

Peter Schmidt

Sozialtechnologie und Handlungsforschung als experimentelle Handlungs- und Forschungsstrategien: Anwendungsprobleme *

Motto 1:

The basic question involves comparing the ethics of gathering information systematically about our large scale programs with the ethics of haphazardly implementing and changing treatments as so routinely happens in education, welfare and other areas. Conducting such investigations is far preferable to the current practice of "fooling around with people" without their informed consent (J. P. GILBERT et al., 1975, 182)

Motto 2:

Keine Experimente; (Wahlslogan der Christlich-Demokratischen-Union im Wahlkampf 1957)

1. Einführung in die Problemstellung

Im Rahmen einer früheren Arbeit (EICHNER, SCHMIDT, 1974) hatten wir vom Standpunkt des Kritischen Rationalismus aus versucht, die wichtigsten Postulate der Handlungs- bzw. Aktionsforschung des Sammelbandes von HAAG et al (1972) zu präzisieren. Darüberhinaus prüften wir die Frage, ob diese Postulate - von uns in verschiedene Ziele und Mittel eingeteilt - dem Anspruch einer neuen Methode gerecht werden. In einer abschließenden Beurteilung kamen wir zu dem Schluß, daß die Vertreter der Aktionsforschung zwar auf wichtige Fragen bei der Lösung praktischer Probleme hingewiesen haben, daß sie aber keine neuen Mittel zur Lösung dieser Probleme liefern. HAAG et al. haben darauf in einem ausführlichen Artikel ge-

* Diese Arbeit ist eine veränderte Fassung eines Papiers, welches auf der Jahrestagung der Deutschen Gesellschaft für Erziehungswissenschaft 1976 in Duisburg in der Arbeitsgruppe "Methodologische Probleme handlungsorientierter Curriculuminnovation" gehalten und diskutiert wurde. Diese wurde von H. HAFT und U. HAMEYER geleitet. Für eine Reihe kritischer Argumente danke ich den Teilnehmern dieser Arbeitsgruppe: Zu meinem Erstaunen hielten mich viele Vertreter der Handlungsforschung, nach Diskussionen im kleinen Kreis, nicht mehr für einen kritischen Rationalisten. Ich weiß nicht, ob ich wegen der Überzahl an Handlungsforschern in dieser Gruppe mich ihnen gegenüber überkonform verhielt oder ob die beiden Ansätze bei näherer Analyse doch nicht so verschieden sind.

Für eine Vielzahl kritischer Hinweise und Verbesserungsvorschläge danke ich UWE HAMEYER, KLAUS HENNING HANSEN, CHRISTIAN LÜDEMANN, JÜRGEN LEHMANN, HANS JOACHIM MÜLLER, RENATE MÜLLER, GABRIELE SZELINSKI, CLAUDIA WAWRZINEK und GISELA WIENECKE.

antwortet (1975), in dem sie die These vertraten, daß die Handlungsfor- schung ein eigenes Paradigma darstelle im Sinne von KUHN.

Diese Kontroverse hat sich nun leider zu einem Musterbeispiel des Aneinandervorbeiredens entwickelt; ich möchte daher in diesem Beitrag die Grundsatzdebatte in ihrer allgemeinen Form nicht fortsetzen, sondern relativ konkrete Vorschläge machen, wie man bei Projekten, die die Auf- gabe haben, praktische Probleme zu lösen, vorgehen kann. Dabei werde ich keine vollständige Methodologie der Evaluation liefern, sondern mich schwerpunktmäßig auf bestimmte Bereiche konzentrieren.

Im ersten Teil werde ich die Frage aufwerfen, welche Bedeutung Theo- rien zur Lösung sozialer Probleme haben. Danach werde ich darauf einge- hen, inwiefern empirisch geprüfte Theorien über Problemlösungsverhalten in Gruppen und organisationssoziologische Hypothesen zur Planung und Or- ganisation von Handlungsforschungsprojekten angewandt werden können. Dies erscheint mir deshalb besonders wichtig, weil in den bisher veröffent- lichten Handlungsforschungsprojekten zumindest nie explizit von dem bisher bekannten Bestand an Hypothesen über Problemlösungsverhalten (vgl. z.B. SHAW, 1971) und organisationssoziologische Tatbestände (vgl. z.B. BLAU, SCHOENHERR, 1972) - einschließlich der Hypothesen über soziale Netzwer- ke - Gebrauch gemacht wurde. Im letzten Teil schließlich werde ich dann näher auf eine Methode eingehen, die mir gerade für Handlungsforschungs- projekte besonders geeignet erscheint (vgl. CAMPBELL, 1969; 1970; 1971; 1975; CAMPBELL, STANLEY, 1970). Es handelt sich um die Methode des Quasiexperiments. Dabei soll gezeigt werden, daß gerade diese Methode für Handlungsforschungsprojekte adäquat ist, wenn man unterstellt, daß es sich bei Handlungsforschung als auch Sozialtechnologie um soziale Experi- mente handelt (vgl. MÜLLER, 1977).

Ich werde diese Methode anhand eines einfachen Anwendungsbeispiels demonstrieren. Daran anschließend werde ich eine Reihe von Kritikpunkten an herkömmlichen Versuchsplänen quasiexperimenteller Art (vgl. CAMP- BELL, STANLEY, 1970) formulieren, um dann die Verwendung eines allge- meinere statistischen Verfahrens vorzuschlagen, das diesen Kritikpunkten Rechnung trägt und außerdem einen Beitrag zum Problem der Generalisier- barkeit leistet. Um dies anschaulich darzustellen, erweitere ich das er- wählte Anwendungsbeispiel, um mit dessen Hilfe dann zu demonstrieren, wie man hierfür einen quasiexperimentellen Versuchsplan aufstellen kann. Abschließend werde ich einen Vorschlag unterbreiten, wie man die subjek- tiven Erfahrungen aus Aktionsforschungsprojektvorhaben mit objektiven Daten sinnvoll verknüpfen kann. Um Mißverständnisse zu vermeiden, soll im folgenden definiert werden, was unter objektiven Daten verstanden wer- den soll:

- (a) bei wiederholter Beobachtung der gleichen Reaktionen durch den gleichen Beob- achter ergeben sich die gleichen Beobachtungen bzw. Daten,
- (b) bei wiederholter Beobachtung der gleichen Reaktionen durch verschiedene Beob- achter ergeben sich die gleichen Beobachtungen bzw. Daten (vgl. hierzu z.B. GALTUNG, 1967, 27-29).

In Form eines Übersichtsschemas (Schema 1) wollen wir unser Vorgehen nochmals illustrieren:

Schema 1: Art des Vorgehens

	Anwendung von Theorien	Anwendung quasi- experimenteller Verfahren
Organisation von Handlungsforschungs- projekten	X	
Lösung sozialer Probleme	X	X

Ein X bedeutet, daß der entsprechende Bereich behandelt wird. Die Diskus- sion der Anwendung von quasiexperimentellen Verfahren auf die Organisa- tion von Handlungsforschungsprojekten erscheint uns deswegen nicht not- wendig, da dort keine neuen Probleme gegenüber der Anwendung quasiexpe- rimenteller Verfahren zur Lösung sozialer Probleme auftreten.

2. Zur Bedeutung von Theorien und experimentellem Vorgehen

Ein wesentliches Merkmal von Handlungsforschung besteht darin, daß man nicht von festen Zielen und Mitteln ausgeht, sondern im Rahmen eines Handlungsforschungsprojektes die Ziele und Mittel erarbeitet; darüberhin- aus sollen zunächst erarbeitete Ziele als auch Mittel durch wiederholte Evaluation revidiert werden.

Dies bedeutet die Aufhebung des herkömmlichen Ziel-Mittel-Schemas, weil Ziele und Mittel gleichermaßen als Handlungsvoraussetzungen kriti- sierbar sind. Diese Aufhebung wird auch vom Kritischen Rationalismus vor- genommen (vgl. ALDRUP, 1971, 124 und MÜLLER, 1977, Abschnitt 3.1.2. und 4.1.).

Im Rahmen unserer früheren Arbeit (EICHNER, SCHMIDT, 1974) hatten wir das Vor- gehen bei der Entwicklung von Sozialtechnologien zu einseitig als statischen rationalen Entscheidungsprozeß dargestellt. Wir meinten damit, daß man bei der Lösung eines praktischen Problems festlegt, was die zu verändernde Größe sei und in welchem Grade diese verändert werden soll. Danach sollte man prüfen, welche unabhängigen Variablen man einführt, um dieses Ziel zu erreichen. In einem weiteren Schritt leitet man im Rahmen einer Ziel-Mittel-Ableitung die optimale Maßnahmenkombination ab. Diese Darstellung ist allerdings sehr vereinfacht.

Experimentelles Vorgehen zur Durchführung sozialer Reformen bzw. zur Lösung sozialer Probleme kann demgegenüber folgendermaßen charakterisiert werden (vgl. CAMPBELL, 1969, und MÜLLER, 1977):

Man probiert neue Programme aus, die spezielle soziale Probleme lösen sollen. In diesem Zusammenhang lernen wir, ob diese Programme effektiv bzw. nicht effek- tiv sind und entscheiden, ob wir sie beibehalten, nachahmen, verändern oder absetzen. Man geht davon aus, daß es kein sicheres Wissen darüber gibt, wie erfolgreich diese Programme sind, sondern daß es darauf ankommt, aus Fehlern zu lernen und Kritik nicht zu unterdrücken bzw. Mängel nicht zu verschleiern. Im Rahmen einer Lösung eines praktischen Problems wird man also die Vorläufigkeit der Festlegung von Zie- len, entsprechenden Mitteln sowie des Wissens um die negativen Nebenwirkungen, möglichst berücksichtigen. Man erwartet nach der Durchführung von Maßnahmen nun

nicht, daß das Ziel erreicht und damit das Problem gelöst ist. Genau das Gegenteil dürfte in der Regel der Fall sein. Man erwartet bei der Lösung beliebiger praktischer Probleme, daß man das Ziel in der festgelegten Form so nicht erreicht und daß unerwünschte Nebenwirkungen auftreten. Daher kommt es darauf an, dies möglichst schnell zu bemerken und entsprechende Korrekturen durchzuführen. Ein objektives Verfahren, um dieses Programm durchführen zu können, bilden die quasiexperimentellen Versuchspläne, auf die wir später näher eingehen wollen.

Gehen wir nun auf die Rolle von Theorien und Alltagstheorien ein. Wir wollen dabei im folgenden zwei Anwendungsbereiche von Theorien und Alltagstheorien unterscheiden:

1. Anwendung bei der Organisation von Veränderungsprozessen bzw. Aktionsforschungsprojekten
2. Anwendung bei der Entwicklung praktischer Maßnahmen.

Was die Verwendung von Theorien unter diesen beiden Aspekten in Handlungsforschungsprojekten angeht, so stimme ich der Auffassung von GRUSCHKA zu, handlungsorientierte Begleitforschung sei z. Zt. noch theorieleiose Praxis bzw. praxislose Theorie (GRUSCHKA in diesem Band).

Welche Konsequenzen sind mit dieser Situation verbunden? Wenn die Wissenschaftler in Aktionsforschungsprojekten - wie dies oft zu beobachten ist - keine wissenschaftlichen Theorien, sondern genau wie die Praktiker nur ihre Alltagstheorien einbringen, um Probleme zu lösen, warum machen sie dann noch Wissenschaft und nicht ausschließlich Praxis? Es stellt sich dann die Frage, was dann noch wissenschaftliche von praktischer Tätigkeit unterscheidet. Von der Wissenschaft kann man unter solchen Voraussetzungen nichts Produktives erwarten. Die Rolle des Wissenschaftlers würde dann auf die Rolle des gruppendynamischen Trainers bzw. des Koordinators reduziert. Man müßte sich dann eigentlich als Wissenschaftler darauf konzentrieren, die Organisation von Problemlösungsprozessen in Gruppen und Organisationen zu verbessern, wobei auch hier bei Nichtanwendung von sozialwissenschaftlichen Theorien klassische Wissenschaftlertätigkeit völlig überflüssig wäre.

