

3. Kapitel

Methodische Ansätze der empirischen Erziehungs- und Sozialisationsforschung

Wolfgang Schneider

1 Einführung

Wenn hier der Versuch gemacht wird, einen Überblick über wichtige methodische Ansätze der empirischen Erziehungs- und Sozialisationsforschung zu geben, so steht von vornherein fest, daß der Anspruch auf Vollständigkeit allein schon aus Raumgründen nicht eingelöst werden kann. Die Vielzahl methodischer Ansätze in der Psychologie läßt sich schon längst nicht mehr in einen einzigen Band integrieren (vgl. etwa Bredenkamp & Feger, 1983; Feger & Bredenkamp, 1983 a, 1983 b).

Macht es angesichts dieser Ausgangssituation überhaupt noch Sinn, einen eigenständigen Beitrag zum Methodenspektrum innerhalb der Erziehungs- und Sozialisationsforschung zu verfassen? Wäre es stattdessen nicht ratsamer, auf das inzwischen gut sortierte Angebot in herkömmlichen Statistikbüchern zu verweisen? Es gibt m. E. einen guten Grund dafür, methodischen Ansätzen in der Erziehungs- und Sozialisationsforschung ein separates Kapitel zu widmen. Im Unterschied zu anderen Teilbereichen der Psychologie dominieren in der empirischen Erziehungs- und Sozialisationsforschung komplexe Fragestellungen, die sich über klassische experimentelle Designs (vgl. Campbell & Stanley, 1966) nicht angemessen beantworten lassen. Die grundsätzliche Komplexität des Forschungsansatzes läßt sich anhand des Definitionsversuchs von Hurrelmann (1986, S.14) verdeutlichen, wonach Sozialisation den Prozeß bezeichnet, „in dessen Verlauf sich der mit einer biologischen Ausstattung versehene menschliche Organismus zu einer sozial handlungsfähigen Persönlichkeit bildet, die sich über den Lebenslauf hinweg in Auseinandersetzung mit den Lebensbedingungen weiterentwickelt“. Eine zentrale Position in diesem Ansatz nimmt der Begriff der Persönlichkeit ein, die nach Hurrelmann als

spezifisch organisiertes Gefüge von Merkmalen, Eigenschaften, Einstellungen und Handlungskompetenzen bezeichnet werden kann. Die Persönlichkeitsentwicklung läßt sich als überdauernde und langfristige Veränderung wesentlicher Elemente dieses Gefüges auffassen.

Angesichts dieses umfassenden theoretischen Modells verwundert es nicht, daß viele empirische Ansätze innerhalb der Erziehungs- und Sozialisationsforschung komplex ausgelegt sind. Die etwa bei Cook und Campbell (1979) näher beschriebenen quasi-experimentellen Designs der sozialwissenschaftlichen Feldforschung, die eine Ausweitung des strikt laborexperimentellen Ansatzes implizieren und für den Fall nicht-äquivalenter Experimental- und Kontrollgruppen erstellt wurden, sind für die Ziele der Erziehungs- und Sozialisationsforschung meist zu restriktiv, also lediglich für eine nichtrepräsentative Teilmenge von Studien anwendbar. Da die Mehrzahl der Arbeiten in diesem Bereich also weder experimentell noch feldexperimentell aufgebaut ist, vielmehr dem Typus der Feldstudie zuzuordnen ist, eröffnen sich hier methodische Probleme und Möglichkeiten, die in den traditionellen Lehrbüchern der psychologischen Methodik nur am Rande behandelt werden.

Es ist ein Ziel der vorliegenden Abhandlung, die Sensitivität des Lesers für die Problematik zu erhöhen, daß komplexe sozialisationstheoretische Modellvorstellungen in angemessene statistische Analysemodelle umgesetzt werden müssen. Ein weiteres, vergleichsweise gewichtigeres besteht darin, in die wesentlichen methodischen Ansätze zur Datengewinnung und -auswertung einzuführen. In einem ersten Schritt wird die typische Ausgangssituation für Forschungsarbeiten im Bereich der Erziehungs- und Sozialisationsforschung charakterisiert. Im Mittelpunkt stehen hier Methoden der Datenerfassung, die natürlich Auswirkungen auf die Qualität der Daten und damit auch auf die Wahl der Auswertungsmethoden haben.

Der Schwerpunkt der Abhandlung liegt auf der Diskussion von Auswertungsmethoden. Üblicherweise werden für die Auswertung von im Feld erhobenen Daten multivariate Verfahren herangezogen. Bei einigen Studien dominiert dabei der Aspekt der Datenreduktion, also etwa die Frage, wie sich aus der Vielzahl von Einzeldaten grundlegende Dimensionen (z. B. der Persönlichkeit, des Erziehungsstils) gewinnen lassen. Je nach Art und Qualität der Daten sind hier unterschiedliche faktoren- und typenanalytische Verfahren relevant, auf die im zweiten Kapitel näher eingegangen wird.

Vergleichsweise häufiger werden in neueren Studien Verfahren zur Modellierung von komplexen Relationen eingesetzt. Es handelt sich hier im wesentlichen um regressionsstatistische Prozeduren unterschiedlicher Komplexität. Die Palette der verfügbaren Verfahren reicht dabei von einfachen Regressions-

modellen bis hin zu Strukturgleichungsmodellen mit latenten Variablen. Die m. E. wichtigsten Ansätze werden im dritten Abschnitt beschrieben.

Während die meisten empirischen Arbeiten auf Querschnittsanalysen aufbauen, also nur die Daten eines einzigen Meßzeitpunkts berücksichtigen, werden in neueren Arbeiten zunehmend Prozeßverläufe über größere Zeiträume erfaßt. Methodische Möglichkeiten und Probleme von Längsschnittstudien sind in der Regel noch wenig bekannt. Auf sie wird im letzten Abschnitt dieses Beitrags näher eingegangen.

Die Darstellung der einzelnen methodischen Ansätze ist im wesentlichen nicht-technisch gehalten, verzichtet also weitmöglichst auf die Formelsprache. Das Ziel der Ausführungen liegt darin, empirisch arbeitende pädagogische Psychologen und Sozialisationsforscher in das Rationale wesentlicher methodischer Ansätze ihres Arbeitsbereiches einzuführen. Da aus Platzgründen nur knappe Erörterungen der einzelnen Verfahren möglich sind, empfiehlt sich für ein besseres Verständnis die genaue Lektüre der angeführten vertiefenden Literatur.

2 Methoden zur Gewinnung von Daten im Bereich der Sozialisations- und Erziehungspsychologie

Die typische empirische Untersuchung im hier relevanten Inhaltsbereich ist als multivariate Querschnittstudie konzipiert. Für eine meist recht große Anzahl von Probanden werden demnach zu einem einzigen Meßzeitpunkt eine Vielzahl theoretisch interessanter Merkmale erfaßt, deren Zusammenhangsmuster in der Folge analysiert werden soll. In der einschlägigen Literatur wird ein solcher Ansatz häufig als Korrelationsstudie, nicht-experimentelle Studie oder als „passive-observational study“ charakterisiert (Cook & Campbell, 1979). Meist bleibt die Analyse jedoch nicht bei der Erörterung von Korrelationsmustern stehen. Aus theoretischen Erwägungen heraus werden einige Merkmale als unabhängige Variablen oder Prädiktormasse, andere als abhängige Variablen oder Kriteriumsmaße definiert. Da in korrelativen Querschnittsstudien keine Effektmanipulationen wie etwa im Labor- oder Feldexperiment vorgenommen werden, kann in der Regel nicht auf Kausalzusammenhänge geschlossen werden (vgl. etwa Wottawa, 1988). Nur mit manchen sehr elaborierten Auswertungstechniken, die als Spezialfälle des mathematischen Modellbaus gelten können, und nur unter Rekurs auf ein probabilistisches Kausalitätskonzept sind solche Rückschlüsse möglich. Problematisch bleibt in jedem Fall, daß der Typus der Feldstudie im Vergleich mit dem Labor- und Feldexperiment in Anbetracht der fehlenden Kontrollmöglichkeiten für Stör-

faktoren und damit für Artefakte am anfälligsten ist (vgl. Gniech, 1982; Wotawa, 1988). Für die Datengewinnung sind in der einschlägigen Feldforschung vor allem die folgenden Techniken herangezogen worden.

2.1 Beobachtungsmethoden

Diesen Verfahren ist gemeinsam, daß Daten aus interpersonell wahrnehmbaren Zuständen oder Verhaltenssequenzen gewonnen werden. Obwohl Verfahren der Selbstbeobachtung durchaus existieren, dominieren im Bereich der Sozialisations- und Erziehungspsychologie Techniken der Fremdbeobachtung (vgl. etwa Feger, 1983; Ischi, 1982; Schumm, 1990; Trollenier, 1985). Beobachtungen können mehr oder weniger strukturiert bzw. standardisiert sein. Sie können weiterhin offen oder verdeckt erfolgen. Erfolgt die Beobachtung nach klar definierten Vorgaben, ist das Verfahren objektiv und überprüfbar, wobei man allerdings in Kauf nimmt, daß der Handlungsspielraum des Beobachters eingeschränkt wird. Unsystematische Beobachtungen sind schwerer replizierbar und lassen sich wohl nur dann rechtfertigen, wenn der Hypothesenbereich näher abgeklärt werden soll (vgl. etwa Petermann, 1987). Die Beobachtungskategorien sind umso zuverlässiger und valider, je enger sie am tatsächlichen Verhalten orientiert sind und je weniger Interpretationen vorgenommen werden (vgl. für einen anderen Standpunkt Köckeis-Stangl, 1980, die hochinferente Beobachtungsdaten für die Sozialisationsforschung höher gewichtet).

Bei der offenen Beobachtung gibt sich der Forscher in seiner Beobachtungsfunktion zu erkennen, während dies bei der verdeckten Beobachtung gerade vermieden werden soll. Obwohl für bestimmte Fragestellungen die offene Beobachtung gerade in Form teilnehmender Beobachtung nützlich sein kann, ist generell dann die verdeckte Beobachtung vorzuziehen, wenn möglichst „natürliche“ Daten gewonnen werden sollen. Es steht dabei jedoch fest, daß diese Form der Beobachtung im Familienkontext oder der Klassenzimmer-Situation äußerst schwer durchführbar ist. Ischi (1982) führt neben der direkten (offenen) und indirekten (verdeckten) Verhaltensbeobachtung noch die apparativ-vermittelte Beobachtung an, bei der im Feld lediglich eine Ton- und/oder Bildspeicherung erfolgt und der Kodierungs- und Registriervorgang im Nachhinein geschieht. Je nach Kategorienschema lassen sich Daten unterschiedlicher Differenziertheit und Qualität generieren, deren Skalenniveau von dichotom oder grobkategorial (vorhanden/nicht vorhanden) über Rangskalen (mehr oder weniger vorhanden) bis hin zum Intervallskalen-Niveau reichen kann, bei dem feine Abstufungen zwischen Verhaltensvarianten möglich werden. Die Liste der verwendeten Instrumente umfaßt Checklisten, bei denen

das Auftreten vordefinierter Verhaltensweisen abgehakt wird, Rating-Skalen, bei denen die Prävalenz bestimmter Verhaltensweisen über festgelegte Zeiteinheiten festgehalten wird, und interaktive Kodiersysteme, die beispielsweise Lehrer-Schüler-Interaktionen innerhalb vorgegebener Zeitgrenzen erfassen (vgl. Stallings & Mohlmann, 1988). Die Mehrzahl der Beobachtungssysteme erzeugt demnach qualitative Daten, die für jeden Probanden die Häufigkeit bestimmter Verhaltenskategorien angeben (z. B. Helmke & Renkl, 1992).

