen Aspekten des Unterrichtsverhaltens in Verbindung gesetzt.

Zweitens: Selbst wenn diese Beobachtungsverfahren eine gültige Dimension des Unterrichtsverhaltens erfassen, ist es unwahrscheinlich, daß diese Dimension eine einfache lineare Beziehung zur Schülerleistung aufweist. Wahrscheinlich ist es naiv zu erwarten, daß eine größere Freiheit der Schüler und ein freundlicheres Verhalten der Lehrer direkte und kontinuierliche Fortschritte in der intellektuellen Entwicklung des Schülers bringen.

Wie aus Soars Untersuchung deutlich hervorgeht, ist Kritik und verbale Schärfe der Lehrer gegenüber den Schülern ein wichtiger Faktor, aber das Ausbleiben dieser Kritik führt nicht automatisch zur Verbesserung des Lernens. Untersuchungen über das Unterrichtsklima sind schließlich Untersuchungen der Bedingungen, *unter denen Lernen stattfindet*. Daß es eine beträchtliche Vielfalt an Bedingungen gibt, die das Lernen nicht verhindern, sollte dabei nicht überraschen.

Die Suche nach signifikanten kognitiven Variablen

In einer neueren Abhandlung über die kognitiven Aspekte im Unterricht, bezweifelte Gage (1966 b) die Nützlichkeit, solche deskriptiven Verfahren wie die Systeme von Bellack und Smith zu entwickeln, bevor etwas über die mögliche Gültigkeit der kognitiven Aspekte des Unterrichts bekannt ist, die analysiert werden sollen. Er zitierte eine Untersuchung, in der offensichtlich signifikante Unterschiede zwischen den Fähigkeiten der Lehrer, den Schülern ein Prinzip zu erklären, auftreten. Gage schlug einen systematischen Vergleich des Verhaltens solcher Lehrer vor, die bekannterweise gut bzw. schlecht erklären können; dieser Vergleich sollte wichtige Hinweise auf die wesentlichen Elemente solchen Unterrichts liefern. Im Anschluß daran berichteten Fortune, Gage und Shutes (1966) über die Ergebnisse eines Vergleichs der »Erklärungsfähigkeit« von 40 Praktikanten. Diese Praktikanten mußten in viertelstündigen Lektionen kleinen Gruppen von Schülern der Sekundarschule eine kontrollierte Themenreihe unterrichten. Als Kriteriumswert für jedes Thema wurde ein Test mit zehn Auswahl-Antwort-Aufgaben benutzt (multiple-choice item test), von denen die Praktikanten die Hälfte schon gesehen hatten, die andere Hälfte ihnen aber unbekannt war. Die Untersuchung war so aufgebaut, daß die Fähigkeit der Praktikanten zu erklären bzw. die Themen zu unterrichten über die verschiedenen Themen und die verschiedenen Schülergruppen hinweg verglichen werden konnte. Die Ergebnisse wiesen darauf hin, daß sich das Erklärungsvermögen eines Lehrers wahrscheinlich mit verschiedenen Themen verändert, bei verschiedenen Schülergruppen jedoch relativ konstant bleibt.

Vor einiger Zeit machten Gage und seine Mitarbeiter Fernsehaufzeichnungen von 43 Lehrern der Sozialkunde, die Schülern der zwölften Klasse Unterricht in zwei standardisierten Themen gaben. Die Themen bezogen sich auf aktuelle ökonomische, politische und soziale Entwicklungen in Jugoslawien und Thailand. Die Lehrer unterrichteten beide Themen in Lektionen von 15 Minuten, während ein drittes Thema (über Israel) allen Klassen in Form einer standardisierten Tonbandlektion vorgeführt wurde. Die Fähigkeit der Schüler, den Inhalt der standardisierten Tonbandlektion zu

lernen, diente dazu, die Kriteriumstestwerte an die anderen zwei Themen anzupassen. Die Schüler in jeder Klasse wurden außerdem darum gebeten, anhand standardisierter Fragebogen jede Lektion zu bewerten und den Grad ihrer Aufmerksamkeit während jeder Lektion anzugeben.

Ist die Erklärungsfähigkeit der Lehrer für verschiedene Themen relativ konstant? Podlogar, Rosenshine und Gage entdeckten (1968) in ihrer Datenanalyse, daß die Korrelation zwischen den durchschnittlichen Testwerten der Schüler für beide Themen zwischen .41 und .47 lag. Diese Korrelation ist signifikant und weist darauf hin, daß zwischen 16 und 20 Prozent der Varianz in der Klassenleistung auf einen Faktor der Fähigkeit des Lehrers zurückzuführen sein könnte. Aus den Bewertungen der Schüler wurde außerdem klar, daß sie die Erklärungsfähigkeit eines Lehrers ziemlich genau einschätzen konnten und daß die Einstufung ihrer eigenen Aufmerksamkeit signifikant mit dem Ausmaß ihres Lernens korrelierte.