Nun stellt sich die Frage, ob man a) die Organisation von Aktionsforschungsprojekten und b) praktische Maßnahmen immer nur mit Alltagstheorien begründen kann, die man durch eine Diskussion in Form eines Diskurses erarbeitet (vgl. MOSER, 1975, 79-115; vgl. aber zur Kritik W. FACH 1974). Auch wenn man nicht übersteigert optimistisch über den jetzigen Stand der Sozialwissenschaft ist, bieten doch eine Reihe von Theorien bzw. Hypotheseninventaren Aussagen an, die man gerade bei der Lösung praktischer Probleme im Bereich der Pädagogik verwenden kann. Ich denke z. B. an die Wert-Erwartungstheorie von ATKINSON (ATKINSON, BIRCH, 1970), die kognitiven Gleichgewichtstheorien (vgl. in einer Übersicht IRLE 1975, S. 294-345), die Lerntheorie (vgl. zur Anwendung in der Soziologie OPP, 1972) und die Theorie des Neugierverhaltens (vgl. BERLYNE, 1960). Zu einer Integration dieser Theorien vergleiche man KAUFMANN (1975; 1977) und zu einer Anwendung auf die Erklärung der Diffusion von Neuerungen die Arbeit von KAUFMANN und SCHMIDT (1976).

-Wir vertreten hier nicht die These, daß diese Theorien in allen Fällen direkt praktisch umsetzbar sind bzw. in allen Fällen den Alltagstheorien von Praktikern überlegen sind. Wir meinen nur, daß, wenn der Wissenschaftler irgendeinen Nutzwert über die Beratung der technischen Anwendung objektiver

Verfahren wie quasiexperimenteller Versuchspläne hinaus haben soll, dann doch den, daß er Aussagen relevanter bestätigter Theorien bei der Problemlösung verwendet.

Eine weitere unerwünschte und nicht weniger wichtige Konsequenz der impliziten Verwendung von Theorien bildet die Ausschaltung einer Möglichkeit der Kritik an Programmen; denn die Explikation der zugrundeliegenden Theorie - unabhängig davon, ob es sich um eine Alltagstheorie oder eine empirisch geprüfte Theorie handelt - ermöglicht eine Kritik am jeweiligen praktischen Programm (vgl. MÜLLER 4.1.).

Im Unterschied zur jetzigen Praxis bei Handlungsforschungsprojekten würden wir also fordern, daß die Wissenschaftler in Handlungsforschungsprojekten in zwei Bereichen theoretisches und methodisches Wissen (z. B. über quasiexperimentelle Versuchspläne) in die Gruppendiskussion einbringen:

- a) bei der Organisation und Evaluation von Aktionsforschungsprozessen
- b) bei der Entwicklung und Evaluation praktischer Maßnahmen zur Lösung sozialer Probleme.

Dies bedeutet allerdings nicht, daß die Wissenschaftler deswegen eine Priester- oder Monopolstellung haben. Die Praktiker können und müssen aufgrund ihrer Kenntnis von speziellen Anfangsbedingungen sowie alternativen Theorien - auch wenn es Alltagstheorien sind - die Konzeptionen der Wissenschaftler kritisieren und korrigieren. So weiß z. B. nur eine Kindergärtnerin, welche Spiele Kindern im Kindergarten Spaß machen, während der lerntheoretisch geschulte Psychologe nicht weiß, welche Spiele in einer konkreten Anwendungssituation belohnend sind (vgl. hierzu allgemein H. WESTMEYER, 1973, 40-100 und 1976).

Den Punkt a) werden wir an konkreten Beispielen im nächsten Abschnitt behandeln, während der Punkt b) von uns im Zusammenhang mit der Aufstellung quasiesperimenteller Versuchspläne diskutiert wird.

Gehen wir jetzt noch kurz auf die Frage der Notwendigkeit objektiver Verfahren ein. Zunächst eröffnen uns diese Verfahren die Möglichkeit der Kritik von Programmen, sie können eine Dogmatisierung bestehender Praxis vermeiden und müssen keineswegs zur Festschreibung von Bestehendem führen.

Eine gut bestätigte empirische Theorie liefert uns ein weiteres Argument. Es ist deswegen notwendig, objektive Verfahren zur Prüfung der Wirkung von Maßnahmen anzuwenden, weil innerhalb von Gruppen nach der Theorie sozialer Vergleichsprozesse (vgl. FESTINGER, 1954 und IRLE, 1975, S. 165-174) Personen zwei unterschiedliche Realitätsprüfungen durchführen, um sich Gewißheit über die Richtigkeit ihrer Alltagstheorien und auch ihrer Wissenschaftstheorien zu verschaffen. Der erste Realitätstest besteht in der unmittelbaren Prüfung von Ereignissen in der Umwelt, die gemäß Erwartung bzw. Hypothese der jeweiligen Person auftreten müssen. Der zweite Realitätstest besteht im Vergleich der eigenen Erwartungen bzw. Hypothesen über Ereignisse mit den Annahmen anderer Personen über diesen Sachverhalt (vgl. IRLE, 1975, 165). Ohne im einzelnen auf die Hypothesen dieser Theorie einzugehen, scheint mir folgender Sachverhalt, der sich aus der Theorie ergibt, sehr wichtig zu sein. Wenn gar keine objektiven Verfahren als Realitätstest erster Art angewandt werden, wird sich die Wahr-

heitsfindung allein durch den Vergleich der eigenen Erwartungen mit denen der anderen Projektteilnehmer ergeben. Damit fehlt aber jegliche empirische Kontrolle, die als Korrektiv der Gruppenmeinung über den Erfolg bzw. Mißerfolg eines Programmes dienen kann. Es ist oft der Fall, daß man sich auf völlig unsinnige "Wahrheiten" per Gruppenmeinung einigt, die sich bei empirischer Prüfung relativ schnell als unsinnig erweisen. Damit ist nun aber nicht gemeint, daß nur in Gruppen ohne begleitende empirische Prüfung unsinnige "Wahrheiten" bzw. Problemlösungsvorschläge erarbeitet werden. Das Gleiche kann natürlich auch für empirische Arbeiten gelten, wenn diese nicht adäquat durchgeführt und interpretiert werden. Wenn nun auch die Durchführung objektiver Verfahren für die Evaluation des Erfolges von Maßnahmen eine wünschenswerte und notwendige 'Korrektur' der Beobachtungen einzelner Mitglieder eines Aktionsforschungsprojektes ist, so wird man jedoch aus Gründen des finanziellen und zeitlichen Aufwandes nicht in jedem Fall den Einsatz eines objektiven Verfahrens für sinnvoll halten. Wenn man jedoch ein längerfristiges Projekt wie die Kollegstufenreform oder die Evaluation der Gesamthochschulkonzeption durchführen will, scheint mir die Durchführung von objektiven Verfahren wie quasiexperimentellen Versuchsplänen notwendig und möglich zu sein. Bei kurzfristigeren und vom Umfang kleineren Problemen dürfte dagegen der Aufwand für die Verwendung objektiver Verfahren oft zu groß sein. Man kann also die Forderung nach Verwendung objektiver Verfahren nicht zum Dogma machen, sondern muß im Einzelfall entscheiden, ob dies sinnvoll ist. Die generelle Ablehnung von objektiven Verfahren scheint mir genauso dogmatisch zu sein, wie die Forderung, daß sie in jedem Falle angewendet werden müßten.

Im folgenden wollen wir nun auf einige ausgewählte Hypothesen eingehen, die meiner Ansicht nach für die Organisation von Aktionsforschungsprojekten verwendbar sein können. Diese Auswahl soll nur exemplarisch zeigen, wie man sozialwissenschaftliche Hypothesen verwenden kann. Für die Organisation solcher Projekte stehen noch wesentlich mehr Hypothesen zur Verfügung als wir hier behandeln können.

3. Zur Anwendung von Theorien bei der Organisation von Handlungsforschungsprojekten

In diesem Abschnitt wollen wir diskutieren, inwieweit man Hypothesen aus der Kleingruppenforschung und der Organisationssoziologie dazu verwenden kann, praktische Empfehlungen für Handlungsforschungsprojekte zu formulieren.

3.1. Hypothesen zur Organisation von Gruppenprozessen

Generell geht es auch in Handlungsforschungsprojekten darum, bestimmte Lösungen für Probleme zu erarbeiten, die man gemeinsam als Probleme identifiziert hat. Demnach könnte man nun prüfen, inwiefern speziell Hypothesen über die Determinanten von Problemlösungsverhalten in Gruppen sinnvoll bei der Organisation von Handlungsforschungsprojekten angewandt

werden können (vgl. hierzu SHAW, 1971; WITTE, 1976, 125-141). Man sollte also bei einem Handlungsforschungsprojekt nicht das vorhandene Wissen über diese Tatbestände einfach ignorieren, sondern das existierende theoretische Wissen als Alternative zur Kritik eigener theoretischer Vorstellungen verwenden.

Gehen wir nun auf eine Anzahl exemplarischer Hypothesen näher ein. Im Rahmen konkreter Projekte wurde immer wieder auf die negative Wirkung unklarer Ziele für die Qualität der Projektarbeit hingewiesen (vgl. z.B. MOSER, 1975, 152 und 156-158). Die Zielklarheit hat sich in Untersuchungen zum Problemlösungsverhalten als wichtige Variable herausgestellt. Es existieren hierzu u.a. folgende Hypothesen (vgl. SHAW, 1971, 325-331):

- H₁ Je größer die Zielklarheit bei den Mitgliedern einer Gruppe, desto positiver sind sie für Problemlösungsverhalten motiviert.
- H₂ Je größer die Zielklarheit bei den Mitgliedern einer Gruppe, desto besser lösen sie die perzipierten Probleme.

Aus H₁ und H₂ ergibt sich, daß eine sehr wichtige Aufgabe der Mitglieder eines Projekts darin besteht, eine große Zielklarheit zu erreichen, weil anderenfalls die Ergebnisse der Gruppenarbeit schlecht sein werden bzw. die Gruppe wegen abnehmender Motivation sogar auseinanderfallen kann (vgl. hierzu auch WEISS, 1974, 49-53). Abgesehen von dem hier nicht behandelten Meßproblem der genannten Variablen ist diese Forderung leichter gesagt als getan. Wie soll man Zielklarheit erreichen? Hierzu benötigen wir eine Anschlußtheorie, die uns Mittel angibt, Zielklarheit zu erreichen. Ein Mittel wäre z.B. die Aufstellung und Anwendung eines umfassenden Projektplans für das jeweilige Aktionsforschungsprojekt, wie ihn MOSER (1975, 159-160) und WELLENREUTHER (1976, 352-355) vorschlagen, sowie dessen Verteilung an alle Projektmitglieder, um eine Diskussion und Entscheidung über die Planungsziele und Planungsschritte eines Projekts zu erreichen. Dies muß in bestimmten Abständen wiederholt werden, da davon ausgegangen wird, daß sich die Ziele verändern können.

Ein wichtiger Aspekt der sog. Aufgabenstruktur innerhalb einer Gruppe ist z.B. der Grad der Schwierigkeit der Aufgaben. In vielen Projekten zeigt sich, daß man sich zu schwierige Aufgaben stellt. Formulieren wir auch hierzu einige Hypothesen (vgl. SHAW, 1971):

- H₃ Je größer die Aufgabenschwierigkeit, desto geringer die Gruppenleistung.

Aus H₃ ergibt sich u.a., daß man bei sehr schwierigen Aufgaben bzw. Problemen extrem viel Zeit braucht. Die hat man aber vielleicht zur Lösung eines aktuellen Problems nicht.

Nach der modifizierten Fassung der Wert-Erwartungstheorie von AMMERMANN (1976), in der aber noch weitere Bedingungen genannt werden, müßte diese Hypothese allerdings so modifiziert werden: (vgl. AMMERMANN, 1976, 144-153).

- H_{3'} Die Beziehung zwischen Aufgabenschwierigkeit und Gruppenleistung ist kurvilinear.

Damit meinen wir folgenden Tatbestand:

Bei Aufgaben mittlerer Schwierigkeit ist die Gruppenleistung hoch, bei Aufgaben sehr geringer und sehr hoher Schwierigkeit ist hingegen die Gruppenleistung niedrig. Demnach müßten sich Gruppen möglichst Aufgaben mittlerer Schwierigkeit auswählen.

Auch wenn wir jeweils im konkreten Einzelfall noch festlegen müssen, wie Zielklarheit bzw. Aufgabenschwierigkeit festgestellt bzw. gemessen wird, so erhält man aus solchen Hypothesen zumindest wichtige Hinweise, wenn auch keine fertige Sozialtechnologie für die Organisation von Projektgruppen.