Die Qualität der unterschiedlichen Beobachtungssysteme läßt sich nicht zuletzt daran ablesen, in welchem Ausmaß die klassischen Gütekriterien erfüllt werden. Die Objektivität der Kodierungen läßt sich über die Beobachterübereinstimmung erfassen, die in der Regel hohe Kennwerte (.8 bis .9) erreichen sollte. Zur Kontrolle der Reliabilität werden mehrere Messungen vorgenommen und miteinander verglichen. Der Reliabilitätskoeffizient läßt sich dann nach einem der aus der klassischen Testtheorie bekannten Verfahren (Split-half, Retest- und Paralleltest-Methode) ermitteln. Eine Erweiterung der klassischen Testtheorie, die sogenannte „Generalisierungstheorie“ ermöglicht zusätzlich die Ableitung eines „Generalisierungskoeffizienten“, der simultan mehrere Fehlerquellen berücksichtigt (vgl. Renkl & Helmke, 1993). Vergleichsweise wenig wurde bisher die Frage der Validität der im Feld erhobenen Daten behandelt (vgl. Helmke & Renkl, 1992; Ischi, 1982). Man begnügt sich hier häufig mit der Augenscheinvalidität der Daten oder Vergleichen mit Experten-Einschätzungen. Einschätzungen der Inhaltsvalidität, also der Frage, ob ein repräsentatives Verhaltens-Sample erfaßt wurde, findet man ebenso wenig wie Überprüfungen der Kriteriumsvalidität, also etwa dem Vergleich mit anderen Maßen zum Themenbereich, deren Validität schon bekannt ist (vgl. Schumm, 1990).

2.2 Befragungsmethoden

Die Befragung läßt sich allgemein als theoriegeleitetes, zielgerichtetes Vorgehen der Datengewinnung definieren, wobei die Befragten entweder innerhalb eines systematisch angelegten Gesprächs oder durch einen schriftlich vorgelegten Fragenkatalog zu Informationsverarbeitungsprozessen veranlaßt werden (vgl. Schwarzer, 1983). Als weitere Abgrenzungskriterien werden häufig die Art der Strukturierung und Standardisierung genannt. Die Strukturierung betrifft dabei den Aufbau der Befragung, insbesondere die Anordnung und den Grad der Geschlossenheit der Fragen. Demgegenüber bezieht sich der Aspekt der Standardisierung auf die inhaltliche Vergleichbarkeit der Befragung, kommt also darin zum Ausdruck, daß die Reihenfolge und Formulierung von

Fragen für die Mehrzahl der Befragten gleich ist. Im folgenden werden die üblichsten Verfahren der Befragung kurz charakterisiert.

(a) *Schriftliche Befragung.* In der Erziehungs- und Sozialisationsforschung werden Fragebögen und Rating-Skalen vergleichsweise am häufigsten eingesetzt. Für beide Verfahren gilt in der Regel, daß sie sowohl hoch strukturiert als auch hoch standardisiert sind. Für Fragebögen gilt letzteres allerdings nur, wenn sie ausschließlich Fragen des geschlossenen Typs beinhalten, also lediglich eine Auswahl zwischen vorgegebenen Antwortalternativen (multiple choice) erfordern. Häufig zitiert werden Fragebogen-Konstruktionen aus dem Bereich der Persönlichkeitsforschung (vgl. Mummendey, 1979; Köckeis-Stangl, 1980). Üblicherweise werden Subskalen für unterschiedliche Persönlichkeitsbereiche formuliert, häufig auch „Lügen-Skalen“ beigefügt, um das Ausmaß von Reaktionen im Sinne der sozialen Erwünschtheit zu erfassen. Problematisch können sich auch sogenannte „response sets“, also systematische Verzerrungen der Reaktionen aufgrund bestimmter Frageformulierungen auswirken.

Je nach Meßniveau der Fragen sind unterschiedliche Auswertungsmethoden möglich. Holm (1975) unterscheidet zwischen Nominal-, Ordinal- und quantitativen Fragen. Bei Nominalfragen werden die Antworten unterschiedlichen Klassen zugeordnet (Beispiel: Einordnung von Befragten nach Studienfächern). Bei Ordinalfragen werden den Antworten Rangplätze zugewiesen (Beispiel: die Einordnung von Schulabschlüssen nach qualitativen Gesichtspunkten). Quantitative Fragen messen in der Regel auf Intervallskalen-Niveau, wobei unterstellt wird, daß zwischen den Antwortalternativen in etwa gleiche numerische Abstände anzusetzen sind. Da bei quantitativen Fragen Auswertungstechniken anwendbar sind, die ein Maximum an Information erbringen, sollte man die Strategie verfolgen, möglichst Fragen dieses Typs zu konstruieren (Holm, 1975). Standardisierte Fragebögen enthalten üblicherweise Angaben zu den klassischen Gütekriterien. Sie sollten nur bei befriedigenden bis guten Reliabilitäts- und Validitätskennwerten herangezogen werden.

Zur Gruppe der schriftlichen Befragungsprozeduren können im Grunde auch standardisierte Testverfahren gezählt werden. Im Bereich der Erziehungs- und Sozialisationsforschung sind darunter insbesondere kognitive Fähigkeitstests und schulische Leistungstests zu subsumieren. Im Vergleich zu den anderen erwähnten schriftlichen Befragungsformen sind hier am wenigsten Probleme mit dem Meßniveau der Daten und ihren Gütekennwerten zu registrieren. Es kann in der Regel von intervallskalierten Daten und guten Reliabilitäts- und Validitätskoeffizienten ausgegangen werden.

Rating-Verfahren erfordern vom Beurteiler Ordinalurteile, die entweder auf numerischen oder graphischen Schätzskalen erfolgen (vgl. Mummendey,

1979). Es werden dabei häufig fünfstufige Urteilsskalen des Likert-Typs verwendet, die bipolar angelegt sind und eine eindeutig neutrale Antwort zulassen. Mummendey weist auf das Problem hin, daß viele Rating-Skalen subjektiv konstruiert sind, also oftmals in der Auswahl der relevanten Eigenschaften recht willkürlich scheinen. Die Reliabilität von Ratings ist schwierig zu schätzen (am ehesten wohl über Intraclass-Koeffizienten). In vielen Feldern der Erziehungs- und Sozialisationsforschung stellen Reliabilitätskontrollen von Ratings noch immer eher die Ausnahme dar.

(b) *Mündliche Befragung.* Das Interview als die mündliche Form der Befragung zählt zu den verbreitetsten Erfassungsmethoden der Sozialforschung (vgl. Hron, 1982). Es ist in der Regel niedrig standardisiert, unterscheidet sich jedoch im Hinblick auf den Strukturierungsgrad. Im sogenannten strukturierten Interview ist die Frageanordnung festgelegt, um soziale Einflußgrößen und andere Aspekte der Datenerhebungssituation möglichst gut kontrollieren zu können. Demgegenüber sind offene Interviews wenig strukturiert. Der Interviewer agiert hier als echter Gesprächspartner, der zuhört, im Bedarfsfall um ergänzende Informationen bittet und auch sonst spontane Bemerkungen macht. Dem offenen Interview wird vor allem explorierende und hypothesengenerierende Funktion zuerkannt.

Herkömmliche Methodenlehren empfehlen den Einsatz strukturierter Interviews, die als Mischung offener und geschlossener Fragen konzipiert sein sollten, um Ermüdungseffekte und Konzentrationsschwächen zu vermeiden. Die grundlegenden methodischen Probleme dieses Verfahrens ähneln denen der standardisierten schriftlichen Befragungen. Standardisierungsprobleme bestehen nicht nur im Hinblick auf die Interaktion zwischen Interviewer und Befragtem, sondern auch im Hinblick auf die Strukturierung von Fragen und Frageabläufen (vgl. Hron, 1982): Variationen des Wortlauts von Fragen wie auch ihrer Abfolge können zu unterschiedlichen Ergebnissen führen. Weitere öfters diskutierte Probleme betreffen Response-Sets, Interviewer-Einflüsse sowie die Kontrolle der Untersuchungssituation. Im Vergleich mit den methodischen Schwierigkeiten des strukturierten Interviews werden die des offenen Interviews als weitaus gravierender angesehen. Nach Hron (1982) besteht ein grundlegendes Problem darin, daß der theoretische Status der über qualitative Erhebungsverfahren erlangten Erkenntnisse noch wenig geklärt ist. Es fehlen insbesondere konsensfähige Verfahren zur Interpretation von Interviewtranskripten. Fragen nach der Objektivität, Wiederholbarkeit und Gültigkeit solcher Erhebungsformen bleiben im wesentlichen unbeantwortet.

Der kurzgehaltene Überblick über die unterschiedlichen Methoden der Datengewinnung im Bereich der Erziehungs- und Sozialisationsforschung sollte verdeutlicht haben, daß in den hier üblicherweise vorfindbaren Feldstudien

eine breite Palette von Erhebungsmethoden verwendet wird, die in unterschiedlichem Ausmaß mit methodischen Problemen belastet sind. Über Beobachtungsmethoden lassen sich zumeist qualitative Daten gewinnen, während Befragungsmethoden mehrheitlich Rangdaten produzieren. Nur in wenigen Fällen (etwa bei Testdaten) ist mit einiger Sicherheit davon auszugehen, daß Intervallskalen-Niveau vorliegt.

Obwohl von Methodikern immer wieder die enge Beziehung zwischen dem Niveau der Datenerhebung und den Methoden der Datenauswertung betont wird, wird dieser Gesichtspunkt bei der Datenanalyse viel zu wenig berücksichtigt (vgl. Mummendey, 1979). Die Darstellung der Erhebungsverfahren hat weiterhin auch deutlich gemacht, daß Arbeiten zu Fragen der Erziehungs- und Sozialisationsforschung in der Regel große Datenmengen produzieren, die oftmals vor der weiteren Verarbeitung erst sinnvoll reduziert werden müssen. Im folgenden Abschnitt werden deshalb zunächst die wichtigsten Verfahren zur Datenreduktion vorgestellt, bevor dann auf die Möglichkeiten eingegangen wird, komplexe Zusammenhangsmodelle zu formulieren und zu testen.

3 Methoden zur Datenreduktion und -gruppierung

Im folgenden werden verbreitete multivariate Verfahren in ihrem Ansatz beschrieben, mit denen im wesentlichen eines von zwei Zielen erreicht werden soll. Das erste Ziel besteht darin, die Anzahl der erfaßten Merkmale auf wenige Grundmerkmale (Faktoren) zu reduzieren und damit zu gruppieren. Das zweite Ziel ist darin zu sehen, daß Personen aufgrund mehrerer Merkmale in homogene Gruppen (Typen) klassifiziert werden. Die Darstellung beschränkt sich im wesentlichen auf Verfahren der Faktoren-, Cluster- und Diskriminanzanalysen, die schon in gängigen Lehrbüchern der Statistik (z. B. Bortz, 1989) ausführlich beschrieben und gleichzeitig meist auch in den weit verbreiteten Computer-Statistikpaketen BMDP, SAS und SPSS enthalten sind. Zusätzlich wird kurz auf die Konfigurationsfrequenzanalyse (KFA) eingegangen, die in der gängigen Computer-Software zwar noch nicht ausreichend repräsentiert ist, dies m. E. aber durchaus verdient hätte (vgl. Krauth & Lienert, 1973; Bortz, Lienert & Boenke, 1990; von Eye, 1992).