Zwei etwas verschiedene Versuche, die wichtigen Komponenten der »Erklärungsfähigkeit« zu isolieren, sind mit diesen Daten gemacht worden. Hiller, Fisher und Kaess (1968) sowie Dell und Hiller (1968) berichteten von einem Versuch, mit Hilfe eines Computers die kritischen Elemente im verbalen Verhalten von Lehrern zu isolieren. Rosenshine (1968) berichtete die Ergebnisse eines Vergleichs innerhalb einer Teilstichprobe, die aus den erfolgreichsten und erfolglosesten Lehrern der Gesamtstichprobe gebildet worden war; dabei wurde sowohl das Verbal- als auch das Handlungsverhalten der Lehrer untersucht. Die Computer-Analyse erbrachte zwei Dimensionen des verbalen Lehrerverhaltens, die verläßlich mit dem Lehrererfolg bei den beiden Lektionen zusammenzuhängen schienen. Die erste dieser Dimensionen wurde als »verbale Flüssigkeit« (verbal fluency) bezeichnet. Sie bestand in einem zusammengesetzten Maß, das auf der Durchschnittslänge des gesprochenen Satzes und auf anderen Kennzeichen der Sprachflüssigkeit, wie etwa dem Anteil der »ähs« in der Rede des Lehrers basiert. Der zweite Faktor wurde »Unbestimmtheit« (vagueness) genannt: gleichfalls ein zusammengesetztes Maß, das sich aus der relativen Häufigkeit solcher Worte wie »fast, im allgemeinen, vielleicht usw.« errechnet. Die Forscher folgerten daraus nicht, daß diese verbalen Faktoren in direktem Zusammenhang mit dem Lernen des Schülers standen:

Wir sind der Meinung, daß das Verhältnis zwischen unserem Maß und den Werten des Kriteriumstests vor allem die Korrelation und nicht die Ursache widerspiegelt. . . . unser Maß der Unbestimmtheit dient wohl eher als Hinweis auf andere Verhaltensweisen, die ihrerseits das Verstehen und Behalten der Lektion kausal beeinflussen (Hiller/Fisher/Kaess 1968, 7).

In Rosenshines Analyse wurden viele Variablen untersucht; auf einige

ist bereits in früheren Untersuchungen hingewiesen worden, andere, z. B. die linguistischen Indikatoren, wurden erstmals hier untersucht. Die folgenden drei Variablen zeigten eine signifikante Beziehung zum Lehrererfolg:

(1) *Gesten und Bewegung*. Die fähigeren Lehrer zeigten eine größere Tendenz, sich frei in der Klasse zu bewegen und dabei Gesten zu machen.

(2) *Regel und Beispiel*. Die fähigeren Lehrer neigten dazu, Regeln vor und nach der Diskussion der Beispiele zu erklären, während die weniger fähigen Lehrer dazu neigten, die Regel nur einmal, entweder vor oder nach den Beispielen zu erklären.

(3) *Erklärende Bindeglieder* (explaining links). Die fähigeren Lehrer benutzten in der Regel häufiger Bindewörter wie »weil, weshalb, um . . . zu, folglich, mittels, usw.«.

Rosenshine beendete seinen Bericht mit der Behauptung, die wichtigsten Variablen seien offensichtlich jene, die mit der Struktur der Kommunikation des Lehrers zu tun hätten. Die sequenzielle Anordnung der Gedanken und ihrer tragenden Elemente bedarf weiterer Erforschung. Man sollte jedoch nicht zu viel in die Ergebnisse dieser von Gage an der Stanford Universität angefangenen Untersuchungsreihe hineinlesen. Es darf nicht vergessen werden, daß, obwohl die Daten auf der Untersuchung einer größeren Zahl von Lehrern beruhen, nur zwei fünfzehnminütige Stichproben von dem Unterricht eines jeden Lehrers gemacht wurden⁴. Die Ergebnisse dieser Analyse lassen sich jedoch ohne weiteres mit jenen anderer Forscher wie beispielsweise Furst (1967) und Soar (1966; 1967) vereinbaren.