Schließlich wollen wir noch darauf hinweisen, daß man aufgrund empirischer Ergebnisse auch normative Forderungen in Bezug auf ihre Realisierbarkeit untersuchen und kritisieren kann. Nehmen wir die Forderung nach Aufhebung der Subjekt-Objekt-Relation zwischen Forschern und Betroffenen. Wenn man darunter das Anstreben von Gleichberechtigung versteht, dann reicht eine solche normative Forderung nie aus, um dieses Ziel zu erreichen. Empirische Untersuchungen zeigen, daß sich selbst in kleinen Gruppen sehr schnell unterschiedliche Statuspositionen bzw. eine Schichtung ergibt (vgl. hierzu LEIK et al., 1975). Die Forderung nach Gleichberechtigung ist ethisch wertvoll, aber nicht praktikabel, da die realen Bedingungen hierfür erst geschaffen werden müssen; z.B. Angleichung der Kompetenz, Angleichung von Spezialkenntnissen etc.. Dabei sei nochmals angemerkt, daß hierfür wesentlich mehr Hypothesen zur Verfügung stehen, als wir in unserer exemplarischen Auswahl berücksichtigt haben. Es genügt also nicht, irgendwelche Postulate aufzustellen, sondern es kommt darauf an:

- a) zu prüfen, ob Theorien vorhanden sind, aus denen Maßnahmen zur Erreichung dieser Postulate abgeleitet werden können z.B. auch zur Erreichung von Gleichberechtigung,
- b) zu untersuchen, ob die Postulate in Handlungsforschungsprojekten erreicht wurden.

3.2. Hypothesen zur Organisation der Beziehungen zwischen Gruppen (sozialen Netzwerken)

In allen Handlungsforschungsprojekten, besonders aber in solchen mit einer umfassenden Fragestellung, wie z.B. dem Kollegstufenversuch-Nordrhein-Westfalen, stellt sich das Problem, wie die Kooperation der verschiedenen beteiligten Gruppen organisiert werden soll. Da es bei solchen Projekten um die Entwicklung und Verbreitung von Innovationen geht, ist hier der von HAVELOCK als Organisationsvorschlag entwickelte Linkageansatz relevant. Es handelt sich dabei nicht um eine explizite Theorie, sondern um eine Beschreibung bzw. Sollvorstellung für die Organisation der Beziehungen zwischen Gruppen (vgl. AREGGER, 1976, 124-136; HAMEYER, 1975; HAVELOCK, 1976). Er hat daher auch für die Bildungsplanung und Curriculumentwicklung Relevanz, bei der es um die Zusammenarbeit von Wissenschaft, Schulpraxis und Politik geht.

Das Problem der Verbreitung von Neuerungen und der damit verbundene 'linkage Ansatz' ist aber auch für die Konzeption des theoretischen Pluralismus von Bedeutung. Vom Standpunkt des theoretischen Pluralismus ist

das Problem der Verbreitung von Neuerungen deswegen wichtig, weil möglichst schnell und umfassend viele Personen über mögliche Alternativen zu bestehenden Problemlösungen unterrichtet werden sollten.

Man kann den Linkageansatz durch folgende Elemente charakterisieren (vgl. HAVELOCK, 1976, 11-16; HAMEYER, 1975, 284-287):

1. Bedürfnisidentifikation
2. Problemdiagnose
3. Suche nach Informationsquellen usw.
4. Informationsbeschaffung
5. Lösungsvorschläge
6. Realisierung des Lösungsvorschlags.

Diese Problemlösungsschritte finden nun als Prozesse zwischen den beteiligten Gruppen statt. Es soll also Kontakt und Wechselwirkung zwischen den Gruppen existieren. Im einzelnen soll dies so ablaufen:

1. Transfer neuen Wissens, neuer Fertigkeiten und Produkte vom resource-System (z.B. Wissenschaft oder Planungsgruppe) auf das user-System (z.B. Schule).
2. Nutzbarmachung von neuen Problemlösungen durch das user-System über lokale Innovations- und Problemlösungsbemühungen.
3. Artikulation und Rückmeldung von Bedürfnissen des user-Systems an das resource-System.
4. Entwicklung neuen Wissens, neuer Fertigkeiten und Produkte durch das resource-System.

Wenn erklärt werden soll, wie sich unterschiedliche soziale Netzwerke und Beziehungen zwischen Gruppen z.B. auf die Diffusion von Neuerungen (also Punkt 1) auswirken, benötigt man Theorien*. So gesehen sind HAVELOCK's Vorschläge und der Linkage-Ansatz noch Beschreibung einer Sollvorstellung ohne Theorie, zumindest ohne explizite Theorie.

Zur Erklärung der Verbreitung von Neuerungen können wir zwei Hypothesen von J.A. DAVIS (1968) verwenden, die er aus der kognitiven Gleichgewichtstheorie ableitet (zur Kritik vgl. aber KÖRNER, 1976).

^H4 Innovationen tendieren dazu, schnell innerhalb von Cliques zu diffundieren, aber langsam zwischen ihnen.

Die balancetheoretische Erklärung für diese Hypothese ist folgendermaßen: Weil Cliques dadurch definiert sind, daß die Mitglieder starke positive Bindungen haben, wird die Innovation nach Übernahme durch ein Mitglied schnell von den anderen Mitgliedern der Clique akzeptiert. Weil jedoch Mitglieder anderer Subgruppen und isolierte Mitglieder weniger positive oder sogar negative Bindungen zu den Mitgliedern der innovierenden Clique haben, werden sie sich den Neuerungen widersetzen oder diese sehr viel später übernehmen.

* Die Theorie sozialer Netzwerke ist noch sehr wenig entwickelt. Eine Übersicht geben WHITE et al., 1976.

H₅ Im Gegensatz zu einfachen Gruppen tendiert die Annahme einer Innovation in dichotomisierten oder polarisierten Gruppen dazu, schnell zu beginnen, aber sich dann zu verlangsamen oder sogar zu stoppen, bevor die Annahme universell ist.

In dichotomisierten Gruppen wird die erstmalige Annahme in einer der beiden Cliques erfolgen. Wegen der stark positiven Bindungen wird die Annahme schnell innerhalb dieser Subgruppe erfolgen. Die Mitglieder der anderen Subgruppe werden jedoch dazu tendieren, die Annahme wegen ihrer negativen Bindungen zu der annehmenden Gruppe abzulehnen.

H₆ Im Gegensatz zu einfachen Gruppen beginnt die Annahme einer Innovation innerhalb sich überschneidender Gruppen langsam, aber sie nimmt nach und nach zu und wird dann vollständig.

Die balancetheoretische Erklärung besteht darin, daß hier die am Anfang übernehmende Clique klein sein wird und ihre mittlere positive Beziehung nicht so groß wie bei der polaren Gruppe ist. Da es jedoch hier weniger Cliques gibt, die sich gegenseitig ablehnen, wird der Prozeß der Diffusion stetig vorwärts gehen und durch einen Schneeballprozeß die meisten der Gruppe erfassen.

Aus diesen Hypothesen ergibt sich, daß für die vollständige Diffusion sich überschneidende Gruppen von Wichtigkeit sind. Dies kann man dadurch erreichen, daß einige Wissenschaftler bei den Praktikern ("user-System") und einige Praktiker bei den Wissenschaftlern in Gruppen mitarbeiten. Diese sind dann Personen mit multipler Gruppenzugehörigkeit. Durch diese Personen erreicht man also, daß Informationen und Neuerungen zwischen Gruppen ausgetauscht und bei letzteren die Wahrscheinlichkeit der Annahme erhöht wird.

Diese Hypothese liefert Argumente für den Aufbau von Planungsgruppen mit Mitgliedern, die multiple Mitgliedschaften haben. Es könnte sich z.B. um Lehrer handeln, die gleichzeitig an der Schule unterrichten und mit Wissenschaftlern in einer Planungsgruppe zusammenarbeiten.

Hiermit wollen wir unsere exemplarische Diskussion der Anwendung von Hypothesen zur Organisation der Beziehungen zwischen Gruppen abschließen. Im folgenden wollen wir nun eine Methode näher behandeln, mit der man empirisch prüfen kann, ob man mit bestimmten Maßnahmen bestimmte Organisationsprobleme von Aktionsforschungsprojekten bzw. bestimmte soziale Probleme gelöst hat. Im Mittelpunkt der nächsten Teile steht allerdings die Anwendung dieser Methode zur Evaluation von Maßnahmen zur Lösung sozialer Probleme.

4. Quasiexperimentelle Versuchspläne als eine Methode experimenteller Handlungs- und Forschungsstrategie zur Lösung sozialer Probleme

In diesem Teil werden zunächst die üblicherweise angeführten zwei wichtigen Kriterien für die Qualität von Methoden und Versuchsplänen, nämlich interne und externe Gültigkeit, einer kritischen Analyse unterzogen. Dabei

bezieht sich die Anwendung dieser Kriterien zum einen auf den Vergleich der Qualität verschiedener Methoden wie Experiment, Quasiexperiment und Feldforschung. Zum anderen vergleicht man auch innerhalb jeder der genannten Methoden die Qualität unterschiedlicher Versuchspläne, so z.B. einer Zeitreihenuntersuchung mit und ohne Kontrollgruppe. Unter Verwendung von Argumenten von GADENNE (1976) wird dann eine alternative Methodologie zum Vergleich der Methoden und Versuchspläne vorgeschlagen, die der fallibilistischen Konzeption entspricht.

Danach gehen wir detaillierter auf die verschiedenen Versuchspläne ein und diskutieren sie anhand unseres alternativen Kriteriums *Strenge eines Tests*, d.h. des Tests, der mit einer bestimmten Versuchsanordnung möglich ist.

Im Anschluß daran formulieren wir ein einfaches Anwendungsbeispiel und führen einige weitere wichtige Kritikpunkte an der Arbeit von CAMPBELL, STANLEY (1970) im einzelnen auf. Unsere Modifikationsvorschläge entsprechen unserer deduktiv orientierten Methodologie und zielen darauf ab, folgendes zu integrieren:

1. Anwendung statistischer Verfahren der Kausalanalyse, die es erlauben, unterschiedlich präzise Hypothesen zu testen. Damit meinen wir Verfahren, die es erlauben, entweder nur ein Vorzeichen oder ein bestimmtes Intervall oder sogar einen Koeffizienten vorzugeben und unter dieser Nebenbedingung ein Gleichungssystem aufzulösen.
2. Im Rahmen dieser Modelle durch Zerlegung von Effekten potentiellen Nebenwirkungen auf die Spur zu kommen.
3. Durch Einbeziehung der verwendeten Beobachtungstheorie, d.h. der Meßhypothesen in den Test, diese nicht per Konvention als richtig zu unterstellen.
4. Durch Verwendung alternativer Methoden spezielle Mängel einzelner Methoden, sogenannte Methodeneffekte, festzustellen.

CAMPBELL (1969, 1971), CAMPBELL, STANLEY (1970) haben sehr eindringlich darauf hingewiesen, daß zur Überprüfung der Wirkung von Reformmaßnahmen und Innovationen eine Abwandlung des klassischen Experimentes methodisch notwendig sei. Diese Methode - quasiexperimentelle Versuchsplanung genannt - unterscheidet sich sowohl vom klassischen Experiment als auch von einer normalen Feldstudie.

Die drei Methoden unterscheiden sich in dem Ausmaß, in dem Faktoren kontrolliert werden können, die die interne und externe Gültigkeit beeinflussen (CAMPBELL, STANLEY, 1970). Was ist nun unter diesen zwei Begriffen - interne und externe Gültigkeit - zu verstehen? Unter *interner Gültigkeit* oder *interner Validität* wird das Ausmaß der Sicherheit, mit der man die Veränderung der abhängigen Variable wirklich der Veränderung der unabhängigen Variable kausal zurechnen kann, verstanden. Eine hohe interne Validität besagt dann, daß man bei dem betreffenden Versuchsplan recht sicher sein kann, daß nur die getroffenen Maßnahmen, d.h. bestimmte Reformen bzw. Innovationen die beobachteten Veränderungen herbeigeführt haben und diese nicht durch andere Einfluß- bzw. Störgrößen hervorgerufen wurden.

Unter *externer Gültigkeit* oder *externer Validität* wird die Sicherheit verstanden, mit der man die gefundenen Ergebnisse generalisieren kann. Diese Generalisierbarkeit ist gegeben wenn die verwendeten Stichproben von Personen, situativen Bedingungen und Variablen für die Populationen repräsentativ sind, auf die generalisiert werden soll.