3.1 Faktorenanalyse

Der Begriff „Faktorenanalyse“ bezieht sich auf eine Gruppe von multivariaten statistischen Techniken, deren Ziel generell darin besteht, einen Variablensatz über eine kleine Zahl hypothetischer Variablen abzubilden. Alle diese Verfah-

ren gehen von der Annahme aus, daß eine kleinere Anzahl von den Daten zugrundeliegenden Faktoren die Kovariation (Korrelation) zwischen den beobachteten Variablen erklären kann. Über Faktorenanalysen läßt sich demzufolge etwa prüfen, welche Anzahl von latenten Persönlichkeitsdimensionen über einen bestimmten Persönlichkeitsfragebogen abgebildet wird. Die Ausgangsbedingungen mögen dabei unterschiedlich sein. Vielfach wird der Forscher nicht von vornherein wissen, wieviele latente Dimensionen in seinen Daten enthalten sind. In diesem Fall wird das Ziel der Analyse darin bestehen, die Minimalzahl von Faktoren herauszufinden, die für die beobachteten Kovariationen zwischen den Variablen verantwortlich sind. Man spricht dann von exploratorischen Faktorenanalysen, die in der gegenwärtigen Forschungspraxis sicherlich die Mehrzahl der Anwendungsfälle ausmachen. Verfügt der Forscher über ein reichhaltigeres Ausgangswissen, kann die Methode aber auch im Sinne einer konfirmatorischen Faktorenanalyse zum Testen und zum Bestätigen von Hypothesen eingesetzt werden.

(a) *Exploratorische Faktorenanalyse.* Sicherlich am häufigsten wird in der psychologischen Forschung die sogenannte Hauptkomponenten-Analyse eingesetzt. Die Ausstattung einer Person mit einer Komponente (Faktor) wird als Faktorwert, die Bedeutung eines Faktors für ein gemessenes Merkmal als Ladung des Merkmals auf dem entsprechenden Faktor bezeichnet. Die Summe der quadrierten Ladungen einer Variablen wird „Kommunalität“ genannt. Sie gibt an, in welchem Ausmaß eine Variable durch die Faktoren aufgeklärt wird. Die Faktoren in der Hauptkomponentenanalyse weisen die Eigenschaft auf, daß sie (a) wechselseitig voneinander unabhängig sind und (b) sukzessiv maximale Varianz in den Merkmalen aufklären. Das Verfahren arbeitet demnach so, daß durch den ersten Faktor ein maximaler Varianzanteil aufklärt. Von der verbleibenden Varianz klärt der zweite Faktor wieder ein Maximum auf, etc.

Die Ermittlung der Faktoren in der Hauptkomponentenanalyse gewährleistet nicht immer, daß die extrahierten Faktoren sinnvoll interpretiert werden können (dies ist immer dann ein Problem, wenn viele Variablen auf dem ersten Faktor laden). Deshalb wird häufig im Anschluß an die Hauptkomponenten-Analyse eine Rotationstechnik eingesetzt (vgl. Bortz, 1989, S.665 für eine Auflistung möglicher Rotationstechniken). Grundsätzlich werden schiefwinklige (oblique) Rotations-Techniken, bei denen die resultierenden Faktoren interkorreliert sind, von orthogonalen Techniken unterschieden, bei denen die Unabhängigkeit der Faktoren gewahrt bleibt. Rotationsmethoden folgen dem Prinzip der Einfachstruktur, was meint, daß auf jedem Faktor einige Variablen möglichst hoch und andere möglichst niedrig laden sollen. Dies führt allgemein zu einer besseren Interpretierbarkeit der Faktoren.

Ein grundsätzliches Problem der exploratorischen Faktorenanalysen (weitere Techniken sind etwa bei Bortz, 1989 oder Kim & Mueller, 1978, beschrieben) besteht darin, daß sie zu interpretativ mehrdeutigen Ergebnissen führen, die demnach keine Überprüfung spezifischer inhaltlicher Hypothesen zulassen. Letzteres ist nur über konfirmatorische Ansätze möglich.

(b) *Konfirmatorische Faktorenanalyse*. Die Limitationen des exploratorischen Ansatzes sind vor allem darin zu sehen, daß alle Faktoren (je nach Rotations-Technik) als korreliert oder aber unkorreliert angenommen werden und weiter unterstellt wird, daß alle Variablen in unterschiedlichem Ausmaß direkt von allen Faktoren affiziert sind. Im konfirmatorischen Modell ist es dagegen möglich anzunehmen, daß nur bestimmte Faktoren interkorreliert sind, andere dagegen nicht, und daß bestimmte Variablen nur auf bestimmten Faktoren laden, auf anderen dagegen nicht. Die theoretischen Modellannahmen lassen sich über einen sogenannten „goodness-of-fit“-Anpassungstest prüfen, der die Kompatibilität der hypothetisch angenommenen Ladungsmatrix mit der empirisch ermittelten Ladungsmatrix vergleicht. Ist die Datenanpassung schlecht, läßt sich das Verfahren durchaus im nächsten Schritt exploratorisch einsetzen, indem Alternativmodelle miteinander verglichen werden (vgl. für weiterführende Informationen Jöreskog & Sörbom, 1979; Long, 1983).

Exploratorischen und konfirmatorischen Faktorenanalysen ist gemeinsam, daß sie das Kriterium der multivariaten Normalverteilung voraussetzen und davon ausgehen, daß die Daten Intervallskalen-Niveau aufweisen. Obwohl die Verfahren relativ robust gegenüber leichteren Verletzungen dieser Voraussetzungen zu sein scheinen, ist noch wenig darüber bekannt, zu welchen systematischen Ergebnisverzerrungen diese Abweichungen führen.

3.2 Clusteranalyse

Unter dem Begriff „Clusteranalyse“ läßt sich eine Vielzahl von Verfahren einordnen, bei denen der Klassifikationsgesichtspunkt im Mittelpunkt steht. Bei diesem ebenfalls heuristischen Verfahren werden beispielsweise Personen, für die ein bestimmter Variablensatz (etwa Persönlichkeitstests) erhoben wurde, nach Maßgabe ihrer Ähnlichkeit in Gruppen oder „Cluster“ eingeteilt (Bortz, 1989).

Obwohl die Technik der Clusteranalyse seit ihren Anfängen als „Faktorenanalyse des kleinen Mannes“ belächelt wurde, erfreut sie sich gerade in letzter Zeit immer größerer Beliebtheit. In Literaturüberblicken (vgl. Aldenderfer & Blashfield, 1984; Eckes & Roßbach, 1980; Milligan, 1981; Schneider &

Scheibler, 1983) wurde immer wieder auf die ständig und rapide steigende Zahl von einschlägigen Veröffentlichungen und neuen Algorithmen verwiesen. Selbst für Kenner der Materie ist es deshalb nicht leicht, sich in dem dabei produzierten Terminologie-Wirrwarr zurechtzufinden.

Eine Grobeinteilung der verfügbaren Verfahren ergibt zwei Hauptgruppen, hierarchische Klassifikationsverfahren und partitionierende Techniken. Hierarchische Klassifikationsverfahren sind in der „agglomerativen“ Variante dadurch charakterisiert, daß nach einer Serie sukzessiver Fusionen von N Elementen (Personen) zu einer Endstufe gelangt wird, in der die Objekte einer einzigen Klasse angehören. Bei der „divisiven“ Form wird genau umgekehrt verfahren, d. h. die Gesamtheit der Elemente wird schrittweise in immer kleinere Klassen ähnlicher Objekte zerlegt. Demgegenüber sind die mit partitionierenden Techniken gebildeten Cluster nicht hierarchisch, sondern wechselseitig exklusiv organisiert: nach einer Anfangsaufteilung (Partitionierung) des Datensatzes in eine vorgegebene Anzahl von Gruppen wird die Homogenität der Cluster iterativ (durch Austausch von Einzelelementen) so lange zu verbessern versucht, bis ein näher zu definierendes lokales Optimum erreicht ist.

Nach einer längeren Durststrecke, während der nur das von Wishart (1978) entwickelte Programmsystem CLUSTAN mit vielen clusteranalytischen Varianten in den meisten Universitäts-Rechenzentren verfügbar war, sind die gängigsten hierarchisch-agglomerativen und iterativ-partitionierenden Techniken auch in BMDP und SAS implementiert (bei SPSS finden sich lediglich die hierarchisch-agglomerativen Verfahren). Die Auswahl der „richtigen“ Clusteranalyse fällt dem Anwender insofern schwer, als nur wenige Informationen über die Bewährtheit von Verfahren vorliegen. Probleme sind insbesondere damit verbunden, daß verschiedene Algorithmen unterschiedliche Ergebnisse für den gleichen Datensatz liefern. Um die Güte der Lösungen unterschiedlicher Algorithmen besser beurteilen zu können, sind in jüngerer Vergangenheit sogenannte „Monte-Carlo-Studien“ durchgeführt worden (vgl. Milligan & Cooper, 1987; Scheibler & Schneider, 1985). Die meisten vorliegenden Evaluationsversuche deuten darauf hin, daß die Kombination einer hierarchisch-agglomerativen Prozedur (Ward-Methode) und einen iterativ-partitionierenden Verfahrens (k-means) besonders robuste Ergebnisse liefert.

Abschließend sei bemerkt, daß für die Clusteranalyse ähnliche Voraussetzungen wie für die Faktorenanalyse gelten. Die Merkmale sollten in der Regel multivariat normalverteilt sein und hohes Skalenniveau besitzen, unabhängig davon, ob als Ähnlichkeitsmaße Korrelationskoeffizienten oder Euklidische Distanzen verwendet werden.

3.3 Diskriminanzanalyse

Während über die unterschiedlichen Clusteranalyse-Algorithmen Strukturen bzw. Gruppierungen in den Daten gesucht werden, die nicht ohne weiteres durch visuelle Inspektion erkennbar sind, verfolgt die Diskriminanzanalyse eine Identifikationsstrategie: Über diese Methode werden keine neuen Gruppen geschaffen, sondern Objekte (Personen) bereits bestehenden Gruppen zugeordnet. Wie schon für die Verfahren der Faktoren- und Clusteranalyse gilt auch hier, daß der Sammelbegriff mehrere statistische Varianten bezeichnet. Nach Klecka (1981) lassen sich dabei grob zwei Hauptrichtungen unterscheiden. Eine Variante fokussiert auf der Interpretation von Gruppenunterschieden, die andere auf der Klassifikation von Fällen in Gruppen (für eine differenziertere Betrachtung vgl. Bortz, 1989; Lachenbruch, 1975).