Ein Versuch mit Variationen in der sequenziellen Anordnung von Unterrichtsinhalten

Eine Untersuchung, die aus der Arbeit von Smith und Meux (1967) entstand, wirft etwas Licht auf die Bedeutung der Gedankensequenz des Lehrers. In dieser Untersuchung (Nuthall 1968) wurden vier alternative Strategien zum Unterrichten von Begriffen (concept teaching strategies) bzw. Sequenzen von »Begriffs-Impulsen« (sequence of conceptual moves) miteinander verglichen (vgl. die Beschreibung von Smiths Analyse von Begriffs-Impulsen). Es waren vier Unterrichtsstrategien, die Smith und seine Mitarbeiter in ihren Tonbandaufzeichnungen von Unterrichtsstunden in Sekundarschulen identifiziert hatten. Der verbale Inhalt dieser Strategien wurde in programmierter Textform einer Stichprobe von 412 Sekundarschülern dargeboten. Eine Varianzanalyse der Ergebnisse nachträglicher Kriteriumstests (delayed criterion tests) deutete an, daß die alterna-

tiven Strategien signifikant andere Wirkungen hatten. Die weitere Analyse ließ vermuten, daß die Wirksamkeit einer Unterrichtsstrategie davon abhängig ist, in welchem Maße sie auf dem schon vorhandenen Wissen der Schüler aufbaut und wie sehr sich Unterrichtsstrategie und vorhandenes Wissen gegenseitig beeinflussen. Bei den beiden Begriffen, die während dieses Versuchs unterrichtet wurden, war der Gebrauch von Beispielen und von Vergleichen mit anderen Begriffen je nach dem Vorwissen der Schüler verschieden.

Einige Schlußfolgerungen und Hinweise für künftige Forschung

Wie zu Beginn dieses Beitrags erwähnt wurde, haben viele Untersuchungen der letzten Jahre Versuchscharakter gehabt. Man sollte nicht versuchen, endgültige Schlußfolgerungen zu ziehen, auf denen Lehrer ihre Praxis aufbauen oder woraus Forscher experimentelle Hypothesen ableiten könnten. Einige Ergebnisse scheinen jedoch in den Schlußfolgerungen vieler Forscher wiederzukehren. Alle beziehen sich auf verbales Lehrerverhalten als einer signifikanten Ursache für Schülerleistung. Solange der Lehrer den Unterricht von Feindseligkeit und übermäßiger Kritik freihalten kann, wird wahrscheinlich die Wirksamkeit seines Einflusses auf die Schülerleistung von solchen Dingen abhängig sein wie:

- (1) seiner Fähigkeit, den verbalen Kontext, in dem die Interaktion zwischen Lehrer und Schülern stattfindet, vorzubereiten bzw. zu »strukturieren«;
- (2) seiner Fähigkeit, bei der Darbietung der Unterrichtsinhalte Gedanken mit einem Maximum an logischem Zusammenhang und einem Minimum an Unbestimmtheit und Ziellosigkeit zu organisieren;
- (3) seiner Fähigkeit, die Schüler zur Teilnahme an Diskussionen anzuregen und sie für die Entwicklung und Erweiterung von Ideen zu interessieren.

Diese Themen sollten als Anregungen zu weiterer Forschung betrachtet werden, insbesondere für die Entwicklung experimentell brauchbarer Theorien des Unterrichtsverhaltens.

Dem kritischen Leser muß es aufgefallen sein, daß vielen der erwähnten Untersuchungen eine feste Richtung und eine kontrollierte Anordnung fehlen; dieser Mangel kann nur durch eine angemessene Theorie beseitigt

werden. Es drängt sich folgende Frage auf: Wenn diese Untersuchungen das Lernen der Schüler im Unterricht erfassen sollen, warum werden dann so wenige von ihnen durch die bekannten Lerntheorien beeinflusst? Wann wird eine Verbindung zwischen den etablierten psychologischen Theorien und dieser neuen Richtung der Erforschung des Unterrichtsverhaltens hergestellt?

Nach Meinung dieses Autors gibt es nur eine Antwort auf diese Fragen, daß nämlich die traditionelle psychologische Theorie keinen großen Wert haben kann, bevor die Unterrichtsforscher ihrerseits nicht signifikante theoretische Erklärungen des Unterrichtsgeschehens gefunden haben. Man braucht keine weitere Anpassung und Ausweitung bekannter Theorien, sondern die Schaffung einer neuen Theorie, die direkt auf das wirkliche Verhalten, das sie erklären soll, bezogen ist. Die in diesem Beitrag behandelten Untersuchungen weisen darauf hin, daß wir jetzt schon genug über einige wahrscheinlich wichtige Variablen wissen, um ihre Funktion wenigstens teilweise zu erklären.