Kommen wir nun zum Vergleich des Quasiexperiments mit den anderen Methoden:

- a) Quasiexperimenteller Versuchsplan - Klassisches Experiment
 Während im Idealfall bei beiden eine Zufallsauswahl von Personen aus einer definierten Grundgesamtheit gefordert wird, unterscheiden sie sich dadurch, daß beim klassischen Experiment darüber hinaus gefordert wird, daß Personen zufällig der Versuchs- und Kontrollgruppe zugeordnet werden, während dies für Quasiexperimente nicht notwendig ist (vgl. RIECHEN, BORUCH, 1974, 31-32). Damit sind nach der üblichen Auffassung die Kontrollmöglichkeiten beim Quasiexperiment zwar schlechter als im Experiment, aber besser als in einer Feldstudie.
- b) Quasiexperimenteller Versuchsplan - Feldstudie
 Bei einer klassischen Feldstudie wird im Gegensatz zum Quasiexperiment die unabhängige Variable nicht direkt vom Forscher verändert, sondern man mißt und beobachtet Variablen in natürlichen Situationen. Wenn man inferenzstatistische Schlüsse auf die Grundgesamtheit ziehen will, benötigt man auch hier Zufallsstichproben von Personen.

Im folgenden wollen wir die drei Verfahren entsprechend unseren Kriterien übersichtlich in Form eines Schemas darstellen:

Schema 2: Vergleich der Methoden

	Direkte Veränderung der Variablen	Zufallszuordnung zu Versuchs- und Kontrollgruppe (Randomisierung)	Zufallsauswahl aus einer definierten Grundgesamtheit
Experiment	X	X	X
Quasiexperiment	X		X
Feldstudie			X

Ein X in der Tabelle bedeutet, daß das entsprechende Merkmal vorhanden ist.

Allerdings muß angemerkt werden, daß das dritte Merkmal, die Zufallsauswahl aus einer Grundgesamtheit, bei keiner der drei Methoden zwingend ist und auch in der empirischen Praxis - speziell bei Experimenten - selten vorkommt. Dies betrifft aber nur die Anwendbarkeit von Signifikanztests, um Schlüsse von Stichproben auf definierte Grundgesamtheiten durchzuführen und ist ansonsten für den Vergleich der Methoden nicht relevant.

Man kann nun die drei Methoden in bezug auf die Kriterien der externen und internen Validität vergleichen, da es als Ziel einer empirischen Studie gilt, eine hohe interne und externe Validität zu erreichen. Je mehr man sich jedoch bemüht, eines der beiden Ideale zu erreichen, desto mehr muß man das andere vernachlässigen (CAMPBELL, STANLEY, 1970, GADENNE, 1976, 23). Dies liegt darin begründet, daß die interne Validität eine möglichst vollständige Kontrolle erfordert, welche man nur erreicht, wenn man eine Untersuchung in einer isolierten Situation durchführt*. Diese wird

* Der Begriff interne Gültigkeit hat hier eine andere Bedeutung als im üblichen testtheoretischen Sinn.

aber nur durch ein Experiment gegeben. Hierdurch wird allerdings die Repräsentativität, die "Lebensnähe", beeinträchtigt. Die Feldstudie zeichnet sich zwar durch Repräsentativität aus, habe aber den Mangel fehlender Kontrolle von Störvariablen*. Da aber Untersuchungen, die nicht intern valide sind, wie z.B. Feldstudien, auch für Generalisierungen nicht brauchbar sind, werden die drei Methoden in bezug auf ihren Wert folgendermaßen eingeordnet (CAMPBELL, STANLEY, 1970 und 1975; kritisch hierzu GADENNE, 1976, 24; OPP, 1973):

Wenn möglich, soll man Experimente verwenden. Wenn dies nicht möglich ist, Quasiexperimente und nur wenn auch dies ausgeschlossen ist, sollte man Feldstudien durchführen.

GADENNE hat nun die Theorie der internen und externen Gültigkeit, auf der diese Einordnung der Verfahren beruht, einer sehr gründlichen Kritik unterzogen (GADENNE, 1976). Seine Hauptthese lautet, daß die Theorie der internen und externen Validität nicht eine adäquate Theorie psychologischer Methoden darstellt. Sie ist, wie er detailliert ausführt, den Möglichkeiten empirischer Forschung unangemessen und steht mit Ergebnissen der Wissenschaftslehre im Widerspruch. Interne und externe Gültigkeit kann man nur erreichen, wenn man davon ausgeht, daß man entweder eine vollständige Kontrolle der Störfaktoren erreicht oder die Möglichkeit induktiver Schlüsse zuläßt. Das erstere ist aber deswegen nicht möglich, weil eine störende Bedingung nicht in einem äußeren Ereignis oder in unterschiedlichen Eigenschaften von Experimental- und Kontrollgruppe bestehen muß, sondern bereits in den Manipulationen, die der Versuchsleiter vornimmt, enthalten sein kann. Eine Übersicht über die entsprechenden Ergebnisse (zu Versuchsleitereffekten u.v.a.m.) findet man bei ROSENTHAL (1966) und MERTENS (1975). Wir wollen in diesem Zusammenhang darauf nicht näher eingehen. Induktive Schlüsse wird man hingegen ebenfalls nicht zulassen wollen, da Schlüsse dieser Art nicht haltbar sind (vgl. z.B. VETTER, 1967).

Da man niemals davon ausgehen kann, alle Gesetze zu kennen, die eine "störende Bedingung" auslösen, ist eine Entscheidung über die Frage, ob in einem bestimmten Untersuchungsfall Störfaktoren kontrolliert wurden, nur im negativen Sinn möglich: Es kann nachgewiesen werden, daß ein unkontrollierter Störfaktor vorhanden war, nicht aber, daß alle Störfaktoren kontrolliert wurden (vgl. GADENNE, 1976, 33).

Im Rahmen einer deduktivistischen Theorie empirischer Untersuchungen zeigt nun GADENNE (1976, 49 ff.), daß die Forderungen von Kontrolle und Repräsentativität auch deduktivistisch begründet werden können. Aus dem Falsifikationsprinzip folgt die Forderung nach strengen Prüfungen von Theorien. Wie kann man nun eine Verbindung zwischen der Kontrolle störender Bedingungen und der Strenge der Prüfung einer Theorie herstellen? GADENNE (1976, 64) schlägt hierfür folgende Regel, die er als Prinzip der Kontrolle bezeichnet, vor. Diese lautet:

Prinzip der Kontrolle (PK):

Gegeben seien zwei Untersuchungen U_1 und U_2 zur Prüfung einer Theorie T (A, B). Diese Theorie bestehe aus einer einfachen Hypothese folgender Art: Wenn A, dann B. In beiden Untersuchungen wurde A hergestellt bzw. beobachtet. U_1 ist in diesem

* Zu einer inhaltlichen Diskussion der Störvariablen vergleiche man Schema 4 und die Diskussion der Bedeutung einer Art von Störvariable im Zusammenhang mit der Evaluation von Programmen zur kompensatorischen Erziehung.

Fall genau dann eine strengere Prüfung als U_2 , wenn die beiden folgenden Bedingungen erfüllt sind:

1. In U_1 wurden alle die potentiell störenden Bedingungen erster und zweiter Art kontrolliert, die in U_2 kontrolliert wurden.
2. Es gibt mindestens eine potentielle störende Bedingung erster oder zweiter Art, die in U_1 , nicht jedoch in U_2 kontrolliert wurde.

Die Bedeutung der störenden Bedingungen erster und zweiter Art wird von GADENNE so definiert: Ein Ereignis F ist eine störende Bedingung erster Art, wenn F das Ereignis B auszulösen und somit einen Effekt des Ereignisses A vorzutäuschen vermag.

Ein Ereignis F' ist eine störende Bedingung zweiter Art, wenn A nur zusammen mit F' hinreichend für das Auftreten von B ist, d.h. weder A noch F' vermögen allein B zu verursachen, beide zusammen sind jedoch hinreichend für B. Die störenden Bedingungen der ersten Art beziehen sich auf die interne Validität, während die störenden Bedingungen der zweiten Art die externe Validität betreffen.

Als wichtige Konsequenz dieser methodologischen Regel (des Prinzips der Kontrolle) wird die Trennung zwischen einem internen Problem der Kausalinterpretation (interne Validität) und einem externen Problem der Generalisierung (externe Validität) aufgehoben. Sowohl störende Bedingungen erster Art als auch solche zweiter Art haben mit der Strenge von Prüfverfahren zu tun, und die eben genannte methodologische Regel ermöglicht nach dem Vorschlag GADENNEs die vergleichende Bewertung verschiedener Untersuchungen, sofern diese sich auf die gleiche Fragestellung beziehen. Daraus ergibt sich eine Folgerung, die im Gegensatz zu den Thesen von CAMPBELL über den Wert der drei angeführten Methoden steht: Nicht die Wahl einer bestimmten Methode ist entscheidend für die Güte einer Untersuchung, sondern die Strenge eines Tests im Rahmen einer gegebenen Untersuchung. Von daher gesehen ist die Anwendung quasiexperimenteller Verfahren kein Notbehelf gegenüber Experimenten, weil allein die Strenge der Prüfung der jeweils verwendeten Hypothesen entscheidend ist (vgl. hierzu auch CAIN, 1975, 304-305).

Zusammenfassend ergibt sich somit:

Nach der Falsifikationstheorie erfordert ein Erkenntnisfortschritt u. a. die Durchführung möglichst strenger Prüfungen. Dies bedeutet, daß Untersuchungen so angelegt werden müssen, daß die Chance der Falsifikation einer falschen Theorie möglichst groß ist. Hierzu ist die Kontrolle potentieller störender Bedingungen erster und zweiter Art notwendig (GADENNE, 1976, 66). Kompliziertere Versuchspläne, die viele der potentiell störenden Bedingungen erster und zweiter Art kontrollieren, sind somit kein überflüssiger methodischer Luxus, sondern geben uns eine erhöhte Chance festzustellen, ob ein Programm wirklich wirksam ist und somit zu 'lernen'.

Warum empfiehlt sich von den drei Verfahren gerade die Methode des Quasiexperiments für Handlungsforschungsprojekte? Es sind zwei Gründe:

Da man bei Handlungsforschungsprojekten und Sozialtechnologien direkt etwas verändern will und dann prüfen will, ob die durchgeführte Änderung die gewünschten Ziele erreicht hat, kommt keine Feldstudie in Frage, in der man nur beobachtet und mißt.

Da in den meisten Fällen in realen Situationen keine zufällige Zuordnung von Personen auf Versuchs- und Kontrollgruppen möglich ist, kommt auch meistens kein klassisches Experiment in Frage.

Als weitere Frage ergibt sich, welche Vorteile diese Methode insgesamt für Handlungsforschungsprojekte hat. Hier scheinen uns folgende Argumente relevant:

1. Man hat die Möglichkeit, neben den subjektiven Eindrücken im Rahmen von Handlungsforschungsprojekten auch eine objektive empirische Kontrolle der Wirkung von Maßnahmen durchzuführen und kann, wie wir später vorschlagen, diese beiden miteinander verbinden. Von CAMPBELL, STANLEY selbst stammt der Hinweis, daß hier keineswegs ein Ausschließlichkeitsanspruch beabsichtigt ist. Sie verweisen darauf, daß die Ergebnisse einer solchen quasiesperimentellen Versuchsordnung nur kombiniert mit den Beobachtungen und Eindrücken der jeweils Betroffenen sinnvoll interpretiert werden können (vgl. CAMPBELL, STANLEY, 1970, 500).
2. Wenn man vor der Durchführung der Maßnahmen explizit das zu Grunde liegende theoretische Wissen formuliert, d.h. den deduktiven Ansatz verwendet, kann man ohne Schwierigkeiten die Maßnahmen auch auf andere Bereiche anwenden, sofern sie sich nicht bereits im Rahmen des Projekts als falsch erweisen. Dabei spielt es keine Rolle, ob es sich bei den angewandten Theorien um Alltagswissen oder empirisch geprüfte Theorien handelt.

Als Bedingung für die Anwendung von quasiexperimentellen Versuchsplänen im Sinne der Aktionsforschung führt MOSER (1975, 128) an, daß der ganze quasiexperimentelle Versuchsplan von allen Beteiligten ausgearbeitet werden soll. Dies fordern aber auch bereits CAMPBELL, STANLEY (1970).

Im folgenden gehen wir nun auf die verschiedenen möglichen quasiexperimentellen Versuchspläne näher ein.

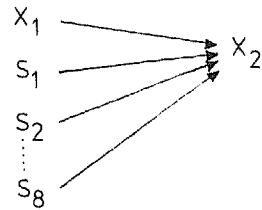
4.1. Überblick über quasiexperimentelle Versuchsordnungen

Von CAMPBELL, STANLEY (1970) werden 14 verschiedene quasiexperimentelle Versuchspläne vorgestellt. Diese unterscheiden sich in dem Ausmaß, in dem Faktoren kontrolliert werden können, die die interne und externe Gültigkeit des Quasiexperiments beeinflussen. Aus der Diskussion in unserem vorigen Abschnitt ergibt sich allerdings, daß sie sich nicht in dem Ausmaß der internen Gültigkeit, sondern in der Strenge der jeweils untersuchten Theorie bzw. des untersuchten Programms unterscheiden.