(a) *Interpretation von Gruppenunterschieden.* Um die Frage nach der Natur von Gruppenunterschieden zu klären, werden in einem ersten Schritt „diskriminierende Variablen“ erfaßt, von denen angenommen wird, daß sie prinzipiell gut zwischen den vorgegebenen Gruppen (z. B. Schüler mit hoher versus niedriger Leistungsängstlichkeit) differenzieren können. Im nächsten Schritt werden „kanonische Diskriminanzfunktionen“ berechnet, die Linearkombinationen der diskriminierenden Variablen darstellen und bestimmte Bedingungen erfüllen. Die Koeffizienten für die erste Diskriminanzfunktion werden etwa so bestimmt, daß die Gruppenmittelwerte auf dieser Funktion so weit wie möglich differieren. Die Koeffizienten für die zweite Funktion werden ebenfalls so abgeleitet, daß maximale Gruppendifferenzen resultieren, allerdings unter der Zusatzbedingung, daß ihre Werte mit denen der ersten Funktion nicht korrelieren. Die maximale Anzahl von Funktionen, die sich auf diese Weise bestimmen läßt, entspricht der um Eins reduzierten Anzahl der Gruppen oder Variablen (abhängig davon, ob weniger Gruppen oder Variablen existieren). Für den Fall von drei Gruppen und sechs Variablen lassen sich maximal zwei Diskriminanzfunktionen berechnen.

Für die Interpretation der Diskriminanzfunktionen können unterschiedliche Maße herangezogen werden. Standardisierte Koeffizienten geben die relative Bedeutung einer diskriminierenden Variablen für den resultierenden Diskriminanzwert der Funktion an. Über sogenannte „Strukturkoeffizienten“ läßt sich weiterhin die Ähnlichkeit zwischen einer diskriminierenden Variablen und der Diskriminanzfunktion bestimmen, die als Produkt-Moment-Korrelation zwischen beiden Merkmalen definiert ist. Für den Fall, daß die Korrelation sehr hoch ist, beinhalten Variable und Funktion annähernd die gleiche Information, so daß die Funktion nach der Variablen benannt werden kann. Um die Frage beantworten zu können, wieviele Diskriminanzfunktionen bedeutsam

sind, empfiehlt es sich, die Eigenwerte (Lambdas) genauer zu analysieren. Je höher der Wert für Lambda ausfällt, desto besser lassen sich die Gruppen über die betreffende Diskriminanzfunktion trennen. Überprüfungen der statistischen Signifikanz von Diskriminanzfunktionen werden häufig anhand von Wilks' Lambda, einem multivariaten Maß zur Bestimmung von Gruppenunterschieden, vorgenommen.

(b) *Klassifikationsprozeduren.* Im Unterschied zur Interpretation der Gruppenunterschiede anhand von Diskriminanzfunktionen soll durch Klassifikationsprozeduren geklärt werden, zu welcher vorgegebenen Gruppe ein bestimmter Fall (eine bestimmte Person) mit größter Wahrscheinlichkeit gehört. Bei diesem Vorgang werden entweder die diskriminativen Variablen oder die Diskriminanzfunktionen verwendet, um die Gruppenzugehörigkeit vorherzusagen. Es läßt sich etwa auf der Basis der Diskriminanzfunktionen für jede Person einer Stichprobe ihre Distanz zu den Zentroiden der bestehenden Gruppen berechnen. Die Person würde dann in die Gruppe eingeordnet, zu deren Zentroid insgesamt die geringste Distanz besteht.

Eine andere Variante der Klassifikation besteht darin, die Genauigkeit der Zuordnungsprozedur zu testen. In diesem Fall wird geprüft, welcher Prozentsatz der Gruppenmitglieder aufgrund der Diskriminanzfunktion wieder der gleichen Gruppe zugeordnet würde. Die Klassifikationsmethode stellt damit einen anderen Zugang zur Erfassung der Güte von Trennfunktionen dar. Ist die verfügbare Stichprobe genügend groß, läßt sich eine „split-sample“-Validierungstechnik einsetzen, indem die Stichprobe nach Zufall in zwei Gruppen aufgeteilt wird. Nachdem die Diskriminanzfunktionen für die erste Stichprobe abgeleitet worden sind, lassen sie sich dann in der zweiten Stichprobe zur Prädiktion von Klassifikationen verwenden.

3.4 Konfigurationsfrequenzanalyse (KFA)

Bei der Beschreibung der unter 3.1 bis 3.3 aufgeführten Verfahren wurde immer wieder darauf aufmerksam gemacht, daß sie multivariate Normalität der Variablen-Verteilungen unterstellen und davon ausgehen, daß die Variablen auf möglichst hohem Meßniveau (Intervallskalen-Niveau) erfaßt worden sind. Das Verfahren der KFA unterscheidet sich von den oben aufgeführten Klassifikationsverfahren dadurch, daß es als sog. verteilungsfreie Methode keine Verteilungsvoraussetzungen hinsichtlich der verwendeten Meßwerte erfordert.

Die KFA wurde von Lienert (1969) als heuristische Methode konzipiert und später zusammen mit Krauth (Krauth & Lienert, 1973) als inferenzstatistische Methode vorgestellt. Eine ausführliche neuere Diskussion des Ansatzes findet sich bei von Eye (1990 a). Die KFA läßt sich zusammengefaßt als Methode zur posthoc-Klassifikation von Individuen darstellen, wobei als Basis eine Datenmatrix mit t Merkmalen und N Personen fungiert. Die Merkmale sollten binär kodiert sein, also lediglich angeben, ob das Merkmal bei bestimmten Personen vorhanden ist oder nicht. Die von den t binären Werten gebildeten N Muster bzw. Konfigurationen werden im folgenden genauer analysiert. Die beobachteten Frequenzen werden mit ihren Erwartungswertschätzungen unter der Nullhypothese totaler Unabhängigkeit der t Merkmale verglichen. „Signifikant“ überfrequentierte Muster lassen sich als Typen, „signifikant“ unterfrequentierte Konfigurationen als Anti-Typen fassen. Probleme mit der „Signifikanz“ sind darin zu sehen, daß bei der KFA mehrere Prüftests an dem gleichen Datensatz vorgenommen werden. Es sind hier deshalb alpha-Adjustierungen vorzunehmen, bei denen z. B. die vorgesehene Signifikanzschranke durch die Anzahl der insgesamt vorgenommenen Prüftests dividiert wird.

Will man Typen und Antitypen aus Stichproben verschiedenen Umfangs hinsichtlich ihrer Ausgeprägtheit oder Prägnanz vergleichen, so läßt sich ein zwischen 0 und 1 variierender „Prägnanzkoeffizient“ berechnen, der seinen numerischen Werten nach dem parametrischen Determinationskoeffizienten entspricht (vgl. Bortz et al., 1990). Dieser Koeffizient gibt also die praktische Bedeutsamkeit der Lösung an und ist insbesondere bei großen Fallzahlen dem Signifikanzkriterium vorzuziehen.

Im Unterschied zu anderen Klassifikationsverfahren muß bei der KFA die Zahl der Individuen im Verhältnis zur Zahl der Merkmale größer sein, da sonst die Erwartungsschätzungen u. U. zu niedrig ausfallen. Weiterhin sollten die Merkmale in höherem Ausmaß als bei anderen Verfahren als „typologisch relevant“ bekannt sein (vgl. Krauth & Lienert, 1973). Um solchermaßen relevante Merkmale herauszufiltern, kommt die sogenannte hierarchische KFA zur Anwendung, bei der die Zahl der Merkmale jeweils um Eins reduziert wird und mit den verbleibenden $t-1$ Merkmalen einfache KFA durchgeföhrt werden. Man entscheidet sich dann für diejenige Merkmalskombination, die das höchste Gesamt-Chi-Quadrat ergibt.

Sollen in einer KFA Typen mit variierenden Anzahlen von Merkmalen identifiziert werden, können Einzelkonfigurationen in einer agglutinierenden KFA zusammengefaßt werden (vgl. Krauth & Lienert, 1973). Zwei überfrequentierte Vierer-Konfigurationen, die sich nur in einem Merkmal unterscheiden (z. B. $+++-$ und $+-++$) lassen sich in einer Dreierkonfiguration zusammenfassen, bei der das Unterscheidungsmerkmal durch einen Punkt charakterisiert ist

(+ +.-). Diese Maßnahme empfiehlt sich bei nur schwach ausgeprägten Konfigurationstypen und vermeidet zu niedrige Erwartungswertschätzungen für die Beurteilung der Konfigurationsfrequenzen.

Hierarchische und agglutinierende KFA sind beides heuristische Suchverfahren und bedürfen der Kreuzvalidierung. Zu diesem Zweck lassen sich Zwei- oder Mehrstichproben-KFA einsetzen, wenn unterstellt werden kann, daß die verschiedenen Stichproben aus der gleichen Population stammen. Ergeben sich für die unterschiedlichen Teilstichproben ähnliche Konfigurationen von Typen und Antitypen, kann man davon ausgehen, daß es sich um einen robusten Befund handelt.

4 Modellierung komplexer Relationen

Die in Abschnitt 3 beschriebenen multivariaten Verfahren bewähren sich in der Erziehungs- und Sozialisationsforschung immer dann, wenn es darum geht, die Vielzahl von erfaßten Merkmalen auf eine überschaubare Zahl von grundlegenden Dimensionen zu reduzieren. Obwohl diese Verfahren angesichts der häufig vorfindbaren „Datenberge“ zweifellos eine wichtige Funktion besitzen, sind multivariate Prozeduren zur Überprüfung von Zusammenhängen zwischen Variablen und Variablenmustern im Hinblick auf ihre Popularität und Verbreitung innerhalb der Sozialwissenschaften wohl nicht minder relevant. Die hier gemeinten regressionsanalytischen Verfahren bauen allesamt auf der Korrelationsstatistik auf und werden insbesondere zur Merkmalsvorhersage eingesetzt. Im folgenden werden zunächst Varianten der multiplen Regressionsanalyse besprochen, die auf beobachteten Merkmalen basieren und gerade in der schulischen Sozialisations- und Erziehungsforschung sehr verbreitet sind. Im Anschluß daran wird in das Rationale von Strukturgleichungsmodelle mit latenten Variablen eingeführt, die eine Verknüpfung des faktoren- und regressionsanalytischen Ansatzes darstellen.

4.1 Möglichkeiten und Probleme multipler Regressionsanalysen

Der Einsatz von multiplen Regressionsanalysen empfiehlt sich immer dann, wenn individuelle Unterschiede in einem bestimmten Merkmalsbereich (z. B. schulische Leistungen) durch individuelle Unterschiede in einem anderen Bereich (z. B. Intelligenz oder Ängstlichkeit) vorhergesagt oder erklärt werden sollen. Es ist hier wichtig hervorzuheben, daß über diese Technik die beiden grundlegenden Anliegen der sozialwissenschaftlichen Forschung, die Vorher-

sage und die Erklärung von Zusammenhängen gleichermaßen realisiert werden können. Der erstgenannte Aspekt steht dann im Mittelpunkt, wenn etwa aus einer Gruppe potentiell wichtiger Merkmale solche selegiert werden sollen, über die das Kriterium am besten vorhergesagt werden kann. Dies läßt sich über unterschiedliche Routinen der schrittweisen Regression ermöglichen. Der explanatorische Aspekt wird dann bedeutsam, wenn es um den relativen Anteil bestimmter Prädiktoren an der Varianzaufklärung im Kriterium geht. Im klassischen Modell der multiplen linearen Regression wird der Wert einer abhängigen Variablen (dem Kriterium) durch die gewichtete Summe von N unabhängigen Variablen (den Prädiktoren) so zu schätzen versucht, daß die Summe der quadrierten Residuen, also der Differenzen zwischen dem aktuellen Kriteriumswert und seinem Schätzwert, ein Minimum ergibt (vgl. für detaillierte Einführungen Bortz, 1989; Lewis-Beck, 1983). Die standardisierten Regressionskoeffizienten oder beta-Gewichte geben den relativen Vorhersagewert der Prädiktorvariablen für das Kriterium wieder. Über das Quadrat der multiplen Korrelationskoeffizienten, den sogenannten Determinationskoeffizienten (R -Quadrat) läßt sich der durch die Prädiktorvariablen im Kriterium aufgeklärte Varianzanteil bestimmen.