Im Rahmen einer Unterrichtstheorie verdienen folgende Elemente des Unterrichtsverhaltens besondere Aufmerksamkeit:

(1) Nicht alle Schüler beteiligen sich immer am Unterricht; dennoch wird erwartet, daß alle lernen, und im allgemeinen lernen alle etwas. Eine Theorie des stellvertretenden Lernens (*vicarious learning*) beziehungsweise des Lernens bei nachlassender Teilnahme ist daher erforderlich.

(2) Die Bedeutung der Schülerbeteiligung sollte unabhängig von Begriffen, Reaktion (*response*) und Verstärkung (*reinforcement*) eingeschätzt werden. Viele Autoren (z. B. Smith 1961) haben angedeutet, daß Lehrer die Beteiligung ihrer Schüler an Diskussionen als Informationsquelle über den Wissensstand und die intellektuellen Prozesse der Schüler benutzen. Das heißt, sie diagnostizieren die Schülerreaktionen und treffen spontane Entscheidungen aufgrund dieser Diagnosen. Jacksons neuere Veröffentlichung (1968), die auf Interviews mit erfolgreichen Lehrern beruht, weist darauf hin, daß Lehrer ihre sichersten Erfolgszeichen der subtilen »Rückmeldung« entnehmen, die sie von den Schülern ihrer Klasse erhalten. Als Praktiker sind sie sehr mißtrauisch und voller Zweifel gegenüber den Ergebnissen standardisierter Tests. Soweit diesem Autor bekannt ist, hat bisher noch keiner die Vorstellungen vom Lernen zu erklären versucht, aufgrund derer ein fähiger Lehrer die Diskussion beeinflusst. Wie interpretiert er die Reaktionen seiner Schüler? Welche Zeichen benutzt er, um den Verlauf seiner Handlungen zu bestimmen? In den Antworten auf diese Fragen muß die Erklärung für Ursache und Wirkung der Diskussion im Unterricht liegen.

(3) Die Bedeutung des Schülers als einer unabhängigen Größe für Verän-

derungen im Unterricht muß berücksichtigt werden. In einem neueren Artikel hat Turner (1968) Material aufgearbeitet, welches die Hypothese unterstützt, daß Schüler zumindest einige Aspekte des Lehrerverhaltens selbst bestimmen. Schüler können ihrerseits ziemlich genau den Erfolg ihrer Lehrer bewerten (vgl. Podlogar/Rosenshine/Gage 1968); Beobachter des Unterrichtsverhaltens werden bald darauf aufmerksam, daß Schüler die subtilen Hinweise ihrer Lehrer begreifen können und es auch tatsächlich tun. Jede Theorie, die darauf zielt, das Lernen im Unterricht zu erklären, sollte die Schüler auch als aktiv auswählende Teilnehmer berücksichtigen.

(4) Die Doppelrolle des Lehrers, die darin besteht, z. T. intellektuell und psychisch auf die Schüler Einfluß zu nehmen (Gage, 1966a) und z. T. ihre Motivation zu erhöhen, bedarf weiterer Untersuchungen. Die aktive Teilnahme des Schülers zu erreichen ist eine oft genau so kunstvolle und schwierige Aufgabe wie die Führung der Gedankengänge von Schülern. Das Problem der Motivation im Unterricht muß bei jeder Untersuchung der gedanklichen Interaktionen mitberücksichtigt werden. Hier ist die Erzeugung von verdeckten geistigen Reaktionen (covert mental responses) ähnlich schwierig wie die Stimulierung offener verbaler Reaktionen.

Schließlich ist kein kritischer Bericht über die Unterrichtsforschung ohne die Warnung vollständig, die auch verschiedene andere Autoren geäußert haben: Die Frage nach den besten Unterrichtsmethoden und den wirksamsten Mitteln, das Lernen zu fördern, darf nicht mit den Fragen darüber gekoppelt werden, wie sich Unterrichten und Lernen im Unterricht in Wirklichkeit abspielen. Erst wenn wir über die Beziehungen zwischen Unterrichtsverhalten, Lernen und Leistung der Schüler genauere Kenntnisse haben, wird es vielleicht möglich sein, wohlbegründete Unterrichtsprinzipien aus diesem Verständnis abzuleiten. Dieser kritische Bericht sollte deutlich machen, daß ein solches Verständnis noch nicht erreicht worden ist. Wahrscheinlich werden die Bemühungen um ein ausreichendes Verständnis des Unterrichts zusätzlich erschwert, wenn dieser Wunsch nach einem besseren Verständnis mit dem nach einem verbesserten Unterricht verwechselt wird.