Der Vergleich der verschiedenen Versuchspläne kann mit Hilfe des vorher explizierten Prinzips der Kontrolle durchgeführt werden. Dieses Prinzip stellte eine Beziehung zwischen dem Ausmaß an Kontrolle störender Bedingungen und der Strenge eines Tests her. Von den beiden Autoren werden acht solcher störenden Bedingungen erster Art herausgearbeitet, die nach ihrer Interpretation die interne Gültigkeit betreffen. Dies sind alternative Hypothesen zur Erklärung der jeweiligen empirischen Ergebnisse. Eine Kontrolle dieser alternativen Hypothesen erlaubt einen strengeren Test einer Theorie bzw. eines Programms als eine Nichtkontrolle dieser Störvariablen erster Art.

Das eben genannte Problem der Kontrolle von Störvariablen sowie sein Zusammenhang mit der Strenge eines Tests läßt sich durch folgendes Pfaddiagramm anschaulich darstellen:

Abb. 1: Störvariablen 1. Art



wobei: X_1 = veränderte unabhängige Variable
(die Maßnahmen oder die Innovation)
 X_2 = abhängige Variable (Zielvariable,
z. B.: Veränderung in einem best.
sozialen System in einer bestimm-
ten Richtung)
 $S_1 - S_8$ = Störvariablen 1. Art

Wenn nun $S_1 - S_8$ Variablen sind, die ebenfalls X_2 erklären können, dann gilt folgendes:

Ein quasiexperimenteller Versuchsplan prüft umso strenger eine durchgeführte Maßnahme (X_1), je mehr Störvariablen kontrolliert werden. Eine solche Kontrolle kann dann auch dazu führen, daß man herausfindet, daß eine durchgeführte Maßnahme überhaupt keine Wirkung hatte. Wenn man hingegen die Störvariablen nicht kontrolliert, würde die Möglichkeit nicht bestehen, dies herauszubekommen und dadurch zu lernen, daß eine getroffene Maßnahme ein Fehlschlag war. Eine Kontrolle der Störvariablen kann aber auch dazu führen, daß man feststellt, daß die Wirkung einer Maßnahme (X_1) unter bzw. überschätzt wurde. Diese Konsequenz, technisch als Spezifikationsfehler durch weggelassene Variablen bezeichnet, ergibt sich dann, wenn die Maßnahme mit den Störvariablen korreliert ist (vgl. hierzu CAIN, 1975, 300-303).

Eine produktive Konfrontation von Handlungsforschung und der Methodologie quasiexperimenteller Versuchspläne müßte nun die Kritik der Handlungsforscher an der empirischen Sozialforschung dahingehend aufarbeiten und explizieren, ob sie weitere Störvariablen enthält, die bei CAMPBELL, STANLEY noch nicht enthalten sind. Dies wäre ganz im Sinne der Argumentation von GADENNE, daß wir nie alle Störvariablen kennen können und daher immer nach alternativen Theorien suchen müssen, um auf neue, unberücksichtigte Störvariablen zu kommen. Bisher haben wir allerdings in den Schriften der Handlungsforschung keine zusätzlichen Störvariablen gefunden, die in der Liste von CAMPBELL, STANLEY noch nicht enthalten sind.

Für die externe Gültigkeit werden vier Störgrößen von den Autoren aufgeführt; die verschiedenen Versuchspläne werden wiederum daraufhin verglichen, inwieweit diese Störgrößen jeweils zu kontrollieren sind. Für diese Störvariablen $S_9 - S_{12}$ gilt das Gleiche wie für $S_1 - S_8$.

Unter Anwendung des vorher formulierten Prinzips der Kontrolle gilt allgemein: Ein Versuchsplan 1 ist genau dann eine strengere Prüfung als Versuchsplan 2, wenn die beiden folgenden Bedingungen erfüllt sind:

1. In U_1 wurden alle die potentiell störenden Bedingungen 1. und 2. Art kontrolliert, die in U_2 kontrolliert wurden (z. B. $S_{11} - S_{10}$)
2. Es gibt mindestens eine potentiell störende Bedingung 1. oder 2. Art, die in U_1 , nicht jedoch in U_2 kontrolliert wurde (z. B. S_{11} und S_{12}).

Es würde an dieser Stelle zu weit führen, näher auf alle 14 verschiedenen Versuchspläne sowie ihren Vergleich einzugehen; wir verweisen daher auf CAMPBELL, STANLEY (1970) sowie ZIMMERMANN (1972).

Anhand dreier quasiexperimenteller Versuchspläne, von denen wir einen auf eine inhaltliche Fragestellung anwenden, wollen wir nun exemplarisch die Brauchbarkeit, aber auch die Probleme quasiexperimenteller Versuchspläne demonstrieren.

Schema 3: Aufbau der drei quasiexperimentellen Versuchspläne

- a) Kontrollgruppenanordnung
ohne Randomisierung, d.h. ohne zufällige Zuordnung
der Individuen zu Versuchs- und Kontrollgruppe

Versuchsgruppe				X				
Kontrollgruppe								
								M_1
								M_2

- b) Zeitreihen

Versuchsgruppe				X				
								M_1
								M_2
								M_3
								M_4
								M_5
								M_6

- c) Mehrfache Zeitreihen

Versuchsgruppe				X				
Kontrollgruppe								
								M_1
								M_2
								M_3
								M_4
								M_5
								M_6

Hierbei bedeuten:

- $M_1 - M_6$ Messung der abhängigen Variablen zu den Zeitpunkt $t_1 - t_6$
- X Veränderung der unabhängigen Variable

Im Versuchsplan a) wird laut diesem Schema eine unabhängige Variable verändert, d.h. eine Innovation bei der Versuchsgruppe eingeführt, was durch das X symbolisiert wird. Daran anschließend werden bei der Versuchsgruppe und der Kontrollgruppe wiederum die abhängige Variable gemessen symbolisiert durch M_1 und M_2 .

Im Versuchsplan b) werden laut diesem Schema drei Messungen, M_1 , M_2 und M_3 , in der Versuchsgruppe zu drei verschiedenen Zeitpunkten gemacht. Dann wird eine Neuerung eingeführt bzw. eine Maßnahme durchgeführt (X). Danach werden wiederum in der Versuchsgruppe drei Messungen, M_4 , M_5 und M_6 , zu drei verschiedenen Zeitpunkten gemacht.

Der Versuchsplan c) entspricht Versuchsplan b), jedoch wird in c) zusätzlich eine Kontrollgruppe untersucht.

Da man wissen will, ob eine durchgeführte Innovation bzw. Maßnahme tatsächlich einen Effekt auf die Zielvariable X_2 hat, ist es wichtig abzuschließen, ob die Veränderung in X_2 tatsächlich auf X_1 (die Maßnahme) und nicht auf $S_1 \dots S_{11}$, d.h. die Störvariablen zurückzuführen ist. Um dies sicherzustellen, kann man entweder versuchen, die Variablen oder einen Teil der Variablen direkt zu messen und sie dann im Rahmen einer multi-

variatsanalyse zu kontrollieren* , oder bereits durch die Art der Versuchsanordnung diese Variablen konstant zu halten.

Um welche Variablen es sich bei den 11 Störvariablen handelt, ist dem folgenden Schema zu entnehmen (vgl. ZIMMERMANN 1972, 276 und CAMPBELL, STANLEY 1970).

In welchem Ausmaß die einzelnen Störvariablen kontrollierbar sind, stellen wir für die drei ausgewählten Versuchspläne nun im folgende Schema dar:

Schema 4: Kontrolle der Störvariablen in den Versuchsanordnungen a-c

		a) Kontrollgruppen- anordnung ohne Randomisierung	b) Einfache Zeitreihen	c) Mehrfache Zeitreihen
S ₁	Zeiteinflüsse	+	-	+
S ₂	biolog. -psych. Veränderungen	+	+ ?	+
S ₃	Meßeﬀekte	+	+	+
S ₄	Veränderung im Meßinstrument	+	?	+
S ₅	Regressionseﬀekte	?	+	+
S ₆	Auswahlverzerrung	+ ?	+	+
S ₇	Ausfälle	+	+	+
S ₈	Interaktion von S ₆ mit S ₂	-	+	+
S ₉	Interaktion von Meßeﬀekten und x	-	-	-
S ₁₀	Interaktion von Aus- wahlverzerrung und x	?	?	?
S ₁₁	Reaktive Eﬀekte exp. Behandlung	?	?	?

* Der Zusammenhang zwischen experimentellem und nichtexperimentellem Vorgehen und die Rolle multivariater Verfahren wird ausführlich bei A. D. MILLER (1971) behandelt. Diese Strategie der Messung und Kontrolle der Störvariablen setzt eine Theorie voraus, mit deren Hilfe solche Störvariablen identifiziert werden können.

Im Folgenden wollen wir nun dieses Schema näher erläutern:

- (+) in der Tabelle bedeutet, daß der Störfaktor in der entsprechenden Versuchsanordnung kontrolliert werden kann
- (-) in der Tabelle bedeutet, daß der Störfaktor in der entsprechenden Versuchsanordnung nicht kontrolliert werden kann
- (?) in der Tabelle bedeutet, daß die Schwierigkeiten die Störgröße zu kontrollieren, in dem betreffenden Versuchsplan bestehen bleiben.

CAMPBELL, STANLEY (1970) bzw. ZIMMERMANN (1972, 76-79) und MERTENS (1975, 24-29) definieren die Störgrößen folgendermaßen:

- (1) Unter Zeiteinflüssen werden diejenigen Ereignisse verstanden, die zwischen der ersten und der zweiten Messung zusätzlich zur experimentellen Variable auftreten und eine Wirkung auf die abhängige Variable haben.
- (2) Die zweite Störvariable, meistens auch Reifung genannt, umfaßt alle diejenigen biologischen und psychischen Vorgänge, die sich unabhängig von besonderen äußeren Ereignissen im Lauf der Zeit ändern. Zwischen zwei Messungen können Schüler z. B. älter, müder, gelangweilter geworden sein. Praktisch ist (2) ein Spezialfall von (1).
- (3) Mit Meßeﬀekten werden die Auswirkungen einer ersten Messung oder eines ersten Tests auf eine zweite Messung bzw. einen zweiten Test verstanden. Diese können z. B. durch eine zwischenzeitliche Diskussion zwischen Befragten entstehen.
- (4) Die Veränderungen in Meßinstrumenten beziehen sich auf autonome Veränderungen der Meßinstrumente, die durch Beobachtungsfehler, Codierungsfehler etc. verursacht sein können.
- (5) Als Regressionseﬀekt bezeichnet man den Tatbestand, daß Personen mit extremen Meßwerten bei einer Wiederholung der Messung mehr zum Mittelwert hinstreben.
Dieses Problem (5) trat z. B. bei der Evaluation amerikanischer Programme zur kompensatorischen Erziehung auf. Zunächst suchte man aufgrund von Testergebnissen die extrem guten und schlechten Schüler aus. Nach der Durchführung des Programms waren nun die guten Schüler schlechter und die schlechten Schüler besser geworden. Während das letztere Ergebnis akzeptabel ist, wird man hingegen das Absinken der Leistungen der guten Schüler als eine negative Nebenwirkung des Programms ansehen. Man kann dieses Ergebnis aber gar nicht sinnvoll inhaltlich interpretieren, da es durch ein methodisches Artefakt - den Regressionseﬀekt - zustande gekommen ist. Wie oben angeführt bewirkt er, daß bei Extremgruppenauswahl immer die schlechten Schüler besser und die guten schlechter werden. Entsprechende Ergebnisse können also durch die Anwendung dieser Methode zustande kommen und nicht durch Anwendung eines Programms für kompensatorische Erziehung (vgl. CAMPBELL, ERLEBACHER, 1970).
- (6) Unter Auswahlverzerrung oder Selektionseﬀekt versteht man die unterschiedliche Auswahl von Personen für Versuch. und Kontrollgruppe. Sie unterscheiden sich dann in ihren Meßwerten bereits vor Ausführung eines Experiments. So nehmen z. B. an einem Schulversuch vielleicht gerade besonders motivierte Lehrer teil, die sich bezüglich ihrer Motivation von Lehrern der normalen Schulen stark unterscheiden.
- (7) Unter Ausfällen versteht man die unterschiedliche Ausfallquote von Personen bei Versuchs- und Kontrollgruppe.
- (8) Wechselwirkung zwischen Auswahl und Reifung kommt z. B. vor, wenn Reifung vor allem in der Versuchsgruppe, aber nicht in der Kontrollgruppe auftritt.
- (9) Wechselwirkungen zwischen dem Testen und experimenteller Einwirkung können auftreten, wenn spezifische Testverfahren die Aufmerksamkeit von befragten Personen gegenüber bestimmten Merkmalen hervorrufen, was sich dann auch auf die abhängige Variable auswirken kann.