Die Vielzahl der inzwischen verfügbaren regressionsanalytischen Varianten kann in diesem Rahmen nicht angemessen beschrieben werden (vgl. dafür Bortz, 1989; Lewis-Beck, 1983; Pedhazur, 1982). Wenn auch das Standardmodell der linearen multiplen Regression sicherlich am weitesten verbreitet ist, findet man neuerdings auch vermehrt Anwendungen von polynomialen Regressionen, bei denen zusätzlich Prädiktoren höherer Ordnung in die Gleichung aufgenommen werden, oder multiple nichtlineare Regressionsansätze, die etwa neben den einfachen Prädiktoren deren Produkte als separate Vorhersagemerkmale benutzen.

Während vielfach davon ausgegangen wird, daß Prädiktorvariablen in jedem Fall Intervallskalen-Niveau besitzen sollten, gibt es dennoch keine Probleme damit, auch dichotom erfaßte Merkmale (dummy-Variablen) oder differenzierter gestufte kategoriale Merkmale in die Vorhersagegleichung aufzunehmen (vgl. etwa Lewis-Beck, 1983; Pedhazur, 1982). Handelt es sich sowohl bei den Prädiktoren wie bei dem Kriterium um qualitative Daten, ist beispielsweise das Verfahren der Prädiktionsanalyse nach Hildebrand, Laing und Rosenthal (1977) angemessen (vgl. auch Rudinger, Chaselon, Zimmermann & Henning, 1985). Die Palette von Anwendungsmöglichkeiten ist in den gängigen Software-Paketen angemessen repräsentiert. Obwohl die Technik der Regressionsanalyse im pädagogisch-psychologischen Bereich äußerst beliebt ist und flexibel eingesetzt wird, ist sie mit einer Reihe von ernstzunehmenden Problemen behaftet, die kurz diskutiert werden sollen.

Ein erstes Problem ergibt sich daraus, daß in den Modellannahmen lediglich das Kriterium als (meßfehlerbehaftete) Zufallsvariable verankert ist, während für die Prädiktoren fehlerfreie Messungen unterstellt werden (vgl. Draper & Smith, 1981; Maxwell, 1975). Diese Annahme mag für einige Anwendungsbereiche der Regressionsanalyse (etwa die Volkswirtschaft) durchaus angemessen sein, ist für sozialwissenschaftliche Daten in der Regel jedoch nicht erfüllt. Dies führt dann dazu, daß die Varianzschätzungen der Prädiktoren verzerrt ausfallen, was sich in fehlerbehafteten multiplen Korrelationskoeffizienten niederschlägt („upward“- oder „downward“-bias, vgl. Pedhazur, 1982).

Ein zweites Problem multipler Regressionsanalysen besteht darin, daß die spezifische Anteile von Prädiktoren an der Kriteriumsvarianz nur dann exakt bestimmbar sind, wenn die Prädiktoren unkorreliert, also voneinander unabhängig sind. Da auch diese Forderung in den meisten sozialwissenschaftlichen Anwendungen nicht erfüllt ist, ergibt sich ein weiterer Grund für Verzerrungen in den Regressions-Schätzparametern.

Sind die Interkorrelationen zwischen den Prädiktoren insgesamt hoch (Multikollinearität), resultieren daraus ernsthafte Probleme für die Schätzung. Hohe Multikollinearität wirkt sich extrem negativ auf die Schätzung der Standardfehler der Regressionskoeffizienten aus (vgl. Berry & Feldman, 1985; Pedhazur, 1982). Allgemein gilt, daß die Verzerrung in den Schätzwerten der multiplen Regressionsanalyse umso größer ausfällt, je geringer die Reliabilität der Prädiktoren (d. h. je größer ihr Meßfehler), und je höher die Korrelation zwischen den Prädiktoren ausfällt. Beide Probleme tauchen in Untersuchungen zur Erziehungs- und Sozialisationsforschung immer wieder auf. Es ist deshalb bei der Berechnung von multiplen Regressionsanalysen äußerst wichtig, die Statistiken zur Kollinearität, die in einigen Software-Paketen (etwa SAS) ausgegeben werden können, sorgfältig zu analysieren.

Das Kollinearitätsproblem hat auch Folgen für Versuche, die relative Bedeutung (also den Erklärungsgehalt) von Einzel-Prädiktoren und Prädiktor-Gruppen für Kriteriumsmerkmale zu bestimmen. Gemeint sind hier Prozeduren, die als Varianzpartitionierungsverfahren in der pädagogisch-psychologischen Forschung populär geworden sind. Ein prominentes Mitglied der Partitionierungstechniken, die sogenannte Kommunalitätenanalyse, hat in der schulischen Sozialisationsforschung große Bedeutung erlangt und wird auch heute noch gerne eingesetzt (vgl. etwa Helmke, 1992), obwohl sie in der Methodenliteratur durchaus umstritten ist. Die grundsätzliche Problematik wird im folgenden kurz skizziert.

4.1.1 Probleme und Möglichkeiten der Kommunalitätenanalyse

Das Prinzip der Kommunalitätenanalyse besteht kurzgefaßt darin, bei der Varianzaufklärung einer gegebenen Kriteriumsvariablen durch eine Reihe von Einzelprädiktoren oder Prädiktor-Blöcken die spezifischen und die gemeinsamen Varianzanteile der Prädiktoren an der Kriteriumsvarianz voneinander zu separieren (vgl. Cooley & Lohnes, 1976). Ein solches Vorgehen ist im pädagogisch-psychologischen Bereich etwa dann von Interesse, wenn die relative Bedeutsamkeit kognitiver und nichtkognitiver Schülermerkmale für den Schulerfolg bestimmt werden soll. Das Verfahren zerlegt die Kriteriumsvarianz in zwei unterschiedliche Komponenten: (a) den spezifischen Anteil eines gegebenen Prädiktors oder Prädiktorblocks an der Kriteriumsvarianz, den dieser Prädiktor zusätzlich zu den anderen Prädiktoren aufklärt; (b) die Kommunalität („konfundierte“ Varianz) als denjenigen Anteil an der Kriteriumsvarianz, der durch zwei oder mehr Prädiktoren gemeinsam „erklärt“ wird. Proponenten der Kommunalitätenanalyse wie etwa Cooley und Lohnes (1976) empfehlen diese Technik nachdrücklich für praktische Anwendungen. Diese Sicht der Dinge scheint jedoch allzu optimistisch, da Kollinearitätsprobleme systematisch eingehen: je höher die Interkorrelation der Prädiktoren, desto niedriger wird der spezifische Anteil und desto höher wird die Kommunalität ausfallen. Kritiker der Kommunalitätenanalyse (etwa Pedhazur, 1982) machen dann auch auf gravierende Interpretationsprobleme aufmerksam, die sich aus diesem Dilemma ergeben.

Das Interpretationsproblem wird zusätzlich dadurch erschwert, daß man nicht weiß, wie Kommunalitäten höherer Ordnung (die bei mehr als zwei Prädiktoren oder Prädiktor-Blöcken entstehen) inhaltlich gedeutet werden sollen. Da sie auf der Korrelation von Prädiktoren basieren, lassen sie sich nicht als Interaktionen im üblichen (varianzanalytischen) Sinn auffassen. Das Problem der Interpretation wird schließlich besonders evident, wenn negative Vorzeichen bei den Kommunalitäten auftauchen. Dies ist dann der Fall, wenn einige Prädiktorvariablen miteinander positiv, andere miteinander negativ korrelieren. Angesichts dieser vielfältigen Schwierigkeiten ist der Einschätzung von Pedhazur (1982) zuzustimmen, daß das Verfahren der Kommunalitätenanalyse für den Zweck der Erklärung im Sinne der Bestimmung relativer Bedeutsamkeiten von Prädiktoren und Prädiktor-Blöcken im wesentlichen ungeeignet ist. Dagegen wird seine Relevanz für die Vorhersagefunktion, also etwa für die Klärung der Frage, welche Prädiktoren für die Vorhersagegleichung selektiert werden sollen, bedeutend höher eingeschätzt.

4.2 Strukturgleichungs-(Kausal-)Modelle

Weiterentwicklungen der Regressionsanalyse, die insbesondere dem Zweck dienen, Abhängigkeitsbeziehungen zwischen Prädiktoren genauer zu modellieren, sind im wesentlichen als Pfadmodelle mit manifesten Variablen und als Kausalmodelle mit latenten Variablen in die Literatur eingegangen.

4.2.1 Pfadanalysen mit manifesten Variablen

Der wohl wesentlichste Unterschied zwischen multiplen Regressionsmodellen und Pfadmodellen mit manifesten (beobachteten) Variablen besteht darin, daß bei multiplen Regressionsanalysen alle Prädiktoren als exogene (unabhängige) Variablen behandelt werden, während in der Pfadanalyse die Prädiktoren zusätzlich in exogene und endogene unterschieden werden. Die Pfadanalyse bietet also zusätzlich die Möglichkeit, theoretisch postulierte Zusammenhangsmuster zwischen Prädiktorvariablen in ihrer Einwirkung auf ein gegebenes Kriterium genauer zu überprüfen. Dies erlaubt eine Zerlegung der Variablen-Interkorrelationen in direkte und indirekte Einflüsse. Sogenannte Pfadkoeffizienten, die den Regressionskoeffizienten in ihrer Bedeutung vergleichbar sind, geben das Ausmaß der Effekte von theoretisch vorgeordneten auf nachgeordnete Modellvariablen wieder. Von direkten Effekten wird dann gesprochen, wenn die vorgeordnete Variable das nachgeordnete Merkmal unmittelbar beeinflusst (z. B.: IQ beeinflusst das Leseverständnis), während indirekte Effekte dann vorliegen, wenn der Einfluß einer vorgeordneten Variablen über „Umwege“ erfolgt (z. B.: IQ hat keinen direkten Pfad zu Leseverständnis, beeinflusst aber direkt die Lesefertigkeit, für die wiederum ein direkter Effekt auf das Leseverständnis beobachtbar ist).