- (10) Wechselwirkung zwischen verzerrter Auswahl und experimenteller Einwirkung kann auftreten, wenn bei bestimmten Experimenten sich immer eine bestimmte Art von Personen meldet.
- (11) Reaktive Effekte der experimentellen Behandlung sind Verhaltensänderungen der Betroffenen als Folge der experimentellen Anordnung.
Die 12. Störgröße ist für die drei genannten Versuchspläne nicht relevant.
Daher haben wir sie weggelassen.

Vergleichen wir nun die drei Versuchspläne:

Im Versuchsplan a) werden folgende Störvariablen 1. Art nicht kontrolliert bzw. es bleibt unklar, ob sie kontrolliert werden können:

1. Interaktion von Auswahlverzerrungen und biologisch-psychologischen Veränderungen
2. Regressionseffekte
3. Auswahlverzerrungen

Im Versuchsplan b) den einfachen Zeitreihen, ergaben sich folgende nicht kontrollierte Faktoren:

1. Zeiteinflüsse
2. Biologisch-psychische Veränderungen
3. Veränderungen in den Meßinstrumenten

Im Versuchsplan c) hingegen kann man alle hier aufgeführten Störvariablen durch die Versuchsanordnung kontrollieren. Da sich die drei Versuchspläne hinsichtlich der Kontrolle der Störvariablen 2. Art nicht unterscheiden, ist Versuchsplan c) - nach dem vorher formulierten Prinzip der Kontrolle - der beste. Zwischen Versuchsplan a) und b) kann man hingegen keinen Qualitätsunterschied erkennen. Dies gilt allerdings nur dann, wenn man annimmt, daß sich sonst die drei Versuchspläne nicht unterscheiden.

Welchen Schluß können wir daraus ziehen? Sicher nicht den, in allen Anwendungssituationen Versuchsplan anzuwenden. Denn neben der Qualität der Versuchspläne muß man ja auch berücksichtigen, ob man überhaupt eine Kontrollgruppe bilden kann und ob aufwendige Messungen zu verschiedenen Zeitpunkten durchführbar sind. Hat man hierzu keine Möglichkeit, muß man die schwächeren Versuchspläne verwenden, die dann natürlich auch nur einen weniger strengen Test erlauben.

Bei der Anwendung solcher Versuchspläne sieht MOSER (1976, 128) die folgenden Probleme:

1. Wieweit werden die Ergebnisse dadurch verändert, daß die Untersuchten nicht mehr die Unwissenden sind?
2. Wieweit treffen die methodologischen Prämissen der empirischen Verfahren noch zu?
3. Ergeben sich dadurch Rollenprobleme, daß der Forscher gleichzeitig Forscher und Mitarbeiter (Freund und Feind) ist?

Auf die Beantwortung dieser Fragen geht MOSER nicht ein. Dies wollen wir im folgenden tun.

Das erste Problem bezog sich darauf, daß die Untersuchten nicht mehr die Unwissenden sind. Dies entspricht unserer Störvariablen S_{11} - den

reaktiven Effekten - in Schema 4. Nun ist dies kein spezielles Problem von Handlungsforschungsprojekten. Auch dann, wenn Personen nicht informiert sind, haben sie bestimmte Hypothesen darüber, was getestet werden soll und wie sie sich verhalten sollen (vgl. z.B. MERTENS, 1975, 88-144). Man muß daher das erste von MOSER genannte Problem als ein Problem bezeichnen, das auch in klassischen Designs auftritt und nicht ein spezielles Problem der Anwendung quasiexperimenteller Versuchspläne in Handlungsforschungsprojekten darstellt.

Das zweite aufgeworfene Problem - inwieweit die methodologischen Prämissen der empirischen Verfahren noch zutreffen - ist ebenfalls nach unserer Diskussion der vielfältigen anderen Fehlerquellen mit ja zu beantworten. Es ist also kein spezielles Problem. Dies gilt allerdings nur, wenn wir den Begriff methodologische Prämissen, der bei MOSER nicht präzisiert wird, adäquat als Ausmaß der Kontrolle von Störvariablen interpretiert haben.

Das dritte aufgeworfene Problem bezog sich auf Rollenkonflikte, die dadurch entstehen, daß Forscher gleichzeitig Forscher und Mitarbeiter sind. Hierzu muß man gleichfalls anmerken, daß Rollenkonflikte auch in der normalen Forschung existieren. Zum Beispiel zeigen dies die Untersuchungen zu den Versuchsleitereffekten (vgl. ROSENTHAL, 1966 und MERTENS, 1975). Mithin ist auch dies kein spezielles Problem bei Handlungsforschungsprojekten.

Zusammenfassend ergibt sich, daß die drei genannten Probleme alle nicht spezifisch für Handlungsforschungsprojekte sind, sondern in anderen Arten von Forschung ebenso auftreten.

Um unsere Überlegungen nicht so kompliziert zu gestalten, gehen wir nun auf ein einfaches Beispiel ein. Jedoch lassen sich unsere Überlegungen ohne weiteres auf komplexere Anwendungsbeispiele übertragen.

4.2. Anwendungsbeispiel 1: Ein hochschuldidaktisches Quasiexperiment (vgl. hierzu PORTELE, 1975)

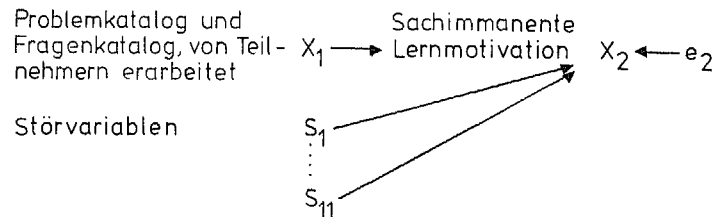
Im Rahmen eines hochschuldidaktischen Quasiexperimentes will man als Zielvariable die sachimmanente Lernmotivation (intrinsische Motivation) erhöhen. Sie ist also unser X_2 . Als ein Mittel, um dieses Ziel zu erreichen, sieht man die Maßnahme an, daß die Teilnehmer der Veranstaltungen selbst von Stunde zu Stunde Probleme, Übungsaufgaben, Fragen zur Kritik von Texten ausarbeiten (X_1 = Partizipation o.ä.). Es handelt sich hier um eine Nominalvariable mit der Ausprägung 0 = keine Erarbeitung von Fragen, Problemen etc. und 1 = Erarbeitung von Fragen, Problemen. Konkret soll dies jeweils von einer wechselnden kleinen Gruppe der Teilnehmer in Zusammenarbeit mit dem Dozenten durchgeführt werden. In diesem Zusammenhang wollen wir offen lassen, ob sich diese Idee per Ableitung aus einer empirisch geprüften Theorie ergab oder aus Alltagstheorien, die von den Teilnehmern der Kurse zusammen vor Beginn der Veranstaltungen erarbeitet wurden. In jedem Fall setzen wir voraus, daß die Hypothese (Partizipation erhöht intrinsische Motivation) wahr ist. Die Messung der sachimmanenten Motivation erfolgt durch ein standardisiertes Meßinstrument, welches aus semantischen Differentialen, also Gegensatzpaaren, besteht. Ein Beispiel dafür:

11 weitere Items dieser Art werden zur Messung des theoretischen Konstrukts "intrinsische Motivation" verwendet.

Unsere Fragestellung lautet: Wird die sachimmanente Lernmotivation durch die geplante Maßnahme erhöht?

Wie wir bereits zu Anfang betonten, kann jedoch auch diese Frage nicht allein durch eine noch so optimale quasiexperimentelle Versuchsanordnung beantwortet werden. Jedoch bildet sie neben den Beobachtungen der Teilnehmer an einem solchen Quasiexperiment eine Art Korrektiv gegenüber den Beobachtungen und Einschätzungen, die durch Gruppen aufgrund von Beobachtungen gemacht werden. In Form eines Pfeildiagramms kann man unsere Fragestellung auch folgendermaßen darstellen:

Abb. 2 Determinanten der sachimmanenten Lernmotivation



Das Residuum e symbolisiert den nichtdeterministischen Charakter unserer Hypothesen. Wir gehen also technisch gesprochen davon aus, daß nicht alle Varianz der Lernmotivation von X_1 bzw. $S_1 \dots S_{11}$ erklärt wird. Nun stellt sich das Problem der Auswahl unter den verschiedenen quasiexperimentellen Versuchsplänen. Wie wir bereits vorher ausführten, hängt aber die Auswahl nicht nur von den Möglichkeiten der Kontrolle der Störvariablen, sondern auch von pragmatischen Gesichtspunkten wie Verfügbarkeit von Kontrollgruppen und Ausmaß der zur Verfügung stehenden Zeit zur Messung der Variablen ab. Aus diesem Grund ist ohne Kenntnis der jeweiligen Bedingungen keine generelle Entscheidung für einen Versuchsplan möglich.

Wenn man nun die M 's in den drei Versuchsplänen durch X_2 und die X durch X_1 ersetzt, erhält man die entsprechenden drei Versuchspläne für unsere Fragestellung. Wir führen sie aber an dieser Stelle nicht nochmals auf.

Im nächsten Abschnitt wollen wir auf einige Kritikpunkte und Verbesserungsvorschläge an den quasiexperimentellen Versuchsplänen eingehen.

4.3. Kritik an der Interpretation und Analyse von quasiexperimentellen Versuchsplänen

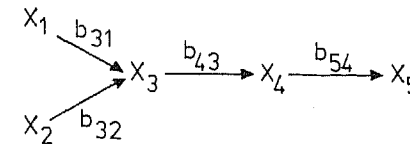
Im Folgenden wollen wir eine Reihe von Kritikpunkten diskutieren, die man an der von CAMPBELL und STANLEY vorgelegten Sammlung von Versuchsplänen, sowie insbesondere der Vorschläge zu deren Analyse machen kann. Deren Berücksichtigung bei der Planung von Quasiexperimenten scheint uns für die Anwendung dieser Methode in Handlungsforschungsprojekten aber auch generell wichtig zu sein.

a) fehlende Generalisierung auf multivariate Modelle

In den meisten realen Situationen (z. B. auch bei dem Projekt Kollegstufenreform) hat man es nicht mit bivariaten Beziehungen, also Beziehungen zwischen zwei Variablen, sondern mit multivariaten Modellen, mit mehreren unabhängigen und mehreren abhängigen Variablen, zu tun.

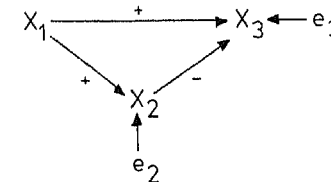
Es handelt sich dann um Modelle des folgenden Typs:

Abb. 3: Multivariates Modell



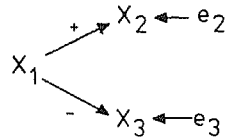
Die b 's seien Regressionskoeffizienten, die die Stärke der Beziehungen ausdrücken. Grundsätzlich kann man die Stärke und Signifikanz der Beziehungen solcher Modelle mit Regressionsanalysen berechnen (vgl. z. B. BREDEKAMP, 1975, OPP, SCHMIDT, 1976). In OPP, SCHMIDT findet man auch konkrete Beispiele, wie aus den verbalen Hypothesen die entsprechenden Strukturgleichungen formuliert werden, durch die die Abhängigkeiten, die in den verbalen Hypothesen enthalten sind, formalisiert werden. In multivariaten Modellen hat man darüber hinaus durch das Verfahren der Zerlegung von Effekten in direkte, indirekte Effekte die Möglichkeit, Nebenwirkungen von Maßnahmen bzw. Variablen festzustellen. Dazu sei folgendes einfache Beispiel gegeben:

Abb. 4: Dreivariablenmodell mit direkten und indirekten Wirkungen



In dem Modell in Abb. 4 wirke X_1 direkt positiv auf X_3 . Eine ausschließliche Berücksichtigung dieser direkten Wirkung würde uns zu dem irrtümlichen Schluß verleiten daß X_1 eine positive Wirkung auf X_3 hat. Nun wirkt aber X_1 auch indirekt über X_2 auf X_3 und zwar negativ. Dies kann man nach der Vorzeichenregel durch Multiplikation der Vorzeichen ersehen. Wenn nun dieser negative indirekte Effekt quantitativ genauso groß ist wie der positive direkte Effekt, dann hebt diese Nebenwirkung den positiven Effekt X_1 auf X_3 vollständig auf. Mit anderen Worten führt demnach die Einführung einer Maßnahme (X_1) zu keiner Veränderung in X_3 wenn X_2 vorliegt (vgl. hierzu ausführlicher OPP, SCHMIDT, 1976, 147-152; 241-250).