In der Literatur werden diese Pfadmodelle oft auch als „Kausalmodelle“ tituliert. Kontroversen um diesen Begriff hat es vor allem deshalb gegeben, weil „Kausalität“ vielfach in Anlehnung an klassische philosophische Traditionen deterministisch konzipiert wird, und immer wieder betont wird, daß sich aus korrelativen Ausgangswerten keine „kausalen“ Effekte dieses Typs ableiten lassen. Im Zusammenhang mit Pfadanalysen läßt sich der Kausalitätsbegriff in der Tat nur unter Rückgriff auf molare, probabilistische „Kausalgesetze“ rechtfertigen. Typisch für solche Gesetze ist, daß sich beobachtete Effekte auf multiple Ursachen zurückführen lassen (vgl. Cook & Campbell, 1979). Für die richtige Einordnung des Kausalitätsbegriffs im Rahmen von Pfad- oder Strukturgleichungsmodellen ist es also wichtig zu wissen, daß es hier nicht um die Überprüfung deterministischer Kausalprinzipien, sondern allenfalls um die Annahme zeitlicher Sukzessionsgesetze gehen kann (vgl. Bagozzi, 1980; Bentler, 1980; Steyer, 1983). Aufgrund seiner wissenschaftstheoretischen Vorbela-

stetheit wird der Kausalitätsbegriff in den folgenden Ausführungen dennoch weitmöglichst vermieden.

4.2.2 Strukturgleichungsmodelle mit latenten Variablen

Strukturgleichungsmodelle der „zweiten Generation“ (Schneider, 1986) unterscheiden sich von den traditionellen pfadanalytischen Ansätzen grundlegend darin, daß zwischen latenten (nicht beobachteten) und gemessenen Variablen differenziert wird. Diese Modelle stellen Kombinationen faktoren- und regressionsanalytischer Ansätze dar und sind daher im Bereich der Erziehungs- und Sozialisationsforschung vor allem zur Analyse von umfangreichen Datensätzen geeignet: während über ein (faktorenanalytisches) Meßmodell die Beziehung zwischen den gemessenen Indikatoren und den latenten Konstrukten definiert wird, dient das Strukturgleichungsmodell dazu, die Abhängigkeitsbeziehungen zwischen den latenten Variablen festzulegen. Das Meßmodell kann im Sinne der schon beschriebenen konfirmatorischen Faktorenanalyse definiert werden, die es dem Forscher ermöglicht, theoriegeleitet Ladungsmuster von Variablen auf Faktoren festzulegen. Das eigentliche „Kausalmodell“ wird dann auf der Ebene der so definierten latenten Konstrukte spezifiziert, die in der Regel mehr interessieren als die zugrundeliegenden gemessenen Variablen.

Eine detailliertere Beschreibung dieses Ansatzes kann im Rahmen dieses Kapitels nicht geleistet werden. In neuerer Zeit sind mehrere Monographien bzw. Sammelbände erschienen, die Einführungen in den Themenbereich bieten. Die meisten basieren auf der von Jöreskog und Sörbom (1989) entwickelten Pfadanalyse mit latenten Variablen (LISREL = Analysis of Linear Structural Relationships) und geben z. T. detaillierte Anleitungen für die Arbeit mit LISREL (s. Andres, 1989; Byrne, 1989; Loehlin, 1987; Pfeifer & Schmidt, 1987). Methodisch anspruchsvollere Abhandlungen finden sich bei Hayduk (1987) oder im Sammelband von Möbus und Schneider (1986).

Die wesentlichen Vorzüge des LISREL-Ansatzes lassen sich wie folgt zusammenfassen (vgl. auch Schneider, 1991):

- a) Implizite Annahmen über Ursache-Wirkungs-Verhältnisse müssen in ein Modell transferiert werden, das sich anhand von empirischen Daten schätzen läßt.
- b) Die Unterscheidung zwischen einem Meß- und einem Strukturmodell erlaubt eine vergleichsweise ökonomische Spezifikation des Gesamtmodells, da jedes der theoretisch interessanten Konstrukte meist über mehrere gemessene Variablen repräsentiert wird.

- c) Die genannte Unterscheidung ermöglicht weiterhin die getrennte Schätzung von Meßfehlern in den manifesten Variablen und Spezifikations- bzw. Gleichungsfehlern im Strukturmodell. Große Spezifikationsfehler deuten darauf hin, daß das Strukturmodell nicht vollständig spezifiziert wurde und theoretisch wichtige Prädiktorvariablen fehlen.
- d) Strukturmodelle lassen sich über LISREL nicht nur schätzen, sondern auch anhand verschiedener „goodness-of-fit“-Tests im Hinblick auf ihre Anpassung an die Ausgangsdaten überprüfen. Existieren alternative Modellkonzeptionen, so können diese in unterschiedliche Strukturmodelle umgesetzt und im Hinblick auf ihre Anpassungsgüte direkt miteinander verglichen werden.
- e) Obwohl Strukturgleichungsmodelle im allgemeinen auf Korrelations- oder Kovarianzmatrizen basieren, lassen sich zusätzlich auch Mittelwertsstrukturen analysieren. Damit ist gemeint, daß immer dann, wenn Modelle für mehrere Gruppen (z. B. Schüler unterschiedlicher Klassenstufen) miteinander verglichen werden, Unterschiede in den geschätzten Mittelwerten von latenten Variablen erfaßt werden können.

In neuerer Zeit sind weitere Strukturgleichungsmodelle entwickelt worden, über die bestimmte Problempunkte von älteren LISREL-Versionen (LISREL VI) überwinden werden sollten. So sind etwa die besonderen Vorzüge von EQS (Equations; Bentler, 1985) darin zu sehen, daß zum einen die multivariate Normalverteilung nicht länger unabdingbare Voraussetzung ist, zum anderen keine Kenntnis der Matrixalgebra erforderlich ist, um Strukturgleichungsmodelle zu spezifizieren. Das Programm kann von daher als vergleichsweise benutzerfreundlich eingestuft werden.

Das von Muthén (1984, 1987) entwickelte Programmsystem LISCOMP (Analysis of Linear Structural Equations using a Comprehensive Measurement Model) ist dann besonders angemessen, wenn für die Analyse gemischte Daten zur Verfügung stehen. LISCOMP ermöglicht die simultane Analyse dichotomer, kategorialer, normaler und nicht-normaler kontinuierlicher Variablen. Für Anwender aus Gebieten der Erziehungs- und Sozialisationsforschung bieten alle drei aufgeführten Programmsysteme insbesondere dann Probleme, wenn komplexe Modellstrukturen analysiert werden sollen. Da alle genannten Prozeduren rechenintensiv operieren, steigen die Bearbeitungszeiten bei Erweiterung des Modells um zusätzliche Variablen bzw. Konstrukte nichtlinear an und erreichen schnell obere Grenzen. Für solche Fälle bieten sich explorative Prozeduren an, die als „soft modeling“-Verfahren bekannt wurden und im LVPLS (Latent Variable Partial Least Squares)-Programm von Lohmöller (1984, 1989) beeindruckend umgesetzt sind (vgl. Helmke, 1992; Schneewind, Beckmann & Engfer, 1983, für Anwendungsbeispiele aus dem Bereich der

schulischen und familiären Sozialisation). LVPLS macht keine Angaben über die Verteilung von Daten, läßt sich also auch für Variablen verwenden, die extrem schiefverteilt sind. Das Verfahren erlaubt allerdings im Unterschied zu den anderen beschriebenen Techniken nur exploratorische Analysen. Dies bedeutet, daß weder Struktur- bzw. Pfadkoeffizienten auf Signifikanz geprüft noch Modell-Testungen durchgeführt werden können.

Abschließend sei noch auf einige Probleme von Strukturgleichungsansätzen mit latenten Variablen hingewiesen. Obwohl über diese Verfahren theoretische Annahmen über Zusammenhangsmuster zwischen den erfaßten Merkmalen sicherlich sehr viel besser in statistische Analysemodelle umgesetzt werden können, als dies für konventionelle regressionsanalytische Verfahren der Fall ist, gibt es dennoch grundsätzliche Restriktionen. So ist ein gerade für Fragen der Erziehungspsychologie und Sozialisationsforschung gravierendes Problem darin zu sehen, daß sich Interaktionen zwischen einzelnen Merkmalen über diese Verfahren nicht angemessen modellieren lassen (vgl. auch Pekrun, 1990). Zwar sind erste Ansätze zur Lösung dieses Problems inzwischen verfügbar (vgl. Hayduk, 1987), doch noch immer mit vielen methodischen Problemen belastet.

Ein weiteres, im wesentlichen noch ungelöstes Problem dieser Verfahren bezieht sich auf die Frage, wie die Ergebnisse, also die für unterschiedliche Strukturmodelle gefundenen Lösungen, angemessen zu bewerten sind. Probleme bei der Modellbewertung entstehen insbesondere dadurch, daß fast alle verfügbaren Indizes zur Beurteilung der Anpassung des Modells an die Daten von der Stichprobengröße abhängig sind: je größer die verfügbare Stichprobe, desto größer auch die Wahrscheinlichkeit, daß die Bewertung des Modells negativ ausfällt. Dies gilt ausdrücklich nicht nur für den bei LISREL gebräuchlichen Chiquadrat-Anpassungs-Index, sondern für annähernd alle in der Literatur verwendeten Kennwerte.

Eine zusätzliche Schwierigkeit bei der Suche nach dem für die Daten angemessensten Modell ist darin zu sehen, daß Strukturgleichungsmodelle eigentlich im Sinne einer konfirmatorischen Hypothesenprüfung eingesetzt werden sollten. Häufig werden jedoch diese Ansätze vorwiegend zu exploratorischen Zwecken verwendet. Das Hauptziel dieser Vorgehensweise besteht demnach darin, die einmal konstruierten Modelle möglichst gut an den gerade verfügbaren Datensatz anzupassen. Bei LISREL kann diese Strategie beispielsweise dadurch „optimiert“ werden, daß man aufgrund der Modifikationsindizes sein Modell sukzessive solange ändert, bis eine befriedigende Datenanpassung gefunden ist. Dies hat natürlich mit theoriegeleiteter Hypothesentestung nichts mehr zu tun, sondern führt lediglich zu einer Überanpassung des Modells an die Daten der gerade verfügbaren Stichprobe. Kreuzvalidierungen eines einmal

„erfolgreichen“ Modells an neuen Datensätzen sind demnach unerlässlich, um genauere Aufschlüsse über seine empirische Bewährung zu erlangen (vgl. Cudeck & Browne, 1983; Opwis, Gold & Schneider, 1987).

5 Längsschnittmethoden

Eine zunehmende Zahl neuerer Arbeiten aus dem Bereich der Erziehungs- und Sozialisationsforschung sind als Längsschnittstudie konzipiert, also so angelegt, daß die gleiche Personen-Stichprobe über einen bestimmten Zeitraum hinweg mehrmals untersucht wird. Es lassen sich je nach Zielvorstellung der Autoren unterschiedliche Untersuchungstypen unterscheiden, je nachdem, ob der Schwerpunkt in der Beschreibung oder Erklärung von Entwicklungsvorgängen liegt (vgl. Baltes & Nesselroade, 1979; Schneider, 1989). Die breite Palette von längsschnittlichen Ansätzen läßt sich grob in zwei Hauptlinien einteilen: Eine Kategorie von Studien legt den Schwerpunkt auf die Analyse intraindividuelle Veränderungen in spezifischen Entwicklungsmerkmalen, was methodisch etwa über Wachstumskurvenmodelle geschieht (vgl. Bryk & Raudenbush, 1987). Hier steht die Frage der Kontinuität/Diskontinuität von Entwicklungsverläufen im Vordergrund. Eine zweite Kategorie von Studien fokussiert auf die Stabilität/Instabilität inter- und intraindividuelle Unterschiede über die Zeit hinweg. Da Fragestellungen dieser Art für die Erziehungs- und Sozialisationsforschung weitaus typischer sind, wird in der Folge lediglich auf Grundprobleme und Möglichkeiten dieses Ansatzes näher eingegangen (Diskussionen zu Einsatzmöglichkeiten von Wachstumskurvenmodellen finden sich etwa bei Appelbaum & McCall, 1983; Schneider, 1993).

Längsschnittstudien zur Analyse intraindividuelle Veränderungen in interindividuellen Unterschieden bauen in der Regel auf Korrelations- oder Kovarianzstrukturen auf. Zur Analyse der Daten werden üblicherweise korrelations- und regressionsanalytische Verfahren herangezogen. Von vielen Autoren wird hervorgehoben, daß sich gerade längsschnittliche Daten besonders gut für den Einsatz von Strukturgleichungsmodellen eignen, da durch den Einbezug der Zeitdimension Fragen nach Abhängigkeitsbeziehungen zwischen Variablen eindeutig beantwortbar sind. Obwohl dies im Grundsatz sicherlich zutrifft und viele der oben beschriebenen pfadanalytischen Ansätze gerade in Längsschnittstudien wertvolle Erkenntnisse liefern können, werden ihre Möglichkeiten teilweise überschätzt. Ein Beispiel aus dem Bereich der schulischen Sozialisation soll dies verdeutlichen.

In einigen neueren Studien zum Thema Schule und Persönlichkeitsentwicklung hat man versucht, komplexe kausale Abhängigkeiten zwischen theoretischen

Konstrukten über die Analyse reziproker kausaler Effekte zu explorieren. Theoretische Ansätze zur Entwicklung des schulischen Selbstkonzepts gehen beispielsweise davon aus, daß die Ausprägung des Selbstkonzepts einerseits schulische Leistungen beeinflusst, andererseits Ergebnisse in Leistungstests sich wiederum auf das Selbstkonzept auswirken. Solche Wechselwirkungen sind gemeint, wenn von „reziproken kausalen Effekten“ die Rede ist. Viele Forscher interessiert nun die Frage, wo die „kausale Prädominanz“ angesiedelt, d. h., von welcher Variable die stärkeren Impulse ausgehen. Wie läßt sich dieses Problem methodisch lösen?

Eine ältere und äußerst problematische Verfahrensweise besteht darin, sogenannte „cross-lagged“-Korrelationen zu berechnen. Bei diesem Ansatz geht es speziell um die Analyse kreuzverzögerter Korrelationen zwischen den interessierenden Merkmalen. Trotz seiner Beliebtheit ist dieser Ansatz wissenschaftlich wertlos, da zahlreiche methodische Unzulänglichkeiten nachgewiesen wurden (vgl. Rogosa, 1980). Ein Hauptproblem ist hier sicherlich darin zu sehen, daß aus Korrelationen kausale Effekte abgeleitet werden.

Die Frage nach reziproken Abhängigkeiten scheint sich über Strukturgleichungsmodelle mit latenten Variablen vergleichsweise besser beantworten zu lassen. Ein übliches Verfahren besteht darin, autoregressive Modelle zu spezifizieren, bei denen (für unser Beispiel) sich Selbstkonzept und Schulleistung jeweils unabhängig voneinander über die Zeit hinweg entwickeln, und diesen Simplex-Modellen Alternativ-Spezifikationen gegenüberzustellen, die zusätzlich kausale Abhängigkeitsstrukturen zwischen den beiden theoretischen Konstrukten beinhalten. Bei der Feststellung kausaler Prädomonanz wird im Prinzip ähnlich wie bei der „cross-lagged panel“-Analyse verfahren, nur daß diesmal die Größe der zeitverzögerten Pfadkoeffizienten analysiert wird.

Welche methodischen Probleme sind bei diesem Ansatz zu beachten? Ein gewisses Dilemma besteht darin, daß die Zeitstabilität der theoretischen Konstrukte entscheidend für die Ausprägungsmöglichkeiten der kreuzverzögerten Beziehungen ist. Sind etwa die in unserem Beispiel zugrundegelegten Schulleistungen und das Selbstkonzept enorm stabil über die Zeit, so wird die Varianz des Merkmals zu t_2 durch dasselbe Merkmal zu t_1 fast vollkommen erklärt. Es verbleibt demnach für das andere Merkmal keine Möglichkeit, substantiell zur Varianzerklärung beizutragen. Die kreuzverzögerten Pfadkoeffizienten fallen also numerisch klein und praktisch unbedeutend aus. Sind die Zeitstabilitäten in den interessierenden Merkmalen dagegen relativ gering, wird es meist zu bedeutsameren kreuzverzögerten Pfadkoeffizienten kommen, die sich in der Größe erheblich unterscheiden können und demnach im Hinblick auf die Frage nach der kausalen Prädominanz inhaltlich deuten lassen. Voraussetzung für eine solche Interpretation ist allerdings, daß die mangelnde Zeitsta-

bilität der Konstrukte nicht auf unzureichende Reliabilität der zugrundeliegenden Testinstrumente zurückzuführen ist. Es ist ein großer Vorzug der Strukturgleichungsmodelle mit latenten Variablen, daß sich diese Frage hier prinzipiell klären läßt (vgl. Rudinger & Wood, 1990).

Anhand der Diskussion dieses Beispiels sollte deutlich geworden sein, daß auch die für die Analyse von Längsschnittdaten prädestinierten Kausalmodelle mit latenten Variablen ihre Grenzen haben, wenn auch das Anwendungsspektrum insgesamt beeindruckend sein kann. Es wurde schon erwähnt, daß sich damit neben Kovarianz- auch Mittelwertsstrukturen analysiert werden können, also längsschnittliche Verläufe in den Mittelwerten latenter Variablen geschätzt werden können (vgl. auch McArdle & Epstein, 1987). Neuere Modellentwicklungen berücksichtigen stärker Situationen, in denen statt der üblichen quantitativen Variablen lediglich diskrete (kategoriale) Merkmale für die Datenanalyse vorliegen (vgl. die Beiträge zu statistischen Längsschnittmethoden im den von von Eye herausgegebenen Sammelbänden (von Eye, 1990 b, 1990c). Die letzten Jahre haben in dieser Hinsicht enorme Fortschritte gebracht, und es sieht wahrlich nicht so aus, als wenn in Zukunft mit Stagnationen in diesem Bereich zu rechnen wäre.

6 *Schlußbemerkungen*

Wenn auch im vorliegenden Beitrag der Versuch gemacht wurde, die wesentlichen methodische Ansätze der empirischen Erziehungs- und Sozialisationsforschung darzustellen, so steht doch außer Frage, daß hier nur eine Auswahl von Datenerhebungs- und -analysemethoden präsentiert werden konnte. Diese Auswahl orientierte sich nicht zuletzt an bestimmten Voreingenommenheiten des Verfassers, der die Mehrzahl der empirischen Studien zum Thema als typische Felduntersuchungen einstuft, die im Vergleich zu experimentellen oder quasi-experimentellen Ansätzen mit weitaus umfangreicheren methodischen Problemen zu kämpfen haben. Aus Platzgründen wurde die Darstellung auf diesen Schwerpunktbereich beschränkt.

Ein gewisses Schwergewicht wurde weiterhin darauf gelegt, die Relevanz von Informationen über die Verteilung und das Meßniveau der Daten für die Auswahl von statistischen Analysemethoden und damit die Gefahr von allzu sorglosem Umgang mit diesem Problem zu verdeutlichen. Selbst die elegantesten Strukturgleichungssysteme bringen irrelevante und irreführende Ergebnisse, wenn sie mit inadäquaten Daten gespeist werden. Es sollte mit den vorliegenden Ausführungen der doch bei vielen Kollegen vorfindbaren Grundhaltung entgegengewirkt werden, daß angesichts der scheinbaren Robustheit

vieler statistischer Analysetechniken gründliche Datenexplorationen überflüssig sind.

Trotz der z. T. kritischen Diskussion vieler multivariater Methoden hat der vorliegende Beitrag zweifellos erkennen lassen, daß sein Verfasser letztendlich an die grundsätzliche Angemessenheit und Entwicklungsperspektive der quantitativ ausgerichteten Erziehungs- und Sozialisationsforschung glaubt. Der Tenor des Beitrags steht demnach durchaus in Kontrast zu früheren Überblicksarbeiten (etwa Köckeis-Stangl, 1980), in denen der „objektivistisch-quantitative Methodenkanon“ äußerst kritisch als „atheoretische Computerjagd nach hohen Korrelationskoeffizienten“ charakterisiert und geradezu als Hemmschuh für die gegenwärtigen Aufgaben der Sozialisationsforschung betrachtet wurde. Aus der hier vorgelegten Übersicht geht jedoch auch hervor, daß bei der Anwendung quantitativer Methoden oftmals Abstriche im Hinblick auf die Repräsentativität und die Validität der Befunde zu machen sind; viele der von Köckeis-Stangl vorgebrachten Kritikpunkte lassen sich in der Tat nicht von der Hand weisen.

Aus Platzgründen war es mir nicht möglich, näher auf methodische Ansätze einzugehen, die als Alternativansätze zum traditionellen Methodenrepertoire anzusehen sind und von Köckeis-Stangl (1980) als „qualitativ-interpretative“ Methoden (etwa die „pragmatischen Diskursanalyse“ oder das „Verfahren pragmatischer Dokumentation“) bezeichnet wurden. Der Leser sei in diesem Zusammenhang auf die Diskussion dieser Ansätze bei Köckeis-Stangl sowie weiterhin auf den Überblick von Gstettner (1980) zu biographischen Ansätzen verwiesen. Vielleicht mag er sich danach meinem (im übrigen auch schon 1979 von Cook und Campbell berichteten) Eindruck anschließen, daß diese Verfahren trotz ihrer vordergründigen Attraktivität in methodischer Sicht wohl ähnlich problematisch sind wie viele ihrer vielgeschmähten quantitativen Gegenstücke. Die Forschungstrends der letzten zehn Jahre haben denn auch m. E. in keiner Weise den Eindruck aufkommen lassen, daß wir in methodischer Hinsicht einem Paradigmenwechsel entgegengehen. Allerdings dürfte Hurrelmann und Ulich (1991) wohl zuzustimmen sein, wenn sie bei Sozialisationsforschern in jüngerer Zeit zunehmend mehr Bereitschaft erkennen, unterschiedliche methodische Ansatzweisen zu akzeptieren und in den eigenen Forschungsarbeiten zu kombinieren. Es hat den Anschein, daß die noch zu Anfang der achtziger Jahre beobachtbare Kluft zwischen „positivistischen“ Positionen und hermeneutischen Methodologien deutlich reduziert, wenn nicht gar überwunden ist.

Autorenhinweis: Für seine kritische Kommentierung einer früheren Fassung dieses Kapitels möchte ich Herrn PD Dr. Andreas Helmke (München) herzlich danken.