Abb. 5: Dreivariablenmodell mit negativer Nebenwirkung der unabhängigen Variable



In dem Modell in Abb. 5 hat X_1 zwar einen positiven direkten Effekt auf die Zielvariable X_2 . Gleichzeitig hat jedoch X_1 einen negativen und unerwünschten Effekt auf die Zielvariable X_3 . Da das Auftreten unerwünschter Nebeneffekte bei Reformen eher die Regel als die Ausnahme ist, bildet die quantitative Ermittlung der Nebeneffekte eine notwendige Voraussetzung um durch weitere Maßnahmen solche unerwünschten Nebenwirkungen zu eliminieren oder wenigstens zu verringern.

b) inadäquate Verwendung von Signifikanztests

In den verschiedenen Arbeiten von CAMPBELL, STANLEY sowie von ZIMMERMANN wird zur Überprüfung der Wirkung von Variablen nur der Signifikanztest angeführt. Dies erscheint uns aber aus folgenden Gründen problematisch:

1. In vielen Fällen verfügt man nicht über echte Zufallsstichproben, womit die Voraussetzung der Anwendung von Signifikanztests wegfällt.
2. Die auf den üblichen Schätzverfahren basierenden Signifikanztests erlauben nur das Testen einer H_0 -Hypothese, ob ein Effekt vorhanden ist oder ob dies nicht der Fall ist. Über die Stärke der Effekte, also die Höhe einzelner Koeffizienten sowie die erklärte und unerklärte Varianz wissen wir dann noch nichts (vgl. auch BREDEKAMP, 1975).

Aus welchem Grund ist die Berechnung von Koeffizienten, sowie der erklärten Varianz der abhängigen Variablen wesentlich wichtiger als die Berechnung des Signifikanzniveaus? Erst mit Hilfe dieser Informationen erfahren wir, wie bedeutsam die Wirkung einzelner Variablen ist (durch Koeffizienten) und wie sicher wir der Wirkung sein können (durch Standardschätzfehler der Koeffizienten und unerklärte Varianz). Im Rahmen neuerer statistischer Ansätze, welche die Varianzanalyse als Spezialfall der Regressionsanalyse entwickeln, kann man diese Maße berechnen (vgl. z.B. KERLINGER, PEDHAZUR, 1973; GAENNSLEN, SCHUBÖ, 1973 und NAMBOODIRI et.al., 1975).

c) a-theoretischer Charakter von Quasiexperimenten

In den Darstellungen der quasiexperimentellen Versuchsordnungen bleibt der genaue Status der Beziehungen zwischen den Variablen oft unklar bzw. implizit (vgl. hierzu CAPORASO, ROOS, 1973, 31 ff.). Da man gerade auch bei Lösung praktischer Probleme bestimmte als allgemein angenommene Gesetze anwendet, liegt eine Explikation der Beziehungen zwischen den Variablen als Kausalbeziehung nahe. Einen entsprechenden Vorschlag haben auch COSTNER (1971) und ALWIN, TESSLER (1974) gemacht. Man kann

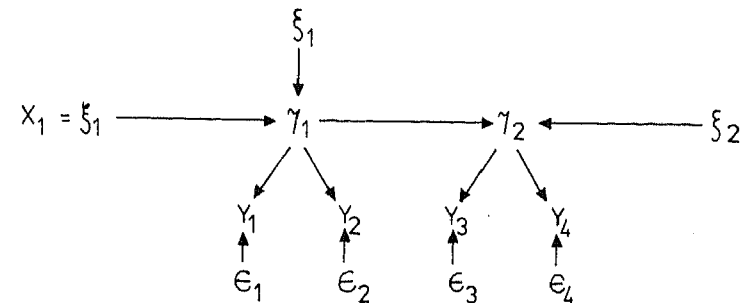
also diesem Kritikpunkt durch eine entsprechende Explikation der Beziehungen zwischen den Variablen als kausale Beziehungen genügen. Solche Hypothesen können dann deduktiv mit Verfahren der Kausalanalyse geprüft werden. Im Falle direkt gemessener Variablen handelt es sich um Pfadanalysen bzw. mehrstufige Regressionsanalysen (vgl. einführend OPP, SCHMIDT, 1976). Die Versuchspläne selber werden davon aber nicht berührt.

Allerdings sei an dieser Stelle auch darauf hingewiesen, daß gerade für die praktische Anwendung eine Reihe zusätzlicher ad-hoc Annahmen expliziert werden muß. Es handelt sich um das was ALBERT (1976, 182) technologische Phantasie nennt.

d) Nicht-Berücksichtigung von Meßfehlern und multiplen Indikatoren

Ein weiterer Kritikpunkt an den Darstellungen von quasiexperimentellen Versuchsplänen ergibt sich daraus, daß implizit angenommen wird, daß der Meßfehler Null sei, da ja nur statistische Verfahren angewendet werden, die von direkt gemessenen Variablen ausgehen, etwa Varianzanalyse und Regressionsanalyse. Da die Annahme eines Meßfehlers von Null empirisch durchweg falsch sein dürfte, benötigt man auch für quasiexperimentelle Verfahren einen allgemeineren statistischen Kalkül, der es erlaubt, Konstrukte mit multiplen Indikatoren einzuführen und damit das Ausmaß des Meßfehlers zu testen. Von COSTNER wurde (1971) hierfür zur Analyse experimenteller Daten die Pfadanalyse mit unbeobachteten Variablen empfohlen. ALWIN, TESSLER (1974) schlagen in Weiterentwicklung von COSTNER eine verallgemeinerte Form der Kovarianzanalyse von JÖRESKOG (1970) vor. Beides sind aber nur Spezialfälle des allgemeinen linearen Modells von JÖRESKOG (1973). Dieses allgemeine lineare Modell erscheint uns als besonders geeignet, um quasiexperimentelle Versuchspläne zu analysieren, da es auch den Kritikpunkten a-c Rechnung trägt, (zu einer deutschsprachigen Einführung vgl. SCHMIDT, GRAFF, 1975). In der einfachsten Form kann man diesen Modelltyp durch folgendes Pfaddiagramm darstellen:

Abb. 6: Kausalmodell eines Experiments mit Konstrukten und Indikatoren



- wobei X_1 = veränderte unabhängige Variable
 ξ_1, ξ_2 = theoretische Konstrukte
 γ_1, γ_2 = Indikatoren der Konstrukte (=Fragebogenitems)
 γ_3, γ_4 = Residuen der theoretischen Konstrukte (=Wurzel aus der unerklärten Varianz)
 ϵ_1, ϵ_2 = Meßfehler der Indikatoren

Diese Verallgemeinerung des quasiexperimentellen Ansatzes hat darüber hinaus noch den bereits früher angedeuteten Vorteil, daß man eher zu allgemein anwendbaren Ergebnissen kommt, da man je nach sozialer Situation wechselnde Indikatoren verwendet hat. Diese messen aber jeweils das gleiche Konstrukt. Daher sind dann zumindest die Beziehungen zwischen den theoretischen Konstrukten generell anwendbar, auch wenn die Korrespondenzregeln, welche die Konstrukte mit direkt meßbaren Indikatoren verknüpfen, u.U. jeweils neu festgelegt werden müssen (vgl. hierzu WESTMEYER, 1973 und 1976).

Um unsere allgemeinen Ausführungen zur Kritik der Analyse quasiexperimenteller Versuchspläne anschaulicher zu machen, werden wir nun im nächsten Abschnitt unser einfaches Anwendungsbeispiel 1 so verändern, daß es diesen Kritikpunkten Rechnung trägt.

4.4. Anwendungsbeispiel 1'

Im Folgenden wollen wir ein modifiziertes Modell zur Erklärung sachimmanenter Lernmotivationen darstellen (vgl. PORTELE, 1975). Es ist multivariat, d.h. es besteht aus mehreren abhängigen Variablen und mehreren unabhängigen Variablen. Es enthält Konstrukte und Indikatoren. Die Beziehungen zwischen den Variablen werden explizit als Ursache-Wirkungsbeziehungen formuliert. Der inhaltliche Unterschied zu Modell 1 besteht darin, daß wir zwei Variablen als intervenierende Variablen zwischen unsere unabhängige und unsere abhängige Variable einfügen. Es sind dies 'Strukturiertheit des Stoffes' und 'Beteiligung', erstere können wir mit Hilfe von semantischen Differenzialen folgender Art messen.

"Denken Sie jetzt bitte an den thematischen Inhalt der bisherigen Veranstaltungssitzungen. Unter thematischem Inhalt verstehen wir die Fakten und Ansichten, mit denen sie sich in der Veranstaltung beschäftigen. Was finden Sie? Der thematische Inhalt paßt zu:"

organisiert 1 - 2 - 3 - 4 - 5 - 6 - 7 unklar

Sechs weitere Items dieser Art werden zur Messung dieses Konstruktes verwendet.

Auf die Messung von 'Beteiligung' wollen wir nicht näher eingehen, sondern hier nur annehmen, daß man hierfür 2 Items verwendet. Wir können nun die angenommenen Beziehungen zwischen den Variablen anschaulich durch folgende Abbildung darstellen.

Im Rahmen des ersten Schrittes ordnen wir allen Personen der Experimentalgruppe eine 1 zu, während alle Personen der Kontrollgruppe eine 0 bekommen. Mit anderen Worten verwenden wir eine sog. Dummy-Variable, die nur zwei Ausprägungen haben kann, zur Kennzeichnung der beiden Gruppen.

Anschaulich lassen sich unsere Überlegungen so darstellen:

X_1	D_1
Versuchsgruppe	1
Kontrollgruppe	0

wobei X_1 die Variable 'Art der untersuchten Gruppe' ist und D_1 die aus X_1 gebildete Dummy-Variable mit den

Ausprägungen 1 = Versuchsgruppe und
0 = Kontrollgruppe

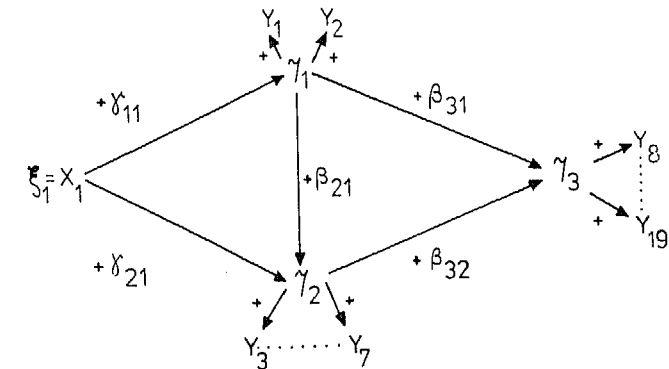
Da wir in der einen Gruppe von den Teilnehmern Aufgaben und Probleme erarbeiten lassen, in der anderen dagegen nicht, ist diese Dummy-Variable identisch mit der Variablen Partizipation. In der Versuchsgruppe ($D_1 = 1$) wird partizipiert. In der Kontrollgruppe ($D_1 = 0$) hingegen nicht. Die Residuen der theoretischen Konstrukte und der Indikatoren haben wir der Einfachheit halber weggelassen.

Wenn wir nun den einfachsten der vorgestellten quasiexperimentellen Versuchspläne wählen, können wir die Analyse folgendermaßen durchführen:

1. Schritt: Bildung einer Dummy-Variable für Versuchs- und Kontrollgruppe
2. Schritt: Aufstellung der Strukturgleichungen für dieses Quasiexperiment
3. Schritt: Berechnung der unbekanntenen Koeffizienten sowie deduktive Überprüfung von deren Vorzeichen

Der dritte Schritt gibt uns die Information, die man für die Evaluation von Maßnahmen braucht: Wirkt sich $x \dots$ tatsächlich so aus, daß unsere Zielvariable 'sachimmanente Lernmotivation' steigt?

Abb. 7:



wobei:

- X_1 = Partizipation
- η_1 = Beteiligung
- η_2 = Strukturiertheit
- η_3 = Intrinsische Motivation
- γ_1, γ_2 = Indikatoren von Beteiligung
- $\gamma_3 - \gamma_7$ = Indikatoren von Strukturiertheit
- $\gamma_8 - \gamma_9$ = Indikatoren von Intrinsischer Motivation

Befassen wir uns nun mit dem zweiten Schritt, nämlich der Aufstellung der Strukturgleichungen.

Wir müssen für jedes der drei theoretischen Konstrukte, Beteiligung, Stukturiertheit und Intrinsische Motivation eine Strukturgleichung aufstellen. Durch diese Strukturgleichungen stellt man in formaler Form die in dem Pfaddiagramm dargestellten Beziehungen zwischen den Variablen dar. Dabei steht in der jeweiligen Strukturgleichung die abhängige Variable, die erklärt werden soll, auf der linken Seite. Auf der rechten Seite der Gleichung stehen die gemäß der Theorie direkt einwirkenden unabhängigen Variablen.

Für die Beziehungen zwischen den Konstrukten und den Indikatoren müssen ebenso die entsprechenden Strukturgleichungen aufgestellt werden.

Wir wollen im Folgenden nur exemplarisch die Strukturgleichungen für die Beziehungen zwischen den theoretischen Variablen ausschreiben.

$$\begin{matrix} \gamma^{(1)} \\ \gamma^{(2)} \\ \gamma^{(3)} \end{matrix}$$

wobei ξ_1 die Dummy-Variablen 'Art der untersuchten Gruppe' bzw. Partizipation/Nichtpartizipation bedeutet. Verbal ausgedrückt besagt z.B. Gleichung (3) folgendes:

Intrinsische Motivation (γ_3) steht als abhängige Variable auf der linken Seite der Gleichung. Sie wird direkt beeinflusst durch Beteiligung (γ_1) und Stukturiertheit des Stoffes (γ_2). Hingegen wirkt Partizipation (ξ_1) nicht direkt auf Intrinsische Motivation was durch die 0 vor ξ_1 symbolisiert ist. Die Koeffizienten β_{31} und β_{32} geben im standardisierten Fall die Höhe des Einflusses von Beteiligung und Stukturiertheit auf Intrinsische Motivation an. Das Residuum ξ_3 symbolisiert den nichtdeterministischen Charakter der Gleichung (3).

In einem dritten Schritt wird man dann die Koeffizienten γ und β sowie die Koeffizienten der Beziehungen zwischen Konstrukten und Indikatoren entweder - bei deduktiver Vorgabe - an Daten testen oder induktiv an Daten gewinnen, wenn man nur die Vorzeichen festgelegt hat. Wegen des technischen Charakters der Vorgehensweise sei hier nicht näher darauf eingegangen. Wir verweisen auf JÖRESKOG (1973) und SCHMIDT, GRAFF (1975).

Wenn nun tatsächlich ein positiver Gesamteffekt von Partizipation auf Intrinsische Motivation sich empirisch ergibt, würde man die Maßnahme als vorläufig bewährt ansehen.

Man kann nun argumentieren, daß die per Fragebogen erhobenen Daten zu sehr durch Störvariable, z.B. soziale Wünschbarkeit beim Antwortverhalten, beeinflusst sind.

Wie kann man feststellen, ob ein solcher spezieller Methodeneffekt existiert?

Ein Vorschlag zur Lösung dieses Problems stammt von CAMPBELL, FISKE (1959) und besteht darin, die gleichen Variablen mit verschiedenen Methoden zu messen, und dann durch die Analyse der Korrelationsmatrix der Variablen zu testen, ob spezielle Methodeneffekte vorhanden sind. Die Matrix der Interkorrelationen der mit verschiedenen Methoden gemessenen Variablen wird Multitrait/Multimethod Matrix genannt. Eine Verallgemeinerung und Integration mit kausalanalytischen Verfahren wurde in jüngster Zeit vorgeschlagen (ALTHAUSER, 1974; ALWIN, 1974). Im folgenden wollen wir nun 2 zusätzliche Methoden zur Messung der Variablen unseres Modells diskutieren:

1. subjektive Eindrücke und Urteile
2. nichtreaktive Verfahren

Wenn man die Vorzeichen und die Höhe der Koeffizienten berechnet und somit die Daten für die Evaluation hat, sollte man diese mit den subjektiven Einschätzungen und Beobachtungen der Teilnehmer eines Handlungsforschungsprojektes vergleichen bzw. diese miteinander korrelieren.

Um eine Vergleichbarkeit zu erreichen, müßten sich diese aber auch auf die vorher genannten Variablen beziehen. Ob und wie man diese subjektiven Daten zusammenfaßt, wollen wir in diesem Zusammenhang offen lassen. Durch den Vergleich der Ergebnisse erzielt man aber eine zusätzliche Prüfung des Programms.

Wenn sich widersprechende Ergebnisse ergeben, wie im Falle des Programms Head Start, verdienen die subjektiven Daten nicht weniger Vertrauen oder Mißtrauen. Vielmehr kommt es darauf an, die Unterschiede theoretisch zu erklären (vgl. CAMPBELL, BORUCH, 1975, 119). Bezogen auf unser inhaltliches Anwendungsbeispiel kann man die entsprechende Multitrait-Multimethod-Matrix mit fingierten Daten so darstellen*: (der Einfachheit halber verwenden wir zunächst nur die Konstrukte)

Schema γ : Multitrait-Multimethod-Matrix mit 2 Methoden* *

	Variable	Methode 1			Methode 2		
		$\gamma'1=A_1$	$\gamma'2=B_1$	$\gamma'3=C_1$	$\gamma'1=A_2$	$\gamma'2=B_2$	$\gamma'3=C_2$
Methode 1	$\gamma 1=A_1$	(0,89)					
	$\gamma 2=B_1$	0,51	(0,89)				
	$\gamma 3=C_1$	0,38	0,37	(0,76)			
Methode 2	$\gamma 1=A_2$	0,57	0,22	0,09	(0,93)		
	$\gamma 2=B_2$	0,22	0,57	0,10	0,68	(0,99)	
	$\gamma 3=C_2$	0,11	0,11	0,46	0,59	0,58	(0,84)

Die Zuverlässigkeitskoeffizienten (eine Variable, eine Methode) sind eingeklammert.

Die Validitätskoeffizienten (eine Variable, verschiedene Methoden) sind unterstrichen.

Innerhalb dieser Matrix haben wir explizit die Korrelationen zwischen den verschiedenen Variablen eingetragen. Man kann nun auf Grund theoretischer Überlegungen Ableitungen machen, mit denen man prüft, ob bestimmte Methodeneffekte vorhanden sind (vgl. ALWIN, 1974; CAMPBELL, FISKE, 1959; LÜCK, 1976, 84).

1. Die unterstrichenen Gültigkeitskoeffizienten sollen sich signifikant von $r = 0,00$ unterscheiden und so hoch sein, daß sich weitere Validitätsuntersuchungen lohnen.

* Die Zahlen stammen aus dem Beispiel von BUNGARD, LÜCK (1976, 84)

** Für die weitere Diskussion haben wir zur Vereinfachung A, B und C statt der griechischen Buchstaben eingeführt

2. Die unterstrichenen Gültigkeitskoeffizienten sollen ferner höher sein als die Korrelationen für verschiedene Variablen, ermittelt mit verschiedenen Methoden in der gleichen Zeile bzw. Spalte (A_1B_2 , A_1C_2 , B_1C_2 , A_2B_1 , A_2C_1 und B_2C_1). Z.B. soll A_1A_2 sowohl höher sein als A_1B_2 und als A_1C_2 als auch B_1A_2 und C_1A_2 übersteigen.
3. Die unterstrichenen Gültigkeitskoeffizienten sollen höher sein als Korrelationen verschiedener Variablen unter Verwendung gleicher Methoden (A_1A_1 , A_1C_1 , A_2A_2 , A_2C_2 , B_2C_2).
4. Die Struktur der Interkorrelationen für verschiedene Variablen bei gleicher Methode soll der Struktur der Interkorrelation für verschiedene Variablen bei Anwendung verschiedener Methoden entsprechen (z.B. soll A_1B_1 , A_1C_1 , B_1C_1 der Struktur von A_1B_2 , A_1C_2 , B_1C_2 entsprechen).

Dieses Verfahren erlaubt also neben einer Kombination von subjektiven Daten und objektiven Daten den Test von Methodeneffekten. Daher scheint es für Handlungsforschungsprojekte besonders geeignet zu sein, in denen oft argumentiert wird, daß Fragebogen allein nicht ausreichend oder sogar irreführend sind. Gehen wir nun noch auf eine Erweiterung unserer eben dargestellten Matrix ein. Um die Methodeneffekte des Fragebogens und der subjektiven Einschätzungen zu kontrollieren, kann man auch sogenannte nichtreaktive Verfahren anwenden (vgl. BUNGARD, LÜCK, 1974, 1975; WEBB, 1975). Hierzu zählt die Verwendung von Archiven, Verhaltensbeobachtungen etc. In unserem Beispiel könnte man nicht nur fragen, wie die Beteiligung war, sondern auch die Beteiligung per Videorecorder aufnehmen. Ebenso könnte man das Ausmaß der intrinsischen Motivation nicht nur durch eine Serie von semantischen Differenzialen messen, sondern auch prüfen, wieviel Teilnehmer von sich aus eigene Papiere in die Gruppendiskussion eingegeben haben. Auch wenn die Entwicklung nicht-reaktiver Verfahren erst am Anfang steht, so wäre ihre Anwendung als alternative Methode sehr sinnvoll, um die Ergebnisse von Maßnahmen adäquat einzuschätzen. Eine Hinzufügung der nichtreaktiv gemessenen Variablen (γ_1'' , γ_2'' , γ_3'') unseres Modells zu der oben dargestellten Matrix würde dann schematisch zu folgender erweiterten Matrix führen:

Schema 7: Multitrait-Multimethod Matrix mit 3 Methoden

	γ_1	γ_2	γ_3	γ_1'	γ_2'	γ_3'	γ_1''	γ_2''	γ_3''
γ_1	Korrelation der								
γ_2	Fragebogenvariablen								
γ_3	untereinander								
γ_1		Korrelation der							
γ_2		Fragebogenvariablen		Korrelation der subjektiven					
γ_3		mit den subjektiven		Einschätzungen	untereinander				
γ_1			Korrelation der				Korrelation der		
γ_2			Fragebogenvariablen				subjektiven Ein-	schätzungen mit	
γ_3			mit den nichtreaktiv				nichtreaktiv ge-	messenen Variab-	len untereinan-
			gemessenen Variablen				der		

In der Hauptdiagonalen findet man nun die Kästchen mit Korrelationen der Variablen, die mit der gleichen Methode gemessen wurden. Außerhalb der Hauptdiagonalen findet man die Kästchen der Korrelationen der Variablen, die mit verschiedenen Methoden gemessen wurden. ALWIN (1974, 91-97) zeigt detailliert, wie man in dieser Art von Modellen die speziellen Methodeneffekte feststellen kann, und wir wollen hierauf in diesem Zusammenhang nicht näher eingehen. Allerdings muß hier noch darauf aufmerksam generell in Handlungsforschungsprojekten anwendbar sind (vgl. HAAG, 1972). Sofern bestimmte nichtreaktive Verfahren, wie z.B. Verwendung versteckter Kameras, es erforderlich machen, die betroffenen Personen nicht zu informieren, kann man diese Methoden nicht anwenden. Dies würde ja dem zentralen Postulat der Handlungsforschung widersprechen, alle betroffenen Personen zu informieren und gemeinsam über die anzuwendenden Methoden zu entscheiden. Mit diesem Vorschlag zur parallelen Anwendung von Fragebögen, subjektiven Einschätzungen und nichtreaktiven Verfahren im Rahmen quasiexperimenteller Versuchspläne wollen wir unsere Arbeit abschließen. Nach meiner Vorstellung wäre jetzt der nächste sinnvolle Schritt, solche Versuchspläne in konkreten Projekten aufzustellen und damit Erfahrungen zu sammeln, d.h. zu experimentieren*.

Schema 7: Multitrait-Multimethod Matrix mit 3 Methoden

	γ_1	γ_2	γ_3	γ_1'	γ_2'	γ_3'	γ_1''	γ_2''	γ_3''
γ_1									
γ_2									
γ_3									
γ_1'									
γ_2'									
γ_3'									
γ_1''									
γ_2''									
γ_3''									

* Einen Anfang in dieser Richtung hat der Verfasser durch Diskussionen mit A. GRUSCHKA von der WBK in Münster gestartet. Diese Diskussion wurde durch die vorhergenannte Arbeitsgruppe in Duisburg in Gang gesetzt.