Literatur

- Aldenderfer, M.S. & Blashfield, R.K. (1984). *Cluster analysis*. Beverly Hills: Sage.
- Andres, J. (1989). *Grundlagen linearer Strukturgleichungsmodelle*. Frankfurt/M.: Lang.
- Appelbaum, M.J. & McCall, R.B. (1983). Design and analysis in developmental psychology. In P.H. Mussen (Ed.), *Handbook of child psychology* (Vol. 1, pp. 415-476). New York: Wiley.
- Bagozzi, R.P. (1980). *Causal models in marketing*. New York: Wiley.
- Baltes, P.B. & Nesselroade, J.R. (1979). History and rationale of longitudinal research. In J.R. Nesselroade & P.B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.
- Bentler, P.M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31, 419-456.
- Bentler, P.M. (1985). *Theory and implementation of EQS: A structural equations program*. Los Angeles: BMDP Statistical Software Corporation.
- Berry, W.D. & Feldman, S. (1985). *Multiple regression in practice*. Beverly Hills: Sage.
- Bortz, J. (1989). *Statistik für Sozialwissenschaftler* (3. Aufl.). Berlin: Springer.
- Bredenkamp, J. & Feger, H. (Hrsg.). (1983). *Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie I: Forschungsmethoden der Psychologie. Band 4: Strukturierung und Reduzierung von Daten*. Göttingen: Hogrefe.
- Bryk, A.S. & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 118, 147-158.
- Byrne, B.M. (1989). *A primer of LISREL*. New York: Springer.
- Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cooley, W.W. & Lohnes, P.R. (1976). *Evaluation research in education*. New York: Wiley.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation*. Chicago: Rand McNally.
- Cudeck, R. & Browne, M.W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, 18, 147-167.
- Draper, N.R. & Smith, H. (1981) *Applied regression analysis* (2nd ed.). New York: Wiley.
- Eckes, T. & Roßbach, H. (1980). *Clusteranalysen*. Stuttgart: Kohlhammer.
- Eye, A. v. (1990a). *An introduction into configural frequency analysis*. Cambridge: Cambridge University Press.
- Eye, A. v. (Ed.) (1990b). *Statistical methods in longitudinal research (Vol. 1): Measuring and structuring change*. New York: Academic Press.
- Eye, A. v. (Ed.) (1990c). *Statistical methods in longitudinal research (Vol. 2): Time series and categorical longitudinal data*. New York: Academic Press.

- Feger, H. (1983). Planung und Bewertung von wissenschaftlichen Beobachtungen. In H. Feger & J. Bredenkamp (Hrsg.), *Datenerhebung, Enzyklopädie der Psychologie, B/1/2* (S. 1-75). Göttingen: Hogrefe.
- Feger, H. & Bredenkamp, J. (Hrsg.). (1983 a). *Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie I: Forschungsmethoden der Psychologie. Band 2: Datenerhebung*. Göttingen: Hogrefe.
- Feger, H. & Bredenkamp, J. (Hrsg.). (1983 b). *Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie I: Forschungsmethoden der Psychologie. Band 3: Messen und Testen*. Göttingen: Hogrefe.
- Gniech, G. (1982). Störeffekte bei psychologischen Untersuchungen im Feld. In J.-L. Patry (Hrsg.), *Feldforschung* (S. 259-276). Bern: Huber.
- Gstettner, P. (1980). Biographische Methoden in der Sozialisationsforschung. In K. Hurrelmann & D. Ulich (Hrsg.), *Handbuch der Sozialisationsforschung* (S. 371-394). Weinheim: Beltz.
- Ischi, N. (1982). Methodologische Probleme systematischer Verhaltensbeobachtung im Feld. In J.-L. Patry (Hrsg.), *Feldforschung* (S. 277-316). Bern: Huber.
- Hayduk, L. A. (1987). *Structural equation modeling with LISREL*. Baltimore: John Hopkins University Press.
- Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Göttingen: Hogrefe.
- Helmke, A. & Renkl, A. (1992). Das Münchner Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostika*, 38, 130-141.
- Hildebrand, D. K., Laing, D. J. & Rosenthal, H. (1977). *Prediction analysis of cross classifications*. New York: Wiley.
- Hron, A. (1982). Interview. In G. L. Huber & H. Mandl (Hrsg.), *Verbale Daten* (S. 119-140). Bern: Huber.
- Holm, K. (1975). Die Frage. In K. Holm (Hrsg.), *Die Befragung 1: Der Fragebogen - die Stichprobe* (S. 32-126). München: Francke.
- Hurrelmann, K. (1986). *Einführung in die Sozialisationsforschung*. Weinheim: Beltz.
- Hurrelmann, K. & Ulich, D. (1991). Gegenstands- und Methodenfragen in der Sozialisationsforschung. In K. Hurrelmann & D. Ulich (Hrsg.), *Neues Handbuch der Sozialisationsforschung* (S. 3-20). Weinheim: Beltz.
- Jöreskog, K. G. & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge: Abt Books.
- Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7. A guide to the program and applications*. Chicago: SPSS Inc.
- Kerlinger, F. N. (1977). *Foundations of behavioral research* (2nd edition). London: Holt, Rinehart & Winston.
- Kim, J.-O. & Mueller, C. W. (1978). *Factor analysis statistical methods and practical issues*. Beverly Hills: Sage.

- Klecka, W.R. (1981). *Discriminant analysis*. Beverly Hills: Sage.
- Köckeis-Stangl, E. (1980). Methoden der Sozialisationsforschung. In K. Hurrelmann & D. Ulich (Hrsg.), *Handbuch der Sozialisationsforschung* (S. 321-370). Weinheim: Beltz.
- Krauth, J. & Lienert, G.A. (1973). KFA. *Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin*. Freiburg/München: Alber.
- Lachenbruch, P.A. (1975). *Discriminant analysis*. New York: Hafner.
- Lewis-Beck, M.S. *Applied regression - An introduction*. Beverly Hills: Sage.
- Loehlin, J.C. (1987). *Latent variable models*. Hillsdale: Erlbaum.
- Lohmöller, J.B. (1984). *LVPLS program manual*. Köln: Zentralarchiv für empirische Sozialforschung.
- Lohmöller, J.B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg: Physica.
- Long, J.S. (1983). *Confirmatory factor analysis*. Beverly Hills: Sage.
- Maxwell, A.E. (1975). Limitations on the use of the multiple linear regression model. *British Journal of Mathematical and Statistical Psychology*, 28, 51-62.
- McArdle, J.J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110-133.
- Milligan, G.W. (1981). A review of Monte Carlo tests of cluster analysis. *Multivariate Behavioral Research*, 16, 379-407.
- Milligan, G.W. & Cooper, M.C. (1987). Methodological review: Clustering methods. *Applied Psychological Measurement*, 11, 329-354.
- Möbus, C. & Schneider, W. (Hrsg.). (1986). *Strukturmodelle zur Analyse von Längsschnittdaten und Zeitreihen*. Bern: Huber.
- Mummendey, H.D. (1979). Methoden und Probleme der Messung von Selbstkonzepten. In S.-H. Filipp (Hrsg.), *Selbstkonzept-Forschung* (S. 171-202). Stuttgart: Klett-Cotta.
- Muthén, B.O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Muthén, B.O. (1987). *LISCOMP - Analysis of linear structural equations using a comprehensive measurement model*. Mooresville, Scientific Software Inc.
- Opwis, K., Gold, A. & Schneider, W. (1987). Möglichkeiten der Kreuzvalidierung von Strukturgleichungsmodellen. *Psychologische Beiträge*, 29, 60-77.
- Pedhazur, E.J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.
- Pekrun, R. (1990). Social support, achievement evaluations, and self-concepts in adolescence. In L. Oppenheimer (Ed.), *The self-concept* (pp. 107-119). Berlin: Springer.
- Petermann, F. (1987). Daten, Dimensionen, Verfahrensweisen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (2.Aufl.; S. 1017-1060). München u. Weinheim: Psychologie-Verlags-Union.

- Pfeifer, A. & Schmidt, P. (1987). *LISREL: Die Analyse komplexer Strukturgleichungssysteme*. Stuttgart: Fischer.
- Renkl, A. & Helmke, A. (1993). Generalisierungstheorie: Prinzip, Nutzen und Anwendungsmöglichkeiten. *Empirische Pädagogik*, 7.
- Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, 88, 245-258.
- Rudinger, G., Chaselon, F., Zimmermann, J.E. & Henning, J.H. (1985). *Qualitative Daten - Neue Wege sozialwissenschaftlicher Methodik*. München: Urban & Schwarzenberg.
- Rudinger, G. & Wood, P.K. (1990). N's, times and number of variables in longitudinal research. In D. Magnusson & L. Bergman (Eds.), *Problems of data analysis in longitudinal studies* (pp. 157-180). Cambridge: Cambridge University Press.
- Scheibler, D. & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods. *Multivariate Behavioral Research*, 20, 283-304.
- Schneewind, K.A., Beckmann, M. & Engfer, A. (1983). *Eltern und Kinder*. Stuttgart: Kohlhammer.
- Schneider, W. (1986). Strukturgleichungsmodelle der zweiten Generation: Eine Einführung. In C. Möbus & W. Schneider (Hrsg.), *Strukturmodelle für Längsschnittdaten und Zeitreihen* (S. 13-26). Bern: Huber.
- Schneider, W. (1989). Problems of longitudinal studies with children: Practical, conceptual, and methodological issues. In M. Brambring, F. Lösel & H. Skowronek (Eds.), *Children at risk: Assessment, longitudinal research, and intervention* (pp. 313-335). New York: de Gruyter.
- Schneider, W. (1991). Methodische Probleme und Möglichkeiten schulbezogener Längsschnittforschung. In R. Pekrun & H. Fend (Hrsg.), *Schule und Persönlichkeitsentwicklung - Ein Resümee der Längsschnittforschung* (S. 57-82). Stuttgart: Enke.
- Schneider, W. (1993). The longitudinal study of motor development: Methodological issues. In A.F. Kalverboer, B. Hopkins & R.H. Geuze (Eds.), *Motor development in early and later childhood: Longitudinal approaches* (pp. 317-342). Cambridge: Cambridge University Press.
- Schneider, W. & Scheibler, D. (1983). Probleme und Möglichkeiten bei der Bewertung von Clusteranalyse-Verfahren. I: Ein Überblick über einschlägige Evaluationsstudien. *Psychologische Beiträge*, 25, 208-237.
- Schumm, W. (1990). Evolution of the family field: Measurement principles and techniques. In J. Touliatos, B.F. Perlmutter & M.A. Strauss (Eds.), *Handbook of family measurement techniques* (pp. 23-36). Newbury Park: Sage.
- Schwarzer, R. (1983). Befragung. In J. Bredenkamp & H. Feger (Hrsg.), *Datenerhebung, Enzyklopädie der Psychologie, B/II/2* (S. 302-320). Göttingen: Hogrefe.
- Stallings, J.A. & Mohlman, G.G. (1988). Classroom observation techniques. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (pp. 469-473). Oxford: Pergamon Press.

- Steyer, R. (1983). Modelle zur kausalen Erklärung statistischer Zusammenhänge. In J. Bredenkamp & H. Feger (Hrsg.) *Strukturierung und Reduzierung von Daten, Enzyklopädie der Psychologie, B/II/4* (S.59-153). Göttingen: Hogrefe.
- Trolldenier, H.-P. (1985). *Verhaltensbeobachtung in Erziehung und Unterricht*. Frankfurt/M.: Fachbuchhandlung für Psychologie.
- Wishart, D. (1978). *CLUSTAN. User manual. Program library unit*. Edinburgh: Edinburgh University.
- Wottawa, H. (1988). *Psychologische Methodenlehre - Eine orientierende Einführung*. Weinheim u. München: Juventa.

Strukturierung und Prozesse
der Persönlichkeitsentwicklung